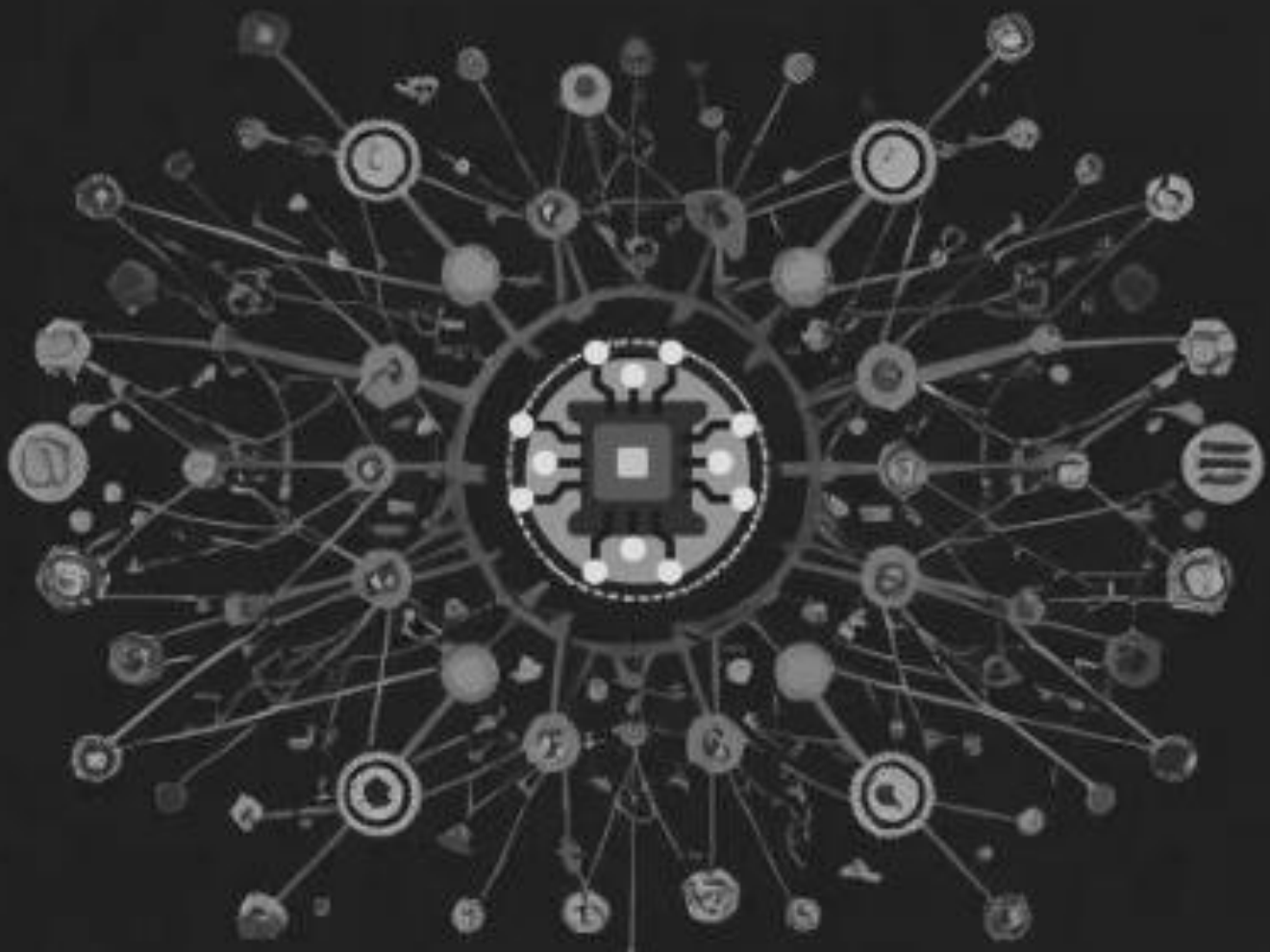
Interdisciplinary Applications of Artificial Intelligence - 1

Editors: Dr. Abdullatif Kaban - Dr. Turgay Demirel

ISTES

Interdisciplinary Applications of Artificial Intelligence - 1

Editors: Dr. Abdullatif Kaban - Dr. Turgay Demirel



ISTES

Editors

Assoc. Prof. Dr. Abdullatif Kaban, Atatürk University, Türkiye
Assoc. Prof. Dr. Turgay Demirel, Atatürk University, Türkiye

Cover Designed by
Canva

ISBN: 978-1-952092-91-6

© 2025, ISTES

The "*Interdisciplinary Applications of Artificial Intelligence - 1*" is licensed under a Creative Commons Attribution-NonCommercialShareAlike 4.0 International License, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Authors alone are responsible for the contents of their papers. The Publisher, the ISTES, shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of their search material. All authors are requested to disclose any actual or potential conflict of interest, including any financial, personal, or other relationships with other people or organizations regarding the submitted work.

Date of Publication

December, 2025

Publisher

ISTES
Monument, CO, USA

Contact

International Society for Technology, Education and Science (ISTES)
www.istes.org

Citation

Kaban, A. & Demirel, T. (Eds.). (2025). *Interdisciplinary Applications of Artificial Intelligence - 1*. ISTES.

CHAPTERS

PREFACE..... viii

Chapter 1- An Overview of Artificial Intelligence and Fundamental Concepts 1

Chapter 2- Generative Artificial Intelligence (GenAI) and Large Language Models..... 18

Chapter 3- Data Ethics, Privacy, and Secure Artificial Intelligence 40

Chapter 4- Machine Learning..... 59

Chapter 5- Artificial Intelligence in Research Processes 83

Chapter 6- An AI-Integrated Platform for End-to-End Meta-Analysis: Design, Implementation, and Evaluation of "tasresearch.org"..... 115

Chapter 7- Artificial Intelligence in Instructional Design and Technologies..... 137

Chapter 8- Use of Generative Artificial Intelligence in Developing Writing Skills in Language Education..... 161

Chapter 9- Artificial Intelligence in Education: Transformation, Applications, and Future Perspective 185

Chapter 10- Artificial Intelligence in Programming and Software Development..... 216

Chapter 11- Artificial Intelligence in the Internet of Things (IoT) and Intelligent Systems.. 233

PREFACE

Artificial intelligence has become an integral component of contemporary scientific research, educational practices, and technological development. Rather than functioning as a standalone technical field, artificial intelligence now operates as a cross-cutting framework that reshapes how knowledge is produced, interpreted, and applied across disciplines. This edited volume was conceived in response to the growing need for a structured, critical, and practice-oriented understanding of artificial intelligence and its expanding role in both research and education.

As editors, we aimed to bring together theoretical foundations, methodological perspectives, and application-oriented studies within a coherent and accessible framework. The chapters included in this book reflect a broad spectrum of approaches, ranging from fundamental concepts of artificial intelligence to advanced topics such as generative artificial intelligence, large language models, data ethics, and secure AI systems. Particular attention has been given to the integration of artificial intelligence into research processes, instructional design, and programming practices, highlighting both opportunities and emerging challenges.

This volume emphasizes that artificial intelligence should be understood not only through its technical capabilities but also through its implications for ethical responsibility, data governance, and human–AI interaction. Several chapters address these concerns directly by examining issues related to privacy, transparency, accountability, and human oversight. By doing so, the book seeks to encourage responsible and informed use of artificial intelligence technologies across diverse contexts.

Another distinguishing feature of this book is its interdisciplinary orientation. Contributions draw on perspectives from computer science, educational sciences, research methodology, and applied technologies, demonstrating how artificial intelligence can function as a unifying tool across domains. The inclusion of an AI-integrated platform for end-to-end meta-analysis further illustrates how artificial intelligence can support complex research workflows in practice.

This book is intended for researchers, graduate students, educators, instructional designers, and practitioners who seek a comprehensive yet conceptually grounded understanding of artificial intelligence. Rather than presenting artificial intelligence as a fully autonomous solution, the chapters collectively highlight the importance of human expertise, critical judgment, and contextual awareness in AI-supported systems.

We would like to express our sincere gratitude to all contributing authors for their valuable expertise and commitment. Their contributions have been essential in shaping this volume and ensuring its academic rigor and practical relevance. We also thank the reviewers and editorial collaborators whose feedback strengthened the clarity and coherence of the chapters.

We hope that this book will serve as a useful reference for ongoing research and practice, and that it will contribute to informed discussions on the responsible and meaningful integration of artificial intelligence into education, research, and software development.

Assoc. Prof. Dr. Abdullatif Kaban
Atatürk University
Erzurum
Türkiye
Contact e-mail: abdullatif.kaban@gmail.com

Assoc. Prof. Dr. Turgay Demirel
Atatürk University
Erzurum
Türkiye
Contact e-mail: turgaydemirel85@gmail.com

Citation

Kaban, A. & Demirel, T. (2025). Preface. In A. Kaban & T. Demirel (Eds.), *Interdisciplinary Applications of Artificial Intelligence - 1* (pp. viii-ix). ISTES.

Chapter 1- An Overview of Artificial Intelligence and Fundamental Concepts

Abdullatif Kaban

Chapter Highlights

- Artificial intelligence is examined through a comprehensive framework encompassing its conceptual foundations, historical development, and interdisciplinary nature.
- Machine learning and deep learning approaches are conceptually introduced, including learning paradigms and fundamental architectures.
- The data-driven operational logic of artificial intelligence systems is presented step by step, from training to inference.
- Human–AI interaction is discussed through human-in-the-loop systems and AI-supported decision-making frameworks.
- The limitations and risks of artificial intelligence are critically analyzed in terms of bias, explainability, and generalization challenges.

Introduction

Artificial intelligence is no longer considered merely a subfield of computer science; rather, it has emerged as a technology with broad interdisciplinary influence, extending from education and healthcare to engineering and the social sciences. The widespread adoption of algorithmic decision-making mechanisms, the integration of data-driven systems into many aspects of everyday life, and the transformation of human–machine interaction have made an understanding of the fundamental concepts of artificial intelligence more important than ever.

The primary aim of this chapter is to provide a conceptual framework for the field of artificial intelligence and to establish a shared terminology for readers. The meaningful evaluation of artificial intelligence applications in an interdisciplinary context requires a foundational awareness of what these technologies are, how they function, and what limitations they entail. For this reason, the chapter adopts an approach that avoids excessive technical detail while maintaining strong scientific foundations.

The chapter begins by examining the historical development and core definitions of artificial intelligence, followed by an explanation of the key components and subfields that constitute artificial intelligence systems. Machine learning and deep learning approaches are discussed under separate headings due to their central role in contemporary artificial intelligence applications. Subsequently, the operational logic of artificial intelligence systems and their modes of interaction with humans are addressed, emphasizing that these systems are not merely technical artifacts but also socio-technical constructs. The chapter concludes with a critical discussion focusing on the limitations, risks, and current debates surrounding artificial intelligence.

This introductory section aims to establish a common conceptual ground for the discipline-specific artificial intelligence applications addressed throughout the book. In doing so, it seeks to enable readers—regardless of their disciplinary background—to evaluate artificial intelligence–based approaches in a more informed, critical, and holistic manner.

The Emergence of the Concept of Artificial Intelligence

Artificial intelligence (AI) is an interdisciplinary field of research that aims to model and realize cognitive processes characteristic of human intelligence through computer systems. These processes include abilities such as learning, reasoning, problem solving, perception, and decision making. Artificial intelligence seeks to simulate these abilities through algorithmic and computational models (Russell & Norvig, 2022).

The scientific foundations of artificial intelligence were established in the mid-twentieth century alongside the development of computer science. One of the intellectual starting points of the field is Alan Turing's seminal 1950 paper "*Computing Machinery and Intelligence.*" In this work, Turing posed the question "*Can machines think?*" and, rather than attempting to answer it directly, proposed the imitation game —later known as the Turing Test— as a means of evaluating whether machines can exhibit intelligent behavior (Turing, 2004). This approach

provided artificial intelligence research with a functional criterion by suggesting that intelligence could be assessed through human-like behavior.

The formal recognition of artificial intelligence as an independent research field occurred with the Dartmouth Summer Research Project held in 1956. Organized by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon, this workshop marked the first systematic use of the term "artificial intelligence." McCarthy and his colleagues argued that every aspect of learning and intelligence could, in principle, be precisely described and simulated by a machine (McCarthy et al., 2006). This event positioned artificial intelligence as a research domain that brought together multiple disciplines, including computer science, mathematics, psychology, neuroscience, and philosophy.

Early artificial intelligence research in the 1950s and 1960s primarily relied on symbolic artificial intelligence and logical reasoning approaches. Systems developed during this period, such as the Logic Theorist and the General Problem Solver, aimed to imitate human-like reasoning processes in tasks involving mathematical proof and general problem solving (Newell & Simon, 2007). These studies were grounded in the assumption that intelligence could be represented through explicit rules and symbols, an assumption that shaped artificial intelligence research for many years.

This early developmental phase of artificial intelligence was not limited to technical advances alone; it also deepened philosophical and cognitive debates concerning the nature of human intelligence. In this context, artificial intelligence gradually evolved beyond being merely an engineering problem and emerged as a genuinely interdisciplinary field of inquiry. Contemporary developments such as generative artificial intelligence, along with ongoing ethical and security debates and application-oriented approaches, are built upon this historical and conceptual legacy.

Definition and Scope of Artificial Intelligence

The concept of artificial intelligence has been defined in various ways in the literature, depending on the perspectives of different disciplines. In general, artificial intelligence is understood as the ability of a machine to perform tasks that normally require human intelligence. These tasks include learning, reasoning, problem solving, perception, natural language use, and decision making (Russell & Norvig, 2022).

Definitions of artificial intelligence are commonly classified along two main axes:

- **Exhibiting human-like behavior**
- **Producing rational behavior**

Within this framework, some definitions emphasize machines that think or behave like humans, whereas others consider it sufficient for machines to produce optimal outcomes without imitating human cognition (Nilsson, 2009). Contemporary artificial intelligence research predominantly adopts the latter approach.

The Distinction Between Weak Artificial Intelligence and Strong Artificial Intelligence

One of the most frequently used distinctions in defining the scope of artificial intelligence is the differentiation between weak artificial intelligence (narrow AI) and strong artificial intelligence, also referred to as artificial general intelligence (AGI). Weak artificial intelligence refers to systems designed for a specific task or problem domain. Face recognition systems, recommendation algorithms, machine translation tools, and generative artificial intelligence models currently in use fall within this category (Boden, 2016).

Strong artificial intelligence, by contrast, denotes systems that possess generalizable cognitive abilities across multiple domains, comparable to human intelligence. It is widely acknowledged in the academic literature that such systems remain largely theoretical and have not yet been fully realized in practice (Goertzel & Pennachin, 2007). Consequently, nearly all contemporary artificial intelligence applications are considered instances of weak artificial intelligence.

Differences Between Artificial Intelligence, Automation, and Programming

The concept of artificial intelligence is often confused with automation and classical programming. Automation refers to the execution of predefined rules in a fixed and unchanging manner, whereas artificial intelligence systems are distinguished by their ability to learn from data and adapt to changing conditions (Domingos, 2015). Similarly, classical programming relies on rules explicitly defined by the developer, while in artificial intelligence systems these rules are largely derived from data. In this respect, artificial intelligence should be regarded not merely as a software technique, but as a data-driven, statistical, and learning-based approach.

Core Components of Artificial Intelligence Systems

The effective functioning of an artificial intelligence system depends on the interaction of four fundamental components:

- **Data:** Forms the foundation of the learning process in artificial intelligence models. Data quality, diversity, and representational power directly influence model performance (Kelleher et al., 2020).
- **Algorithm:** Refers to the mathematical and statistical methods used to extract meaningful patterns from data.
- **Model:** The structure that emerges as a result of training an algorithm on data and that is capable of generating predictions or decisions.
- **Computational Power:** Represents the hardware infrastructure required for processing large datasets and complex models.

The balance among these components is one of the key factors determining the performance and applicability of artificial intelligence systems.

The Expanding Scope of Artificial Intelligence

Initially focused on limited problem domains, artificial intelligence research has expanded significantly through interdisciplinary applications. In fields such as education, healthcare, social sciences, engineering, and research methodologies, artificial intelligence is used to support decision-making, analyze data, generate content, and enhance learning processes. This expansion necessitates viewing artificial intelligence not only as a technical tool but also as a transformative cognitive technology. Within this context, the definition and scope of artificial intelligence provide a conceptual framework for understanding generative artificial intelligence, ethical and security considerations, and application-oriented approaches addressed in subsequent chapters.

Subfields of Artificial Intelligence

Artificial intelligence is not a single method or technology; rather, it represents a broad research ecosystem composed of multiple subfields shaped by different types of problems and application purposes. These subfields differ in terms of both methodological approaches and application contexts, yet they often operate in close interaction with one another. The subfields discussed in this section aim to provide a conceptual roadmap for topics that will be examined in greater detail in subsequent chapters of the book. The relationship between artificial intelligence and the closely related concepts of machine learning and deep learning is presented in Table 1.

Table 1. Conceptual Relationship Between Artificial Intelligence, Machine Learning, and Deep Learning

Feature	Artificial Intelligence (AI)	Machine Learning (ML)	Deep Learning (DL)
Definition	General term for systems that perform tasks requiring human intelligence	AI subfield that improves performance through learning from data	ML approach based on multi-layer neural networks
Scope	Broadest concept	Subset of AI	Subset of ML
Learning	Not mandatory (may be rule-based)	Core characteristic	Core characteristic
Data Requirement	Low–high	Moderate–high	Typically very high
Model Structure	Rule-based or learning-based	Statistical / algorithmic	Artificial neural networks
Examples	Expert systems, planning	Regression, classification	CNN, RNN, Transformer

Machine Learning

Machine learning is one of the most fundamental and widely used subfields of artificial intelligence. In this approach, systems improve their performance on specific tasks by learning from data rather than relying on explicitly programmed rules (T. M. Mitchell, 1997). Machine learning encompasses different learning paradigms, including supervised, unsupervised, and reinforcement learning. Today, machine learning is extensively used in tasks such as prediction, classification, clustering, and pattern recognition.

Deep Learning

Deep learning is a machine learning approach based on multi-layer artificial neural networks and is capable of achieving high accuracy on large-scale datasets. It has enabled significant advancements particularly in computer vision, speech recognition, and natural language processing (LeCun et al., 2015). The widespread adoption of deep learning models, driven by increased computational power and the availability of large datasets, has also formed the foundation of generative artificial intelligence applications..

Natural Language Processing (NLP)

Natural language processing is a subfield of artificial intelligence focused on enabling computers to understand, generate, and interpret human language. Applications such as text analysis, sentiment analysis, machine translation, and chatbots represent core examples of this field (Jurafsky & Martin, 2025). In recent years, the development of large language models has positioned NLP at the center of generative artificial intelligence applications.

Computer Vision

Computer vision is a subfield of artificial intelligence that aims to enable machines to interpret and understand visual data, including images and videos. Typical applications include object recognition, face recognition, medical image analysis, and autonomous vehicles. With the widespread adoption of deep learning–based methods, this field has achieved high levels of accuracy and has assumed a critical role across disciplines such as engineering, healthcare, and security.

Expert Systems

Expert systems are artificial intelligence systems that represent human expert knowledge in a specific domain using rule-based structures. These systems constituted a significant portion of early artificial intelligence research and were widely used for decision support, particularly in medicine and engineering. Although they have largely been replaced by data-driven models, expert systems remain important for understanding the historical development of artificial intelligence.

Generative Artificial Intelligence

Generative artificial intelligence encompasses models capable of automatically producing content such as text, images, audio, and code. This approach has gained substantial momentum in recent years, particularly through large language models and deep learning–based generative architectures. Generative artificial intelligence will be examined in detail in later chapters of this book; here, only a preliminary conceptual framework is provided.

The Interdisciplinary Position of Artificial Intelligence

The subfields summarized above clearly illustrate why artificial intelligence is regarded as an interdisciplinary technology. A wide range of applications—from education and cybersecurity to research methodologies and IoT-based intelligent systems—are built upon different combinations of these subfields.

Introduction to Machine Learning: Types of Learning

Machine learning is a fundamental approach that enables artificial intelligence systems to improve their performance through experience. Mitchell (1997) defines learning as a process in which a computer program improves its performance on a given class of tasks, as measured by a specific performance criterion, through experience. This definition remains one of the most widely accepted conceptual frameworks in the machine learning literature. Machine learning methods are categorized into different types depending on the nature of the data used during learning and the feedback mechanisms involved. In this section, four commonly used learning paradigms are examined.

Supervised Learning

Supervised learning is a learning approach in which input data are provided together with corresponding correct outputs (labels). During the training process, the model learns input–output mappings and uses this knowledge to generate predictions for new data. Classification and regression problems represent the most common applications of supervised learning (Hastie et al., 2009). Supervised learning has been successfully applied across fields such as education, healthcare, and the social sciences such as student performance prediction or medical diagnosis.

Unsupervised Learning

In unsupervised learning, the dataset is unlabeled, and the model is expected to discover hidden structures or patterns within the data. Clustering and dimensionality reduction methods fall under this category. Unsupervised learning plays a crucial role particularly in exploratory data analysis and data preprocessing stages (Aggarwal, 2018). This approach enables the identification of meaningful structures in large and complex datasets without prior knowledge.

Semi-Supervised Learning

Semi-supervised learning is an approach that combines a small amount of labeled data with a large amount of unlabeled data. This method offers significant advantages in situations where labeling costs are high (Zhu & Goldberg, 2009). Semi-supervised learning has gained increasing importance, particularly in interdisciplinary studies where labeled training data are limited.

Reinforcement Learning

Reinforcement learning refers to a learning paradigm in which an agent learns by interacting with its environment through reward and punishment mechanisms. In this approach, the model aims to maximize long-term cumulative reward (Sutton et al., 1998). Reinforcement learning is widely used in applications such as robotics, game-playing artificial intelligence systems, and autonomous decision-making.

Training, Validation, and Testing Processes

In the development of machine learning models, datasets are typically divided into training, validation, and test subsets. This division is used to evaluate the model's generalization capability and to reduce the risk of overfitting (Goodfellow et al., 2016).

Introduction to Deep Learning and Artificial Neural Networks

Deep learning, regarded as a subfield of machine learning, encompasses methods that aim to automatically learn data representations through multi-layer artificial neural networks. Deep learning approaches have assumed a decisive role in artificial intelligence research, particularly due to their superior performance on high-dimensional and unstructured data such as images, audio, and text (LeCun et al., 2015).

Basic Structure of Artificial Neural Networks

Artificial neural networks (ANNs) are computational models inspired by biological neural systems. In their most basic form, a neural network consists of three main components:

- Input layer,
- One or more hidden layers,
- Output layer.

Neurons within each layer compute weighted sums of their inputs and transform these sums through an activation function to produce outputs (Haykin, 2009). This structure enables the modeling of non-linear relationships and allows the learning of more complex patterns compared to classical statistical methods.

Distinctive Characteristics of Deep Learning

The primary characteristic that distinguishes deep learning from traditional machine learning approaches is its ability to perform feature extraction automatically. While traditional methods require features to be manually defined by domain experts, deep learning models can learn hierarchical representations directly from raw data (Goodfellow et al., 2016).

Through this hierarchical structure:

- Lower layers represent simpler patterns,
- Higher layers represent more abstract and complex concepts.

For example, in image processing applications, early layers may learn basic visual features such as edges and corners, whereas higher layers may represent object parts and complete object concepts.

Fundamental Deep Learning Architectures

Some of the most commonly used architectures in deep learning include:

- **Feedforward Neural Networks:** Basic structures in which information flows unidirectionally from the input layer toward the output layer.
- **Convolutional Neural Networks (CNNs)**: Networks designed to capture spatial features and achieve high performance, particularly in image and video data (Krizhevsky et al., 2012).
- **Recurrent Neural Networks (RNNs):** Models used for sequential data involving temporal dependencies. Advanced variants such as Long Short-Term Memory (LSTM) networks aim to mitigate the long-term dependency problem (Hochreiter & Schmidhuber, 1997).

These architectures form the foundation of interdisciplinary applications due to their adaptability to different data types and problem domains.

Limitations of Deep Learning

Despite offering high accuracy, deep learning models exhibit several important limitations:

- The requirement for large amounts of labeled data,
- High computational costs,
- Limited interpretability of model decisions.

These limitations necessitate careful consideration in the design and deployment of deep learning models and have motivated the development of new methods focused on explainability and efficiency (Lipton, 2018).

Operational Logic of Artificial Intelligence Systems

Although the operation of artificial intelligence systems may appear complex on the surface, it is fundamentally based on a data-driven and sequential process. While technical details may vary across different subfields of artificial intelligence, the overall workflow follows a common structure. In this section, the operational logic of artificial intelligence systems is explained through a conceptual framework and then illustrated using an example scenario.

Core Processing Workflow in Artificial Intelligence Systems

The operation of an artificial intelligence system can be described through the following fundamental stages:

Data → Preprocessing → Feature Representation → Model Training → Evaluation → Inference

- **Data Collection:** Data constitute the starting point of artificial intelligence systems. These data may be numerical, textual, visual, or auditory in nature. The representativeness and accuracy of the data directly affect the performance of the developed model (Kelleher et al., 2020).

- **Data Preprocessing:** Raw data are often incomplete, noisy, or inconsistent. Therefore, preprocessing steps such as cleaning, normalization, transformation, and, when necessary, labeling are applied. This stage is critical to the reliability of artificial intelligence systems (Hastie et al., 2009).

- **Feature Representation:** Data must be represented in a form suitable for the model. In traditional machine learning, feature representation is largely defined by humans, whereas in deep learning approaches, features are predominantly learned automatically by the model itself (Goodfellow et al., 2016).

- **Model Training:** At this stage, the algorithm learns patterns within the data and constructs a model. The training process involves optimizing model parameters with respect to a predefined objective function.

- **Model Evaluation:** The performance of the trained model is assessed using previously unseen data. This evaluation process enables the identification of issues such as overfitting and underfitting.

- **Inference:** Once training is complete, the model generates predictions or decisions for new data encountered in real-world applications. The primary functional value of artificial intelligence systems emerges at this stage.

Difference Between Training and Inference Processes

Two concepts that are frequently confused in artificial intelligence systems are training and inference. The training process requires substantial computational resources and large datasets, whereas the inference process is typically faster and can operate in real-time systems. This distinction is particularly critical in large language models and generative artificial intelligence applications (Russell & Norvig, 2022).

Error, Bias, and the Generalization Problem

The performance of an artificial intelligence system cannot be evaluated solely based on accuracy metrics. Biases present in the data may cause the model to produce systematic errors for specific groups. Furthermore, excessive adaptation to training data may result in poor performance on new data, a phenomenon known as the generalization problem (Domingos, 2015).

Example Scenario: An Artificial Intelligence System in Education

Consider an educational institution that aims to predict students' achievement levels in mathematics courses and to identify at-risk students at an early stage.

- **Data:** Students' exam scores, attendance records, and homework performance
- **Preprocessing:** Handling missing data and scaling scores
- **Feature Representation:** Identification of variables reflecting academic performance
- **Model Training:** Development of a classification model using supervised learning
- **Evaluation:** Measurement of model accuracy and error rates on test data
- **Inference:** Risk prediction for students in a new academic term

In this scenario, the artificial intelligence system functions as a decision-support tool rather than replacing the teacher. The collaboration between human expertise and the system forms the foundation of reliable and ethical use.

Human–Artificial Intelligence Interaction and Oversight

Modern artificial intelligence systems are typically designed not as fully autonomous structures but as human-supervised or human-in-the-loop systems. This approach aims to enhance system transparency and reliability (Amershi et al., 2019).

Artificial Intelligence and Human Interaction

The real-world impact of artificial intelligence systems is directly related not only to their technical accuracy but also to how they interact with humans. In this context, artificial intelligence is most often considered within the framework of human–AI interaction rather than as an independent decision-maker. Particularly in complex, uncertain, or ethically sensitive problem domains, the role of humans within the system loop is of critical importance.

Human-in-the-Loop Artificial Intelligence

Human-in-the-loop artificial intelligence (HITL) refers to system designs that require human intervention at specific stages of the decision-making process. In this approach, artificial intelligence models generate recommendations or predictions, while the final decision is made by a human, or the system is continuously updated through human feedback (Amershi et al., 2019).

The HITL approach is particularly effective as a control mechanism in:

- Data labeling processes,
- Critical decision-support systems,
- Applications with a high risk of producing erroneous or biased outputs.

Through this mechanism, both model performance is improved and human oversight is preserved.

AI-Supported Decision-Making Systems

AI-supported decision-making systems are designed not as replacements for human decision-makers but as tools that support them. These systems analyze large datasets to reveal potential scenarios, risks, or trends, thereby reducing the cognitive load on decision-makers (Shrestha et al., 2019).

The primary objective of such systems is:

- Not the full automation of decision-making,
- But the establishment of a balanced collaboration between human intuition and algorithmic analysis.

Research indicates that hybrid systems in which humans and artificial intelligence collaborate can produce more consistent outcomes than approaches based solely on human judgment or purely algorithmic decision-making (Bansal et al., 2021).

Levels and Limits of Autonomy

The autonomy of artificial intelligence systems refers to the extent to which human intervention is required for a given task. Fully autonomous systems are capable of making and executing decisions without human input, whereas semi-autonomous systems require human supervision and approval.

However, higher levels of autonomy may also introduce risks such as:

- Widespread impact of erroneous decisions,
- Ambiguity in accountability,
- Reduced user trust.

For this reason, the level of autonomy must be carefully determined based on the application context and intended use (Parasuraman et al., 2000).

From Autonomy to Responsibility and Accountability

In environments where artificial intelligence systems operate alongside humans, the question of responsibility is becoming increasingly significant. When a decision is made based on an AI-generated recommendation, it may be unclear whether the source of an error lies with the human decision-maker or the system itself. This ambiguity increases the importance of principles such as (a) transparency, (b) traceability, and (c) the ability to provide explanations to users in AI system design (Floridi et al., 2018). Within the context of human–AI interaction, responsibility has evolved into a multi-layered discussion that extends beyond technical considerations to encompass legal and ethical dimensions.

Limitations, Risks, and Contemporary Debates in Artificial Intelligence

While artificial intelligence systems offer significant opportunities across many domains, they also involve various limitations and risks at technical, social, and epistemological levels. For this reason, artificial intelligence should be considered not merely as a technological advancement but as a socio-technical system with multidimensional implications.

Data Dependency and the Problem of Bias

The performance of artificial intelligence systems is largely dependent on the data on which they are trained. Deficiencies, imbalances, or historical biases present in training data may lead to systematic errors in model outputs. This issue has brought concerns of fairness and equity to the forefront, particularly in applications with social implications (Barocas & Selbst, 2016).

Although data-driven learning processes are often perceived as “objective,” decisions regarding how data are collected, which variables are selected, and how labels are defined inherently involve human judgment. Consequently, artificial intelligence systems should not be regarded as entirely neutral or unbiased structures.

The Challenge of Explainability and Interpretability

Despite their high predictive accuracy, deep learning–based models are often described as “black box” systems due to the difficulty of understanding their internal mechanisms. This characteristic leads to challenges in explaining why a model produces a particular output (Lipton, 2018).

A lack of explainability may:

- Reduce user trust,

- Complicate the identification of erroneous decisions,
- Limit the use of such systems in critical domains.

As a result, contemporary artificial intelligence research increasingly emphasizes not only performance improvements but also the interpretability and transparency of model behavior.

Generalization and Context Sensitivity

Artificial intelligence models are typically trained on specific data distributions and may experience performance degradation when applied outside these distributions. Real-world problems, however, are often dynamic, uncertain, and highly context-dependent. This mismatch introduces significant limitations in the generalizability of artificial intelligence systems across varying conditions (M. Mitchell, 2019). Success in a specific task or dataset does not guarantee comparable performance in a different context. Therefore, contextual awareness should be considered a critical factor in the deployment and evaluation of artificial intelligence systems.

Contemporary Debates and Research Directions

Current artificial intelligence research is increasingly focused not only on developing more powerful models but also on designing systems that are more reliable, transparent, and human-centered. Prominent research directions in this context include:

- Explainable artificial intelligence approaches,
- Models capable of learning from limited data,
- Hybrid systems that emphasize human oversight,
- Model designs that are sensitive to ethical principles.

These ongoing debates demonstrate that the future of artificial intelligence will be shaped not only by technical advancements but also by decisions regarding how, where, and within which boundaries these technologies are deployed.

Acknowledgements or Notes

Disclosure of AI usage

During the preparation of this work, the author used ChatGPT and NotebookLM in order to generate text, check grammar, analyze data, or assist with literature search. After using this tool/service, the author reviewed and edited the content as needed and take full responsibility for the content of the published work.

References

Aggarwal, C. C. (2018). Machine Learning for Text: An Introduction. In C. C. Aggarwal (Ed.), *Machine Learning for Text* (pp. 1–16). Springer International Publishing. https://doi.org/10.1007/978-3-319-73531-3_1

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for Human-AI Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300233

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D. (2021). Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3411764.3445717

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *Calif. L. Rev.*, *104*, 671.

Boden, M. A. (2016). *AI: Its Nature and Future*. Oxford University Press.

Domingos, P. (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Goertzel, B., & Pennachin, C. (Eds.). (2007). *Artificial General Intelligence*. Springer. https://doi.org/10.1007/978-3-540-68677-4

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. https://mitpress.mit.edu/9780262035613/deep-learning/

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer series in statistics New-York.

Haykin, S. (2009). *Neural networks and learning machines, 3/E*. Pearson Education India.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jurafsky, D., & Martin, J. H. (2025). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models* (3rd ed.). https://web.stanford.edu/~jurafsky/slp3/

Kelleher, J. D., Namee, B. M., & D'Arcy, A. (2020). *Fundamentals of Machine Learning for Predictive Data Analytics, second edition: Algorithms, Worked Examples, and Case Studies*. MIT Press.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, *25*. https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lipton, Z. C. (2018). The Mythos of Model Interpretability. *Queue*, *16*(3), 31–57. https://doi.org/10.1145/3236386.3241340

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI Magazine*, *27*(4), 12–12. https://doi.org/10.1609/aimag.v27i4.1904

Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. Penguin UK.

Mitchell, T. M. (1997). Does Machine Learning Really Work? *AI Magazine*, *18*(3), 11–11. https://doi.org/10.1609/aimag.v18i3.1303

Newell, A., & Simon, H. A. (2007). Computer Science as Empirical Inquiry: Symbols and Search. In *ACM Turing Award Lectures* (p. 1975). Association for Computing Machinery. https://doi.org/10.1145/1283920.1283930

Nilsson, N. J. (2009). *The Quest for Artificial Intelligence*. Cambridge University Press.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Russell, S. J., & Norvig, P. (with Chang, M., Devlin, J., Dragan, A., Forsyth, D., Goodfellow, I., Malik, J., Mansinghka, V., Pearl, J., & Wooldridge, M. J.). (2022). *Artificial intelligence: A modern approach* (Fourth edition, global edition). Pearson.

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational Decision-Making Structures in the Age of Artificial Intelligence. *California Management Review*, *61*(4), 66–83. https://doi.org/10.1177/0008125619862257

Sutton, R. S., Barto, A. G., & others. (1998). *Reinforcement learning: An introduction* (Vol. 1). MIT press Cambridge.

Turing, A. (2004). Computing Machinery and Intelligence (1950). In B. J. Copeland (Ed.), *The Essential Turing* (pp. 433–464). Oxford University PressOxford. https://doi.org/10.1093/oso/9780198250791.003.0017

Zhu, X., & Goldberg, A. (2009). *Introduction to Semi-Supervised Learning*. Morgan & Claypool Publishers.

Author Information

Abdullatif Kaban

https://orcid.org/0000-0003-4465-3145

Ataturk University

Yakutiye / Erzurum

Turkiye

Contact e-mail: abdullatif.kaban@gmail.com

Citation

Kaban, A. (2025). An overview of artificial intelligence and fundamental concepts. In A. Kaban, & T. Demirel (Eds.), *Interdisciplinary Applications of Artificial Intelligence - 1* (pp. 1-17). ISTES.

Chapter 2- Generative Artificial Intelligence (GenAI) and Large Language Models

Önder Yıldırım

Chapter Highlights

- This chapter examines generative artificial intelligence (GenAI) and large language models (LLMs) within their historical, theoretical, and technical contexts.
- It demonstrates that large language models are the outcome of an evolutionary trajectory extending from rule-based and statistical language models to contemporary transformer-based architectures.
- The chapter emphasizes that training, fine-tuning, and alignment processes constitute the core mechanisms shaping LLM behavior and reliability.
- Prompt engineering is discussed as a central interaction and control layer in the design of LLM-based systems.
- The chapter explains agent-based LLM systems and their roles in financial, industrial, and enterprise applications.
- The opportunities and limitations of LLMs in the contexts of healthcare, education, and human–machine interaction are examined from an interdisciplinary perspective.
- The computational costs, environmental impacts, and sustainability challenges associated with large language models are evaluated through efficiency-oriented approaches.
- By addressing issues of ethics, digital inequality, and societal transformation, the chapter highlights the importance of a responsible and human-centered AI paradigm.
- Finally, the chapter argues that the future of LLMs depends not on building ever-larger models, but on developing more informed, more efficient, and more responsible systems.

Introduction

Technological advances in artificial intelligence have become transformative not only in terms of technological innovation, but also since they have altered the conditions under which knowledge is generated and processed, decisions are made and human-machine-interaction takes place. GenAI (Generative Artificial Intelligence), lies at the heart of this change – Instead of being your regular AI system that is generally used for classification/prediction, GenAI refers to models that generate any type of content (Text, Image, Audio and Code). In this wave of AI development, Large Language Models (LLMs) are perhaps the most prominent - and certainly highest profile subsystems in the GenAI lair (Naveed et al., 2025; Shanahan, 2024; Xiao & Zhu, 2025).

GPT-2 and other language modules are based on models that have been trained on massive textual corpora, and whose task is to predict the next word of a sentence, given a context. This method allows LLMs to model complex linguistic patterns with high fidelity and to produce outputs that look humanly generated. But it has also led to many myths and popular articles that claim these models "understand", "think" or have an agency. Indeed, current research highlights the fact that LLMs are essentially statistical (or probabilistic) systems devoid of consciousness, intentionality or conceptual cognition as found in human beings (Shanahan, 2024).

The societal effects of generative AI and big language models go beyond CS to include education, health care, law, economics, social science research and the workplace. In an educational context, LLMs open up exciting possibilities to create learning resources, deliver automated feedback, and support tailored learning (Baldassarre et al., 2023). Currently, in the medical domain, they are increasingly promoted as adjuncts to summarising clinical documentation, aiding patient–doctor interaction and helping with decision support. Likewise, the use of LLMs are becoming increasingly popular across fields like law, finance and public administration, where we see applications in automating and augmenting text-heavy tasks (Storey et al., 2025; Naveed et al., 2025; Teubner et al., 2023).

However, the AI and LLMs that GenAI offers also come with a variety of risks and concerns that are at the center of current academic discussions. Problems like hallucination – i.e., the production of things wrong or made up, data-driven biases, lack of explain ability, and high computational demands are often seen as important barriers to the trustworthy and responsible deployment of such technologies (Chang et al., 2024; Gallegos et al., 2024). Furthermore, the consequences of generative AI for labor markets, environmental sustainability and social equity require that these technologies be analyzed not only as technical systems but also ethico-socio-economic phenomena.

The focus of this chapter is to discuss generative artificial intelligence and large language models in relation to the guiding concept, challenges, consequences in terms of fields and interdisciplinary settings (and long-term compensation), interconnections between behind-the-scenes processes product-placed on a wide variety of domains where those methods have their biggest impact, but also keep things by talking about what they can not do. Hence, the chapter follows a systematic move through historical and theoretical linkages of LLMs first, then

examines imitation learning training processes and alignments, application domains, ethical issues involved and future research trends.

The Path Toward Large Language Models: Conceptual and Historical Background

While the rise of giant language models are typically cast as rapid technological breakthroughs, their development have built on decades of deeper knowledge in natural language processing, statistical modeling and human cognition. Modern work places large language models as the latest artifact in a long arc of evolution from rule-based systems to statistical language model to deep learning–based methods (Naveed et al., 2025; Zhao et al., 2023). By the same token Understanding LLMs also entails beyond their architectures and focuses on other conceptual and historical underpinnings of which today s architectures, in turn, are based on (Xiao & Zhu, 2025).

From Rule-Based and Statistical Language Models to Neural Approaches

Early rule-based system of natural language processing. In these systems, language was modeled by means of fixed syntactic and semantic rules but such rule-based approaches were insufficient to cope with the intrinsic flexibility and ambiguity of natural language. The limitations of rule-based system inspired researchers with the probabilistic views of language, hence statistical language models (Xiao & Zhu, 2025).

Statistical methods such as language models calculate the likelihood of a sequence of words given earlier text. In particular, the n-gram-based models were marked by a major breakthrough in local contextual correspondence at that time; however they suffered from two structural drawbacks: i) short-range based and ii) were unable to manage dependencies over long range of text. These shortcomings pointed to the necessity of modeling language not as mere surface-level statistics, but rather more abstract and generalizable representations. This was solved by taking into account the accumulation of disabilities of language in context and learning language representation in continuous vector space, which leads to a huge landmark in NLP (Zhao et al., 2023).

Language, Meaning and Modelling: Theoretical Perspectives

Language model development has a lot to do with computational power but it is also tightly related to theoretical assumptions on language and meaning. Traditional (and I would suggest obsolete) theories of natural language understanding did not view meaning as a single fixed entity but, rather, as a collection of context-sensitive possible meanings (Bender & Koller, 2020). An example of the model-theoretic generation approach has an explanation for an expression being in its description of the set of assumptions in which that expression is true (Tuomela, 1972). Under this view, meaning is regarded as a dynamic structure activated by context, rather than a stable image.

This paradigm has striking similarities to modern large language models. Instead of generating a unique “correct” meaning, LLMs provide statistical samples over plausible continuations that are conditioned on context. This

feature highlights that production acts over probabilities distributions and not deterministic rules, and that meaning is updated at each token given context history and evidence (Naveed et al., 2025). Thus the operational logic of LLM shares a conceptual history with earlier theoretical models casting meaning as an emergent property of models.

Deep Learning, Transformer Architectures, and Scaling

With the rise of deep learning techniques, the neural language model also made great progress. At first, recurrent neural networks and their LSTM variant had remarkable performances in modeling sequential language structures. But their capability of parallel computation and difficulty in modeling long-range context hindered their scaling ability (Naveed et al., 2025, Xiao & Zhu, 2025; Zhao et al., 2023).

Transformer-like models have played a key role in that regard. Linguistically dependent features were better learned by modeling the linguistic dependencies with attention mechanisms in the transformer architecture (Vaswani et al., 2017). This architectural leap enabled the simultaneous scaling of model size, training data and compute resources, which in turn directly leveraged the rise of large language models (Naveed et al., 2025; Zhao et al., 2023).

Scaling law observations show that the model performance is determined not only by architecture design but also by number of parameters, diversity of data and computation power. As they exceed certain scaling thresholds, qualitatively novel behaviors -often coined as emergent capabilities- have been observed. These qualitative changes are in-context learning, instruction following, and multi-step reasoning (Naveed et al., 2025).

From Historical Continuity to Contemporary Large Language Models

This historical and conceptual landscape suggests that large language models derive from neither explicit rules nor statistics unadorned. And it is not LLMs that stand alone, but sit at the end of a much longer chain of body-of-knowledge from rule-based systems though statistical modelling, to theoretical arguments about meaning, to developments in deep learning architectures. Accordingly, we suggest that large language models should not be conceived of as simply “large and powerful systems”, but rather as the latest articulations of prolonged research traditions in language, meaning and cognition. This view lays the basis for our discussion of generative artificial intelligence (GenAI), underscoring that GenAI is not just a technical invention but rather an answer to established theoretical and methodological practices which precede it (Tuomela, 1972; Naveed et al., 2025; Xiao & Zhu, 2025; Zhao et al., 2023).

Generative Artificial Intelligence (GenAI): Definition, Scope, and Model Types

Recent works have referred to Generative Artificial Intelligence (GenAI) as the concept under which models able to produce new and original content based in previous information can be placed with respect of previous classical approaches of artificial intelligence. The content may be any of a number of types, including text, images, audio,

video and code. The distinct feature of GenAI systems is their capacity to produce novel samples by conditioning on learned probabilistic distributions instead of learning input–output relationships via only classification or regression (Naveed et al., 2025).

The idea of GenAI has been getting a lot of attention recently, especially with the explosion in popularity of large language models. Yet this increased visibility has also led to a degree of conceptual blurring and generalization in the literature. In such a context, specifying the scope and limits of GenAI is essential for discussions in technical as well as moral fields (Xiao & Zhu, 2025).

What Generative Artificial Intelligence Is—and Is Not

Generative AI is sometimes used as an overarching term for all types of AI; rather, it refers to a specific type of model with its own unique generative logic. In traditional taxonomy, artificial intelligence models can be classified as discriminative and generative models. While discriminative models concentrate on the decision boundary between the input and its label, a generative model tries to form an estimate of the underlying stochastic structure of data in order to generate new samples (Zhao et al., 2023).

From this viewpoint, GenAI not only refers to systems that “generate outputs”, but systems where generation is based on learned data distributions. The most sophisticated versions of that approach in the text domain are large language models themselves; however, GenAI is not restricted to LLMs. Particularly, diffusion models and VAE/GAN based image generation methods are the building blocks of the larger GenAI system (Naveed et al., 2025).

This is not to say that it has been proven impossible for GenAI systems to be creative or understanding in a manner that is indistinguishable from human creativity or understanding; however, wild and sweeping assertions to the effect have no place in serious discussion. Recent critical positions point out that meaning does not get "created" but instead is constrained by statistics. If SJWs reproduce in the future as well, pointless things would be left because those systems can't tell you, "I don't have an answer for that," but dumb LTE's (low-level thinking engines) like SJWs can do, thus GenAI is better thought of as pattern generation with high capacity than “creative intelligence” (Shanahan, 2024).

Scope and Core Capability Domains of Generative Artificial Intelligence

What GenAI systems can do very widely both in kind of content generated and application domain . Language generation, summarization, translation, and question answering are some of the most well-known skills displayed by big LMs. Moreover, the use of GenAI systems for code generation and software development is continuously breaking down the borders between natural language and programming language (Teubner et al., 2023).

In addition to textual uses of GenAI, Theall has soared into the visual creation space, in synthesizing speech and multimodal investigations. Such image generation model and text-to-image models based on diffusion may also show that GenAI is not limited to a single modality but can be ported to linguistic domain as well. Existing evidence leads us to believe that this multimodal ability could enhance the effectiveness and generalization capabilities of a GenAI system (Naveed et al., 2025).

Yet, GenAI's versatile functionality sets up considerable challenges for reliability and control. The correctness, source, and context-sensitivity of generated outputs cannot always be guaranteed, therefore such outputs need to be used with caution and restraint especially in knowledge-rich domains (Chang et al., 2024).

Generative Model Types and the Position of Large Language Models

There are different types of generative models based on various generation mechanisms in the GenAI ecosystem. Variational autoencoders (VAEs) generate new samples by introducing latent representations of the data, while generative adversarial networks (GANs) create—what is hopefully realistic—output by having a generator model compete against a discriminator network in an adversarial game. Such diffusion models that produce data via iterative denoising have also shown exceptional performance in video generation (Naveed et al., 2025; Zhao et al., 2023).

Among this family of generative models, large language models are unique members as text-focused system that are powered by transformer architectures. Learning syntactic and semantic patterns of language at a large scale, LLMs can potentially accomplish many tasks within a single unified learning framework. These features place LLMs as generative systems into the overall GenAI ecosystem (Zhao et al., 2023).

Recent research also suggests that the capabilities of very large language models may go beyond only linguistic patterning to include more general pattern learning and generalization. This observation indicates that LLMs might progressively operate not only as text generating systems, but also as materials of more extensive cognitive and decision-support architectures (Naveed et al., 2025).

In this lens, it should be possible to interpret the relationship of generative AI and LLMs as something other than that of a straightforward hierarchical inclusion; genAIs occupy a central but limited place in the GenAI ecosystem. In the next section we will see why this difference is an important conceptual underpinning for the reasons that training, alignment and control mechanisms — to be discussed in Section 3 above — are necessary for the responsible deployment of generative AI systems (Naveed et al., 2025; Xiao & Zhu 2025; Shanahan 2024).

Training and Alignment Processes in Large Language Models

The capacities of large language models are dictated, not just by their architectures, but also by how theyre trained and how aligned they are with human priors. In the modern literature, one often treats the LLM development pipeline as proceeding through a series of complementary stages-- pre-training and fine-tuning, for example.

These stages altogether determine the extent of linguistic competence, task generality and reliable behavior a model can produce (Naveed et al., 2025).

Pre-training: Learning Linguistic Representations

"Pre-training" is the process by which large language models are instilled with the building blocks of language. At this step, models are trained on massive, often unlabelled text data using self-supervision. The most general task is language modeling in which previous context (or tokens) are used to predict the next token (Naveed et al., 2025; Xiao & Zhu, 2025; Zhao et al., 2023).

This objective incentivizes the model to capture linguistics structure, semantics relationships, and context-dependent language usage at large scale. The diversity and size of the pre-training corpus determines the model's extent of general knowledge. Moreover, biases, wrong patterns and representational imbalances existing in the training data can be enacted by the model while outputting, which may pose potential risks to the reliability and fairness of generated outputs (Zhao et al., 2023).

The empirical results suggest that pre-training is important for LLMs, but not enough. Pre-trained models may be able to produce coherent text, but they might lack the ability to still engage in task-oriented conversations, comprehend instructions properly and act safely and predictably (Naveed et al., 2025).

Fine-tuning and Instruction Tuning

Finally, fine-tuning is conducted on top of pre- training to customize the model for certain tasks or scenarios. The second stage of this two-stage approach usually uses a smaller (but better quality) labeled datasets. Although fine-tuning can lead to better performance on specific tasks, it also has negative effects on involving risks such as overfitting, domain dependance and lower cross-task generalization. This stage is a step toward unlocking LLMs from being pipelines that purely "produce language" to interactive tools that are increasingly open to and thus driven by user intention. However, the efficacy of instruction tuning is highly dependent on the quality, diversity and representational coverage of base datasets (Naveed et al., 2025; Xiao & Zhu, 2025; Zhao et al., 2023).

Fine-Tuning for Instruction Following Fine-tuning as a way of conditioning large models to better follow natural-language instructions has been popularized in recent years. In this manner, the model is trained over datasets of question–answer pairs in addition to task descriptions and sample outputs such that more coherent and contextually sound responses are embodied when responding to user inputs. Therefore, instruction tuning is now one of the key work for deploying LLM's as general-purpose assistants (Zhao et al., 2023).

Alignment and Learning from Human Feedback

Alignment is one of the most fundamental methodologies for ensuring that large language models behave in a safe, ethical, and predictable way. Alignment is the attempt to align model outputs with human values, expectations, and societal norms. One popular approach in this direction is Reinforcement Learning from Human Feedback (RLHF) (Chang et al., 2024; Naveed et al., 2025; Xiao & Zhu, 2025).

In RLHF, the outputs of the model are evaluated by human annotators; such evaluations are employed to train a reward model, which leads the optimization of behavior performed by the primary model. While RLHF may help mitigate harmful, misinformation and inappropriate content, it has limitations of its own: Scale, consistency and the subjectivity of human judgment (Chang et al., 2024).

More recent work also hints at different strategies for aligning, in the forms of preference learning, constitutional AI and automated feedback. This stream of research also provides significant indications on how such human-centered AI principles can be transformed into actual technical choices (Naveed et al., 2025).

Limits of Training and Alignment

Though decent and significantly enhancements can be obtained by training and alignment, not one of these processes ensure the absolute reliability. Hallucination (i.e., generation of bogus or incorrect facts) and out-of-context generalization, limited explainability etc. are some of the bottleneck limitations faced by LLMs till date (Chang et al., 2024).

This fact serves to remind us that LLMs are not "systems that 'think like humans,'" but rather tools '" whose use can be beneficial within limits and must therefore be carefully designed, deployed, and overseen. For the responsible and effective application of these technologies, it is essential to understand training and alignment processes (Chang et al., 2024; Naveed et al., 2025; Shanahan, 2024).

The processes outlined in this section also serve to inform why timely engineering and LLM-based system design—covered in the following section—are not just technical questions, but rather are ethical considerations and context-relative operability (Liu et al., 2023; Xi et al., 2023).

Prompt Engineering and LLM-Based System Design

As large language models (LLMs) become prevalent, prompt engineering has become a critical aspect in the development of LLM-based systems which controls how users feed information to the model and use its outputs. The training and alignment procedures described earlier in this paper set up a model's generic behaviorial architecture, and prompt engineering specifies how that architecture is realized within context-specific interactions over time. In this regard, prompt engineering offers a versatile means to direct model behavior without the need of modifying internal model parameters (Liu et al., 2023; Xi et al., 2023).

Prompt Engineering: A Conceptual Framework

Hacking prompt engineering can be thought of as an exercise in eliciting desired outputs from large language models through the careful construction of inputs. This goes beyond the asking of a good question into structuring of context, participant roles, examples and output formats. In the literature, prompt engineering is described as an interim level of abstraction that allows LLMs to move on a scale from general purpose language models toward high performing task specific systems (Xi et al., 2023).

The rise of this paradigm is directly related to the ability for large language models to learn in-context. LLMs can transfer to new tasks without modifying their model parameters, given task examples and explanations. Prompt engineering makes this ability a systematic and repeatable design strategy (Zhou et al., 2022; Zhao et al., 2023).

Prompt Types and Design Strategies

Zero-shot, few-shot and chain-of-thought (CoT) prompting are the generic methods for prompt engineering that have been widely used in literature (Marvin et al., 2023). Zero-shot prompting uses only task descriptions, with no prompts provided, while few-shot prompting aims to improve output quality by providing a few examples. By contrast, chain of thought prompting encourages the model to generate intermediate reasoning steps explicitly, and leads to more coherent and interpretable outputs especially for multi-step problem solving problems (Naveed et al., 2025).

The success of these methods relies on both the semantics carried by but also the structure of the prompt. Role-based descriptions (e.g., “act as an expert”), contextual cues, and explicit output requests help mitigate model uncertainty and promote predictable behavior. Also, previous studies emphasized that prompt engineering is brittle and small linguistic changes can result in a different behavior for the model (Liu et al., 2023).

The Relationship between Prompt Engineering and System Design

Such expeditious engineering should not be treated as only one single (user) interaction practice, but as a formal part of the design of LLM-based systems. Modern LLM applications are commonly composed using layered architecture, where the prompts serve as control and orchestration layer between user inputs and model outputs. In this sense, prompts serve a function similar to interface definitions or business rules in more traditional (i.e., softwarebased) systems (Liu et al., 2023; Xi et al., 2023).

At the professional system level, prompts are frequently combined with up-to-date context information, multi-step task performance and use feedback. In the case of agent-based LLN systems, prompts are not only inputs that a user provides but may also be intermediary instruction produced by the model. This transforms prompt engineering from being a static command-writing exercise to an interactive and adaptive system component (Naveed et al., 2025; Xi et al., 2023).

The Relationship between Prompt Engineering and Fine-Tuning

Mechanical tuning and quick fabrication serve as dual methodologies to sculpt LLM behavior, featuring different cost, time and flexibility tradeoffs. - Fine-tuning: Model parameters are modified when fine-tuning, leading to hightime and data overhead while in prompt engineering, little or no changes in model internals are made because the adaptation is quicker of low cost (Naveed et al.; 2025).

Therefore, rapid engineering is preferable for prototypes, developing new applications quickly, and customized user functionality. However, the literature notes that in high-risk or safety-critical domains, simple prompt-only systems could be insufficient and both fine-tuning and alignment-based proceedings should be combined with prompt-based approaches (Chang et al., 2024).

Limitations and Design Principles

However, prompt engineering has a number of limitations to its flexibility and strength. High prompt sensitivity to contextual variation could lead to unpredictability — and even the instability — of system behavior, rendering LLM-based applications brittle to linguistic drift (Liu et al., 2023; Xi et al., 2023). Furthermore, for complex tasks, longer and more specific prompts do not necessarily achieve an proportional improvement in the performance of model and may even add more sources of ambiguity or cognitive overload into reasoning process inside BERT (Naveed et al., 2025; Chang et al., 2024).

The ideas conceptualized in this part illustrate why the multimodal and interdisciplinary uses of LLMs described in the following section rely not just on model size but also on thoughtful design of interaction and system-level decisions (Xi et al., 2023; Xiao & Zhu, 2025).

Because of this, the literature calls for prompt engineering practices to be guided by principles that emphasize clarity (of the prompt), consistency, and reproducibility. The type of systematic generation, testing and version management protocols for prompts appear to be increasingly seen as essential in order that LLM-based applications are sustainable and resilient (Liu et al., 2023; Xi et al., 2023).

Interdisciplinary Applications: Healthcare, Education, and Human–Machine Interaction

The effects of generative artificial intelligence and large language models (LLMs) are not solely confined to computer science, with parallels observed in human-centered domains such as healthcare, education, and human–machine interaction (Naveed et al., 2025; Shanahan, 2024). These domains provide natural environments for LLM-based applications because they are characterized by a high density of textual and informational content, along with the crucial importance of decision-supporting processes (Teubner et al., 2023; Xi et al., 2023). Moreover, applications active in these domains require not just efficacy but trustworthiness, ethical considerations and enduring human supervision (Chang et al., 2024; Shanahan, 2024).

Large Language Models in Healthcare

Health is an area where significant amount of textual data (clinical notes, patient records, scientific publications and clinical guidelines) are produced. Recent years have seen increasing interest in large language models as auxiliary aids for summarizing, structuring and making sense of these data. In the literature, LLMs are being used increasingly for tasks from automated summarization of clinical documentation, aiding in patient–physician communication to speeding up medical article review (Naveed et al 2025).

From the educational and clinical DSS perspectives, results of studies suggest that LLMs can help improve access to information and provide explanatory material. However, in medical applications, false positives such as hallucination, obsolete information and context-insensitive generalisation are critical risk factors. As a result, the literature has highlighted that LLMs should not be framed as independent decision-making systems but rather support and enhance human expertise (Chang et al., 2024).

Furthermore, health AIs are deeply intertwined with ethical and regulatory issues such as data privacy, patient safety and legal responsibility (Chang et al., 2024; Naveed et al., 2025). These limitations highlight the requirement that specific domain and oversight considerations be addressed in the development and deployment of LMM-based systems (Shanahan, 2024; Teubner et al., 2023).

Generative Artificial Intelligence and Large Language Models in Education

The field of education is one of the most rapidly expanding areas for generative AI pervasion. Most of current educational scenarios make use of large language models, spanning from the production of instructional material, to the personalization of learning content, automated feedback delivery and assessment support. Particularly with regard to the reduction of teacher load and being better able to adapt to students' individual learning requirements, these applications open up great potential (Teubner et al., 2023).

However, debates about the role of LLMs in education go beyond pedagogical advantages. The problems of academic integrity, originality, and the trustworthiness of an educational assessment process are still among the hottest topics in generative AI in education. Recent research report that the LLMs could be supportive of learning in a positive way, as guidance and feedback tools, whereas an unregulated or inappropriate use is likely to have negative effects on learners (Shanahan, 2024).

Consequently, the literature has, increasingly purported that staying clear of generative AI in education is both not feasible and not realistic. Rather, it is considered more feasible to incorporate LLMs in accordance with pedagogic goals. Models that foster students' utilization of LLMs as cognitive tools for critical thinking, problem solving, and self-regulated learning are the cornerstones of AI-supported educational changes (Daugherty & Wilson, 2024).

Human–Machine Interaction and Cognitive Systems

The advent of large language models has also brought a major paradigm shift in the study of human-computer interaction (HCI). Conventional graphical interfaces are being supplemented—or replaced—by natural language–based interaction modes, in which users can interact with sophisticated digital systems using conversational language. This shift sees LLMs both as knowledge-generating means and interactive interfaces (Xi et al., 2023).

HCI research shows that LLMs can be used to poorly model users' intentions and generate contextually relevant responses. But the cognitive consequences of such interactions themselves are far from understood. To the extent that users anthropomorphize LLMs, they are also likely to defer unnecessarily frequently to these systems and neglect their critical faculties (Shanahan 2024).

Nascent work in, for example, language-based brain–computer interfaces (BCI) and cognitive support systems also indicates that large language models could become part of systems that interface more closely with human cognition. These developments suggest that LLMs could be at the heart of human–machine cooperation as we know it in the future (Naveed et al., 2025).

Shared Constraints and Challenges of Interdisciplinary Use

One common challenge of generative AI and large language models becomes clear when compared across application domains, from healthcare2–4 to education5–7 and human-machine interaction: the system must be both safe and beneficial. These range from the problems of output verification, contextuality, ethics, and user perception (Chang et al., 2024; Naveed et al., 2025; Shanahan, 2024). The responsible and effective use of LLMs in these domains, however, demands more than just technical knowledge — it also requires domain expertise from humans (who play a vital role to guide the generative process), human oversight and ethical considerations that inform system design (Teubner et al., 2023; Xiao & Zhu, 2025).

The interdisciplinary cases that we present here are demonstrations of the sort of changes brought about by large language models for good on one hand, while underscoring how such potentials can be actualized only through thoughtful, intelligent and context-sensitive strategies of deployment (Naveed et 37 al., 2025; Shanahan, 2024). The agent based systems and domain-specific applications, which are discussed in the next section, are more sophisticated forms of this cross-fertilisation (Xi et al., 2023; Xiao & Zhu, 2025).

Agent-Based LLM Systems and Domain-Specific Applications

In recent years, large language models have demonstrated the ability to generalize beyond text generation for goal-oriented tasks and even multi-step reasoning and interaction with external environments. This growth of usage has lead to the rethinking of LLMs in terms of the agent paradigm. Agent-based LLM systems treat a language

model as an agent that takes the environment as input and outputs actions to influence it, repeatedly adjusting its behavior to achieve certain goals (Naveed et al., 2025; Xiao & Zhu, 2025).

Such a view allows LLMs to be seen as more than reactive answer parsers, but as cognitive middle-ware agents capable of deconstructing complex tasks and performing them through interaction with external abilities/tools/resources (Xi et al., 2023).

Conceptual Framework of Agent-Based LLM Systems

Classical concepts of agents within artificial intelligence writings are employed by agent-based LLM systems. From this perspective, the agent is a system itself that perceives its environment, decides upon what it perceives and modifies behavior in order to achieve specific goals. Integrated with large language models, agent architectures enable sufficiently established models to become the central means of representation, reasoning and control for these processes (Xiao & Zhu, 2025).

Modern literature characterise the agents that rely on LLM as layered architectures including task planning, subtask generation, tool usage and result evaluation components. In these architectures, the LLM is a control core that orchestrates natural language reasoning as well as system component interactions (Xi et al., 2025; Naveed et al., 2025).

In this environment agent-oriented systems must be engineered quickly to rise above static command generation and into dynamic and reflective interactional loops (Liu et al., 2023).

Single-Agent and Multi-Agent Approaches

Architecture of Agent-Based LLM systems can be structured as single-agent and multi-agent approaches. In single-agent settings, a single LLM takes care of all the planning and execution for a specific goal. Any such method usually works well for relatively simple problems, but often has difficulty dealing with complex and multivariate problems (Xiao & Zhu, 2025).

In multi-agent systems, several LLM agent-based cooperate on separate roles or domains of knowledge. It allows task decomposition, parallel reasoning and mutual verification among agents. The literature shows that in such systems more flexibility and robustness can also be achieved especially in settings that require planning, negotiation or decision support (Naveed et al., 2025).

But multi-agent systems also bring about new challenges, such as the complexity of coordination and consistency problem, along with more computational overhead. Therefore, the agent architecture design should not be based on model capability but also on system engineering and governance (Xiao & Zhu, 2025).

Domain-Specific Applications: Finance, Industry, and Enterprise Systems

Domain specific applications are however, a particularly interesting niche for agent-based LLM systems. In finance, existing systems are used as assistive agents helping the human experts with tasks such as risk analysis, interpretation of reports and generation of scenarios. In this class of applications, LLM agents examine data sources for information, form a synthesis and propose recommendations subject to constraints (Teubner et al., 2023; Naveed et al., 2025).

At the industrial and enterprise levels, agent-bases systems are being increasingly used for process automation, decision support, and knowledge management. Agents enabled by LLMs can reason upon company documents and user requests and act as mediators between disparate information systems (Teubner et al., 2023; Xi et al., 2023).

One of the advantages with LLM agents in these domains is that they enable automation of complicated, text-dependent business rules while maintaining human-in-the-loop and ultimate decision-making control (Xiao & Zhu, 2025).

Limitations and Risks of Agent-Based Systems

Although agent-based LLM systems exhibit flexibility and can be automated to a high degree, there are some serious concerns about reliability and predictability. Short of well-designed, agents can easily lead to a chain-reaction issues under high-stake contexts (e.g., finance, healthcare and public administration) by planning badly due to wrong information source or bias (Chang et al., 2024; Xiao & Zhu, 2025).

Further, higher degrees of agent autonomy aggravate concerns with responsibility and accountability. It is thus suggested in literature that agent-based LLM systems should be developed in the presence of human oversight, controlled authority and transparent control mechanisms (Shanahan, 2024; Naveed et al., 2025).

The agent-based perspectives considered in this section provide critical new perspective towards how large language models might be used in the future, but effective and responsible deployment of these systems needs an integrated treatment of technical, ethical, and governance issues (Chang et al., 2024; Naveed et al., 2025; Xiao & Zhu, 2025). Related and connected matters are discussed further in the next section: bonanzas, sustainability, efficiency and computational expense (Naveed et al. 2025; Teubner et al., 2023).

Efficiency, Sustainability, and Computational Costs

The high capabilities and flexibility of the large language models have a steep computational cost. It is worth to note that the explosion of LLMs number of parameters in recent years has ceased being a technical problem and turned into a multi-faceted one with direct economic, environmental and social consequences. As a result, the

ongoing literature has increasingly focused on development and deployment of large language models from efficiency and sustainability perspectives as well (Naveed et al., 2025).

Computational Costs of Large Language Models

These models relies on cross-product keys and values for its attention mechanisms The training and inference of LLMs are very computational intensive as well. Training models with billions or even trillions of parameters requires dedicated hardware like GPUs or TPUs, long training time and high energy cost. Hence, building such models remains mainly within the reach of big tech companies and represents a high-barrier entry point for academia and small-scale research groups (Zhao et al., 2023).

Costs may not be just on training. The inference process of LLMs is also highly resource-consuming, especially when it comes to the long context windows and multi-step reasoning. Literature in that regard, seems to induce skepticism of the degree that the ever narrower margins on performance are still worth the economy regarding them against higher computational costs (Chang et al., 2024).

Efficient Large Language Models: Technical Approaches

Growing computational expense has driven the studies of efficiency-focused methods in LLM. Commonly used methodologies are model compression, quantization, pruning, and knowledge distillation that attempt to scale down the size and computational requirements of ML models with least performance drop (Liu et al., 2024; Naveed et al., 2025).

Besides, parameter-efficient fine-tuning and task-specific adaptation methods allow for repurposing large models to perform on different tasks withno retraining of the model from scratch (Naveed et al., 2025; Zhao et al., 2023). These strategies are especially helpful in low resource settings, increasing the feasibility of LLMs across a wider variety of settings (Teubner et al., 2023; Chang et al., 2026).

Small, task-specific models have additionally been shown to reach similar or even higher performance than large general purpose models on individual tasks. This finding justifies the belief "bigger is not always better". We believe that goal-specific model design has particular importance (Zhao et al., 2023).

Environmental Impacts and Sustainability Debates

The power consumption of big language models is at the heart of a debate about the environmental effects of artificial intelligence. The carbon emitted during model training and inference is directly related to the energy that supports big data centers. Empirical evidence suggests the sustainability of products such as LLMs, should be seriously considered in their developing and widespread deployment (Delort et al., 2023; Strubell et al., 2019).

In this backdrop, the Green AI paradigm broadens the evaluation metrics beyond model accuracy to also consider energy efficiency and its environmental costs. This view encourages us as researchers and practitioners to focus on solutions that provide high societal benefit with minimal compute, taking their value-cost ratio into account (which is a significant change of perspective towards responsible inclusion and usage of new LLMs) (Strubell et al., 2019).

Economic and Societal Dimensions

There are economic and social implications, not just technical issues with computational cost and durability. The high costs of building large language models could centralise technological power in a handful of institutions and countries, compounding digital inequalities. This problem is commonly discussed in the literature under the rubric of "compute divide", and represents an important threat for inclusiveness for the AI ecosystem (Teubner et al., 2023).

In this context, sustainable and performant LLM approaches will be necessary not just to satisfy ecological concern but also to make sure accessibility and fairness in terms of the scientific pluralism. Both academic and industrial parties are more and more urged to take integrated approaches into account that compare cost and impact in addition to performance point of views (Naveed et al., 2025; Strubell et al., 2019; Teubner et al., 2023).

These discussions of efficiency and sustainability in this section off er the technical underpinning for next chapter on ethics, inequality, and social issues. In the end, the future of big language models will not be a matter only of their force but also depend substantially on how responsibly and sustainably they are built and put to use (Chang et al., 2024; Shanahan, 2024).

Ethics, Inequality, and Societal Impacts

The rapid proliferation of generative AI and large language models has everything to do with ethics, responsibilities, and social ramifications alongside technological breakthroughs. As LLMs become more influential in generating knowledge, aiding decision-making, and automating processes, their accountability towards principles of fairness, reliability and justice needs to be reckoned with. Current literature highlights the fact that societal ramifications of generative AI are due to technical decisions, data sources, usage contexts and governance structures (Sabherwal & Grover, 2024; Shanahan, 2024; Naveed et al., 2025).

Ethical Challenges and Reliability

Among the most frequently discussed set of ethical issues surrounding large language models is the risk associated with hallucination, where a model spits out information that sounds good and respectably linguistically plausible, but which we cannot safely assume to be true. Outputs like this are problematic in critical applications, such as medicine, law and education. In this regard, we would like to highlight that the literature emphasizes that LLM

predictions should be not interpreted as a confirmed knowledge but probabilistic inferences from the available training data (Chang et al., 2024).

A further important ethical consideration is that of biasedness in data-driven systems. Big language models are able to regurgitate — and at times even magnify — historical, cultural and societal biases buried in the training data. This may lead to biased predictions from sensitive attributes like gender, ethnicity, and socio-economic status. As such the ethical AI discussion should move beyond the model output to same light as a critic on broader data ecosystems which contribute to shape them (Gallegos et al., 2024; Li et al., 2024).

Digital Inequality and the Compute Divide

Creating and operationalizing generative AI technology is computationally expensive and capital intensive. This has resulted in a centralized development of LLM capacities at few companies and countries, further possibly increasing digital inequalities worldwide. In the literature, this is often described as the "compute divide", and is considered by many as one of the biggest threats to inclusivity in AI ecosystem (Teubner et al., 2023).

In addition, access to LLM-based resources is dependent not only on technological infrastructure but also on linguistic, cultural and educational background (Shanahan 2024; Teubner et al. 2023). Establishing an inhomogeneous AI is likely to systematically favor or disadvantage certain segments of users -thus resulting in a reduced realization of the wider societal value potential from generative AI (Gallegos et al., 2024; Naveed et al., 2025).

Education, Work, and Societal Transformation

The societal implications of large language models are particularly acute in education and work. In the educational context, LLMs bring important prospects in support of learning and teaching, at the same time that they lead concerns about academic integrity, assessment validity, and student motivation. Current thinking is increasingly to reject outright bans and promote instead the cautious, informed use of generative AI in education (Shanahan, 2024).

In work, generative AI systems are speeding automation and the restructuring of occupations across a wide array of professions. But that will not only turn roles into something else, or elsewhere, but also generate new demands and job profiles. The work effects of GenAI are not one-way street, and these technologies can augment human–machine partnership with the right policy supports and learning strategies (Daugherty & Wilson, 2024; Eloundou et al., 2023).

Critical Perspectives and Responsible AI

These critical AI perspectives redirect focus away from what generative AI systems do and towards for whom, under what conditions, and with what effects they work. From this perspective, LLMs are not politically neutral technologies but systems moulded by the power relations, economic imperatives and social discourses of their context. As a result, responsible AI debates need to involve not only technical countermeasures, but also governance structures, transparency and accountability systems (Bender et al., 2021). The ethical, inequity and societal issues covered in this section indicate that the future of large language models does not only rest on advancing technically, but also increasingly depends on how these technologies are aligned with societal values and norms. The last part of this chapter brings together these discussions by presenting future outlook and research directions, thus rounding the chapter off as a unified whole (Weidinger et al., 2021).

Future Perspectives and Conclusion

Generative AI and LLMs do not simply emerge from technological advancements; they are the latest realization of decades-long theoretical and practical progress in language, cognition, computation (Tuomela, 1972; Naveed et al., 2025; Xiao & Zhu, 2025). The conceptual frameworks, architectural innovations, education and alignment practices, inter- disciplinary uses and ethical debates interrogated in this chapter signal that the shift brought about by LLMs is not one-dimensional : instead these represent a multi-levelled transformation (Chang et al., 2024; Shanahan 2024).

Emerging Trends in the Future of Large Language Models

Recent work suggests that the future of large language models is more than the pursuit for larger and stronger systems alone. Instead, research lines are directed towards efficiency, long-context processing, sensor merging and agent-based systems (Naveed et al., 2025; Xiao & Zhu, 2025). Notably, models effective with extended context windows have opened new prospects for complicated document understanding, multi-step reasoning and multi-discipline knowledge exchange (Naveed et al., 2025).

Concurrently, multi-modal large language models (MLLMs) that go beyond textual-cen-tric architectures are paving the way to more holistic AI systems jointly processing linguistic, visual and auditory information (Naveed et al., 2025; Zhaoet et al., 2023). This trend situates LLMs as not just text generators but more general pattern learners and representatives (Shanahan, 2024; Xiao & Zhu, 2025).

Human–Machine Collaboration and Agentic Architectures

Another frequently addressed topic in futurological debates is the changing nature of human–machine interactions. Agent-based LLM systems allow model plans with specific goals and interact with their environment, raising again the question of AI autonomy versus human control. The literature states that rather than developing

fully autonomous systems, human-centered agent architectures who works in conjunction with human experts provide a sustainable and secure way forward (Shanahan 2024).

In this regard, the next stage for LLMs is unlikely to be the substitution of human cognitive functions, but as in assisting applications where they are a counterpart: assistance with decision making; synthesizing knowledge and generating ideas (Naveed et al., 2025; Shanahan, 2024; Daugherty & Wilson, 2024).

Ethics, Governance, and Responsible Development

The future of generative AI depends as much on ethical and governance frameworks as it does technical developments. Issues such as hallucination, bias, environmental footprint and digital inequality can't be resolved by technical fixes alone. Normative literature has emphasised that the issue of responsible AI should be informed by principles of transparency, accountability and social utility (Chang et al., 2024; Teubner et al., 2023).

We will therefore see the emergence of future research agendas that go beyond narrow performance measures, taking a holistic approach in which assessing impact, analyzing ethical risks and forecasting sustainability become central dimensions of evaluation and development of AI (Chang et al., 2024; Shanahan, 2024; Strubell et al., 2019).

Overall Assessment and Conclusion

This chapter has attempted to bring a comprehensive view on generative AI and large language models, grounding technical advancements in their wider interdisciplinary and social context. Looking across our reviewed articles, the studies indicate that LLMs are not simply "human-like intelligence" or "statistical tools", but they possess meaning and impact depending on their diverse use cases and design decisions *(Naveed et al. 2025; Shanahan 2024; Xiao & Zhu 2025).

In short, the future of large language models will depend on much more than simply building bigger systems but will also require further combining advances across a multitude of disciplines toward intelligent, efficient and responsible AI. Seen through the lens of interdisciplinarity, generative AI holds great promise for making meaningful and sustainable contributions across scientific research, education, health care, work and beyond. But this can only be achieved if the AI ecosystem advances in both technical mastery and awareness of ethical issues (Chang et al., 2024; Strubell et al., 2019; Teubner et al., 2023).

Disclosure of AI usage

During the preparation of this work, the author(s) used [OpenAi-ChatGPT and Google Gemini] in order to [e.g., generate text, check grammar, analyze data, or assist with literature search]. After using this tool/service, the author reviewed and edited the content as needed and take full responsibility for the content of the published work.

References

Baldassarre, M. T., Caivano, D., Fernandez Nieto, B., Gigante, D., & Ragone, A. (2023, September). The social impact of generative ai: An analysis on chatgpt. *In Proceedings of the 2023 ACM Conference on Information Technology for Social Goo*d (pp. 363-373).

Bender, E. M., & Koller, A. (2020, July). *Climbing towards NLU: On meaning, form, and understanding in the age of data*. In Proceedings of the 58th annual meeting of the association for computational linguistics (pp. 5185-5198).

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). *On the dangers of stochastic parrots: Can language models be too big?*. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623).

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., ... & Xie, X. (2024). A survey on evaluation of large language models. ACM transactions on intelligent systems and technology, 15(3), 1-45. https://dl.acm.org/doi/10.1145/3641289

Daugherty, P. R. & Wilson, H. J., (2024). *Human + machine: Reimagining work in the age of AI*. Harvard Business School Press.

Delort, E., Riou, L., & Srivastava, A. (2023). *Environmental Impact of Artificial Intelligence* (Doctoral dissertation, INRIA; CEA Leti).

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130, 10*.

Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., ... & Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics, 50*(3), 1097-1179.

Li, Y., Du, M., Song, R., Wang, X., & Wang, Y. (2024). A survey on fairness in large language models. *Procedia Computer Science, 00*, 1–28. https://doi.org/10.48550/arXiv.2308.10149

Liu, A., Feng, B., Wang, B., Wang, B., Liu, B., Zhao, C., ... & Xu, Z. (2024). *Deepseek-v2: A strong, economical, and efficient mixture-of-experts language model*. arXiv preprint arXiv:2405.04434.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM computing surveys, 55*(9), 1-35.

Marvin, G., Hellen, N., Jjingo, D., & Nakatumba-Nabende, J. (2023, June). Prompt engineering in large language models. *In International conference on data intelligence and cognitive informatics* (pp. 387-402). Singapore: Springer Nature Singapore.

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., ... & Mian, A. (2025). A comprehensive overview of large language models. *ACM Transactions on Intelligent Systems and Technology, 16*(5), 1-72.

Sabherwal, R., & Grover, V. (2024). The societal impacts of generative artificial intelligence: A balanced perspective. *Journal of the association for information systems, 25*(1), 13-22.

Shanahan, M. (2024). Talking about large language models. *Communications of the ACM, 67*(2), 44–49. https://dl.acm.org/doi/pdf/10.1145/3624724

Storey, V. C., Yue, W. T., Zhao, J. L., & Lukyanenko, R. (2025). Generative artificial intelligence: Evolving technology, growing societal impact, and opportunities for information systems research. *Information Systems Frontiers*, 1-22.

Strubell, E., Ganesh, A., & McCallum, A. (2019, July). *Energy and policy considerations for deep learning in NLP*. In Proceedings of the 57th annual meeting of the association for computational linguistics (pp. 3645-3650).

Teubner, T., Flath, C. M., Weinhardt, C., Van Der Aalst, W., & Hinz, O. (2023). Welcome to the era of chatgpt et al. the prospects of large language models. *Business & Information Systems Engineering, 65*(2), 95-101.

Tuomela, R. (1972). Model theory and empirical interpretation of scientific theories. *Synthese*, 165-175.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems, 30*.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., ... & Gabriel, I. (2021). Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359*.

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., ... & Gui, T. (2025). The rise and potential of large language model based agents: A survey. *Science China Information Sciences, 68*(2), 121101.

Xiao, T., & Zhu, J. (2025). *Foundations of large language models. NLP Lab*, Northeastern University & NiuTrans Research. https://github.com/NiuTrans/NLPBook

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, Z., Yang, Z., & Wen, J. R. (2023). A survey of large language models. *arXiv preprint arXiv*:2303.18223. *1*(2). https://arxiv.org/abs/2303.18223

Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., & Ba, J. (2022, November). Large language models are human-level prompt engineers. *In The eleventh international conference on learning representations*.

Author Information

Önder Yıldırım

https://orcid.org/0000-0003-4333-2323

Atatürk University

Oltu, Erzurum

Türkiye

Contact e-mail: *ondery@atauni.edu.tr*

Citation

Yıldırım, Ö. (2025). Generative Artificial Intelligence (GenAI) and Large Language Models. In A. Kaban, & T. Demirel (Eds.), *Interdisciplinary Applications of Artificial Intelligence - 1* (pp. 18-39). ISTES.

Chapter 3- Data Ethics, Privacy, and Secure Artificial Intelligence

Hasan Küçükoğlu

Chapter Highlights

- Data ethics forms the fundamental framework for AI systems, ensuring ethical data collection and processing by integrating individual rights, social norms, and legal procedures.
- The concept of privacy is being redefined to address the challenges of personal data collection and processing, requiring new ethical boundaries for AI-based systems.
- To ensure processes run smoothly and effectively, AI-integrated applications must clarify risk dimensions, monitor accountability, and make trust-building design transparent.
- Defining the ethical and legal boundaries of developed AI systems is critical to ensuring their development is compatible with both individual rights and societal benefits.
- AI technologies should be designed and implemented with a strong focus on both ethical and legal responsibilities to increase societal benefit and ensure sustainable integration.

Introduction

The rapid developments in artificial intelligence (AI) technologies and applications raise new ethical questions regarding personal information security and data processing. This transformation is not limited to technical and functional advances but is also shaped by social and ethical responsibilities, creating not only a legal obligation but also a moral obligation for the effective implementation of AI systems. Therefore, the effective and secure use of AI requires a comprehensive and integrated approach that ensures the correct processing of data, respect for user privacy, and system security from the outset (Alkali et al., 2022).

In this context, developments in information and communication technologies are increasing the adoption and usage rates of AI systems by protecting individuals' privacy while gaining the trust of societies. Although the technologies used as the basis for AI are result-oriented and effective in creating individual and societal benefits, the problems they cause in the background can lead to greater harm. This situation emphasizes the importance of focusing on the process as well as the result in the development of AI technologies (Brasioli et al., 2025).

Data ethics is a crucial element shaping the foundations of AI, requiring that not only legal requirements but also social norms and individual rights be taken into account in the process of data collection, processing, and distribution. Fundamental principles such as transparency, fairness, and consent, which must be considered throughout these processes, form the theoretical foundations for the ethical management of AI systems. Similarly, the fair management of the effects of AI systems on users will be possible within a framework that considers both individual rights and social benefit. In this context, ensuring data security and protecting privacy is directly related to people's sense of trust as well as the functionality of the technology (Kashefi et al., 2024).

Beyond protecting individuals' privacy, ensuring that AI systems are deployed in decision-making processes in a manner that is fair, transparent, and free from bias is also a critical area. Developing AI algorithms with a focus on impartiality, creating models that consider diversity and equality, will help protect the fundamental rights of every individual by increasing social benefit. At the same time, ensuring that algorithms are controlled and traceable will enable users to know how and why they encounter certain results. Such transparency is not only important for increasing security but also for correcting potential errors and preventing risks and dangers. Therefore, the proper training and supervision of AI are essential factors for strengthening social trust and ensuring widespread acceptance of the system (Abolaji & Akinwande, 2024).

In this regard, when evaluated within a broader concept, the variable structure of AI systems necessitates that ethical and legal regulations also be dynamic. The use of AI should not lag behind laws that cannot keep up with the speed of technology; rather, legal infrastructures should be shaped to adapt to these innovations. Consequently, the development of risk management strategies minimizes the negative effects of technologies while ensuring that their benefits are distributed fairly across all segments of society. Ultimately, it should be recognized that strengthening cooperation between technology, law, ethics, and social sciences is a critical factor in adopting an interdisciplinary approach in this process (Gupta et al., 2025).

The Theoretical Foundations of Data Ethics in AI Systems

The sharp and continuous rise in information and communication technologies is accelerating the development process of AI technologies and raising significant ethical concerns, particularly in the areas of data processing, ensuring personal information security, and developing legal regulations. The successful implementation of AI systems requires not only technical advances but also ethical and societal responsibility. Data must be collected and processed ethically, as this is not just a legal obligation but a moral one. The effective and secure use of AI involves ensuring data is handled correctly, respecting user privacy, and safeguarding system security. Technological advancements in AI not only protect individual privacy but also help gain societal trust, enhancing the acceptance of AI systems within communities (Galiana et al., 2024).

Data ethics plays a crucial role in shaping AI systems by incorporating legal requirements, societal norms, and individual rights. Key principles such as transparency, fairness, and consent are foundational to the ethical management of AI systems, ensuring that the development and application of AI respect human rights. The fair management of AI's societal impact can only be achieved within a framework that takes both individual rights and collective benefits into account. In this context, privacy protection and data security are not only tied to the functionality of the technology but also to the public's trust in AI (Jain et al., 2025).

The design and implementation of AI systems require significant social and ethical responsibility, involving more than just technical considerations. The security of AI systems depends not only on their technical infrastructure but also on the management of potential risks, the implementation of protective measures, and the transparent monitoring of these processes. Furthermore, identifying, developing, and implementing individual control methods during the development phase is an important step in ensuring social trust in general. In order for AI to increase social benefit, it must be addressed not only from a technical perspective, but also from an ethical and legal perspective, and it must be developed in a manner that is compatible with the systems it is used in, ensuring that it serves the public interest (Hermansyah et al., 2023).

As AI continues to evolve, the theoretical foundations of data ethics become more intertwined with normative ethical frameworks. These frameworks focus on ensuring that AI systems adhere to specific norms and values, while also considering the societal impacts these systems create. Furthermore, the ethical responsibility of developers and institutions goes beyond evaluating algorithmic outputs; it includes assessing design choices, data selection, and institutional decision-making. Thus, data ethics emerges as an area deeply connected to human decisions, highlighting the inseparable relationship between AI's technical functioning and its broader societal consequences (Stahl et al., 2023).

Data ethics is fundamentally shaped around three main questions (OECD, 2024):

- How and for what purposes is data collected?
- Whom do these data represent, and who is excluded?
- What impacts do the outputs generated from data have on individuals and society?

The Concept of Data Ethics

These questions necessitate the evaluation of data usage not only in terms of legal compliance but also in terms of justice, equality, and societal impact. As AI systems work with large-scale datasets, they have the potential to create systematic ethical problems rather than isolated individual errors. This calls for a broader and more comprehensive approach to data governance, one that incorporates ethical considerations at every stage of AI development (Alonso & Siracuse, 2023).

This situation shapes the framework of data ethics around four fundamental concepts (Huang et al., 2022):

- **Ethical Issues & Risks of AI:** While AI offers significant benefits, it also presents risks such as algorithmic biases, privacy violations, and lack of transparency, which undermine trust and complicate regulation. These challenges require proactive measures to ensure that AI development aligns with ethical standards and societal values.
- **Methods to Evaluate the Ethicality of AI:** Assessing AI ethics involves not only technical evaluation but also social and legal considerations, with tools like bias detection and transparency audits helping to measure societal impact. Effective evaluation methods are essential to mitigate risks and ensure responsible deployment of AI systems.
- **Guidelines & Principles for Ethical AI:** Ethical AI development must adhere to principles like transparency, fairness, and privacy, guided by international frameworks such as those from the OECD and UNESCO. Following these principles ensures that AI technologies benefit society while respecting fundamental rights and freedoms.
- **Approaches to Solve Ethical Problems in AI:** Addressing AI's ethical issues requires diverse approaches, from deontological models to emotional intelligence and participatory ethics, involving various societal stakeholders. This multi-faceted approach is crucial to developing AI systems that align with human dignity and social justice.

Figure 1. AI Ethics (Huang et al. 2022)

AI Ethics Principles

The reliance of AI technologies on the collection of large amounts of data requires the protection of this data through personal information, legal frameworks, and ethical norms throughout the lifecycle of these technologies. In this context, according to the draft of the UNESCO Recommendations on AI Ethics, the ten key principles of AI ethics are outlined as follows (UNESCO, 2022):

1. **Proportionality and Do No Harm**: The design and implementation of AI systems should be proportional to the goal of minimizing societal harm. This principle emphasizes that AI systems should intervene only when necessary, with potential harms identified in advance. AI applications must be carefully monitored to ensure they do not harm individuals, considering possible side effects.

2. **Safety and Security**: AI must be resilient against security vulnerabilities and external attacks; this requires the continuous provision and improvement of system security. The security of AI systems is not only a technical issue but also a societal responsibility, as data breaches can pose significant security risks. Security measures must be implemented from the design phase and tested at every stage.

3. **Right to Privacy and Data Protection**: AI systems must respect the privacy of individuals in the collection, processing, and storage of personal data. Data protection is not only a legal obligation but also an ethical duty to protect individuals' fundamental rights. Privacy rights in AI applications must align with transparency and user consent, minimizing the risk of data breaches. Advanced data protection mechanisms are critical to ensuring the secure and ethical deployment of AI.

4. **Multi-Stakeholder and Adaptive Governance & Collaboration**: The governance of AI systems requires collaboration not only from technical experts and governments but also from societal actors, businesses, and civil society organizations. This multi-stakeholder approach ensures that the widespread societal impacts of AI are managed fairly. Governance processes must be adaptable to consider the evolution of technology and should be flexible enough to address emerging challenges.

5. **Responsibility and Accountability**: AI systems should operate within the boundaries of responsibility defined by their designers and users. This requires the traceability and auditability of decision-making processes in AI applications, as erroneous outcomes or faulty decisions could lead to societal harm. Every AI application should be held accountable for potential damages, and clear accountability mechanisms should be established.

6. **Transparency and Explainability**: AI systems should be transparent and understandable to users and stakeholders, with clear explanations of how algorithms function and which data are used to make decisions. Transparency fosters trust in AI systems and helps prevent misunderstandings. Explainability includes not only understanding the outcomes of algorithms but also responsibly communicating these results.

7. **Human Oversight and Determination**: Human oversight and control should always be central in the design and application of AI systems, as AI should complement human decision-making, not replace it. Humans must have the ability to review AI-generated results and intervene when necessary.

8. **Sustainability**: AI applications should adhere to environmental sustainability principles and be designed to minimize the ecological footprint of the technology. This requires the efficient management of energy and resources used in AI development. Sustainable AI considers not only environmental impacts but also societal sustainability.

9. **Awareness and Literacy**: It is crucial that societies have adequate knowledge and awareness of AI technologies to ensure their safe and ethical use. AI literacy enables individuals to understand the effects of these technologies. This principle encompasses not only technology users but also policymakers, developers, and the academic community.

10. **Fairness and Non-Discrimination**: AI systems should produce fair outcomes for every individual and community, preventing discrimination based on factors such as race, gender, or age. This principle aims to eliminate systematic biases in algorithmic datasets and offer equal opportunities to all users. Fairness should not only be limited to the use of technology but should also ensure equal access to technology across different segments of society.

Figure 2. UNESCO AI Principles (UNESCO, 2022)

In conclusion, data ethics offers a theoretically profound field of evaluation that is inseparable from the technical achievement of AI systems. Ethical principles should be viewed not as external constraints hindering AI innovation, but as guiding frameworks that ensure the social legitimacy, credibility, and sustainability of these systems. This theoretical approach to data ethics makes the interdisciplinary nature of privacy and secure AI research more visible (Paliszkiewicz & Gołuchowski, 2024).

The Redefinition of Privacy in the Context of AI

The rapid increase in the data processing capacity of AI systems significantly challenges traditional definitions of privacy. The classical concept of privacy is primarily based on the protection of personal data from unauthorized access and the preservation of an individual's private space. However, AI-based systems are capable of generating meaningful inferences about individuals not only from explicit identity information but also from indirect and anonymized data. This situation demonstrates that privacy is no longer solely concerned with "data protection," but also with the limitation of information derived from data (Baker & Caton, 2025).

In the context of AI, privacy is a multifaceted concept that should be addressed not only in terms of individual control over personal data, but also in relation to the context, purpose, and outcomes of the data processing processes. Therefore, privacy is viewed as a dynamic concept that is continuously redefined in parallel with technological advancements, rather than a static right. The complex and often opaque decision-making mechanisms of AI systems make this need for redefinition even more evident (Curzon et al., 2021).

In the event of the leak of privacy, the potential harms of AI can be categorized as follows (NIST, 2023):

Figure 3. Potential Harms Related to AI Systems (NIST, 2023)

- **Harm to People:** AI systems can violate privacy by using personal data without proper consent or transparency. Additionally, biased algorithms can result in unfair treatment, impacting individuals' autonomy and trust in technology.

- **Harm to an Organizations:** The use of AI can lead to inefficiencies and errors in decision-making, harming business operations. Furthermore, AI-driven security risks, such as cyberattacks, can cause significant financial and reputational damage.

- **Harm to an Ecosystems:** AI technologies consume large amounts of energy and contribute to environmental degradation. The extraction of resources for AI development can damage ecosystems and hinder sustainability efforts.

Traditional Privacy vs. Contextual Privacy

The traditional concept of privacy is based on the principles that personal data should not be collected, stored, or shared without explicit consent. This approach centers on the protection of data that directly reveals an individual's identity. However, AI systems are capable of accessing sensitive information through behavioral patterns and statistical relationships, without requiring direct identification of the individual. This situation reveals that privacy breaches may occur not only through data leaks or unauthorized access but also as a result of legitimate data processing practices (Wachter & Mittelstadt, 2019).

Contextual privacy, on the other hand, is an approach that suggests the privacy and meaning of a piece of data may vary depending on the context in which it is collected. In other words, a data point may pose a low privacy risk in one context but a high risk in another. This becomes particularly complex when combined with AI and big data analysis. AI systems can collect large amounts of data and perform contextual analysis of these data sets. As a result, the environment in which the data is collected, the algorithms used, and the design of the systems require a more dynamic approach to privacy (Szostak, 2025).

In this context, privacy is now related not only to who sees the data but also to the context, purpose, and outcomes of data processing. Since AI systems can detach data from its original context and apply it to different use cases, the usage terms initially accepted by individuals may lose their meaning over time. This explains why the classical notion of privacy is insufficient in the age of AI (Jobin et al., 2019).

Data Sovereignty and Individual Autonomy

Data sovereignty refers to the control individuals have over their own data. Traditional privacy concepts often attribute ownership of data to the institutions or governments that collect it. However, with the widespread adoption of AI applications, the control individuals have over their data is being reconsidered. AI systems often collect personal data from individuals while leaving them with minimal control over its usage and sharing. This raises concerns about data sovereignty and individual autonomy (Kaur & Singh, 2025).

For individuals to have control over their data in line with their will, transparency and informed consent are required in data collection, sharing, and processing. In this context, individual autonomy refers to the ability of people to make decisions about their data. AI obtains user consent during data collection, but the extent to which this consent is informed and freely given remains a major question (Hagendorff, 2020).

As AI systems become more complex, they often use personal data in ways that are not transparent to the individuals involved. This lack of clarity reduces people's ability to make informed decisions about their data, undermining their autonomy and sovereignty. To protect data sovereignty, it is essential to implement stronger regulations that ensure individuals are fully informed and have meaningful control over how their data is used by AI systems (Tang, 2023).

Legal Regulations and New Approaches

The impacts of AI systems on privacy necessitate a review of existing legal regulations. Classical personal data protection laws aim to ensure transparency, accountability, and the protection of individual rights in data collection and usage processes, without delving into technical details. However, these regulations struggle to keep pace with the rapidly advancing technologies in AI and big data analytics. Concepts such as data sovereignty, contextual privacy, and meaning generation call for a new legal framework beyond the current regulations (Ravindran & Singh, 2025).

Global legal regulations and frameworks can be listed as follows:

Table 1. Regulations in the Field of Trustworthy AI

Organization	Regulation	Scope
EU	EU Artificial Intelligence Act (European Union, 2023)	Regulatory framework for AI in EU member states, transpa safety, ethics, user rights
USA	AI Governance Principles (NIST, FTC, 2023)	Safe and ethical use of AI technologies, transparency, data security, AI interaction and regulation
Canada	AI Ethics Guidelines (Montreal Declaration, 2023)	Justice, transparency, accountability, security, responsible AI development and governance
UK	AI Regulation White Paper (2023)	Identifying AI risks, transparency, accountability, fairness, ethical principles, regulation of AI technologies
AI HLEG	High-Level Expert Group on AI - Ethical Guidelines and Assessment List (2018)	Trustworthy AI, ethical use, transparency, responsibility, principles for safe use of AI
ISO, IEC	ISO/IEC AI Standards (ISO/IEC 27701, ISO/IEC 23894, 2025)	Privacy in AI systems, reliability, international standards, data security
OECD	OECD AI Guidelines (2019)	Ethical and democratic values, AI policies, global cooperation, safe and fair use of AI
GPAI	Global Partnership on AI (2025)	Responsible development of AI, ethical policies, best practices, global cooperation
UNESCO	Recommendation on AI Ethics (2021)	Ethical dimensions of AI, global ethical norms, societal impact of AI, and ethical regulations
WEF	Guidelines for AI Solutions Procurement by the Private Sector (2023)	Transparent and inclusive AI design, market launch, global cooperation, sector leadership, ethical use of AI

Source: Ağdeniz (2024)

Privacy is a concept that needs to be redefined in the age of AI and big data. This concept is not only related to data access, but also to broader social dynamics such as how data is involved in meaning-making processes and data sovereignty. Considering that traditional personal data protection frameworks fall short in light of the new opportunities offered by AI, new paradigms must be established and more effective solutions developed to minimize the social impacts of technology (Boppiniti, 2023).

Risk, Responsibility and Trust-Building in the Formation of Safe AI

As AI systems grow in complexity and autonomy, security must be understood beyond just technical protections. Safe AI requires a comprehensive approach that not only safeguards systems from external threats but also ensures their predictability, auditability, and societal acceptance. This approach considers AI security as a holistic quality, involving design, governance, and responsible usage practices (Bandyopadhyay et al., 2025).

Additionally, the social and institutional impacts of AI outcomes must also be considered, as these systems profoundly affect people's lives. Therefore, security, risk management, accountability, and trust-building are redefined. Safe artificial intelligence should address AI security from a wide range of social, ethical, and governance challenges and establish a solid foundation for the future by emphasizing the necessity of ethical decision-making and transparent processes throughout the AI lifecycle (Correia et al., 2025).

The Human and Planet dimension represents human rights as well as the broader welfare of society and the planet. AI actors within this dimension form a distinct Artificial Intelligence Risk Management Framework (AIRM) that primarily informs a specific audience. These AI actors may include trade associations, standard-setting organizations, researchers, advocacy groups, environmental organizations, civil society groups, end-users, and potentially affected individuals and communities (Zhang, 2022).

These actors can (NIST, 2023):

- Help ensure the context and understanding of potential and actual impacts;
- Serve as a source of formal or semi-formal norms and guidance for AI risk management;
- Define the boundaries of AI operations (technical, societal, legal, and ethical); and
- Promote discussions on the necessary trade-offs to balance civil liberties and rights, equality, the environment and planet, and societal values and priorities related to the economy.

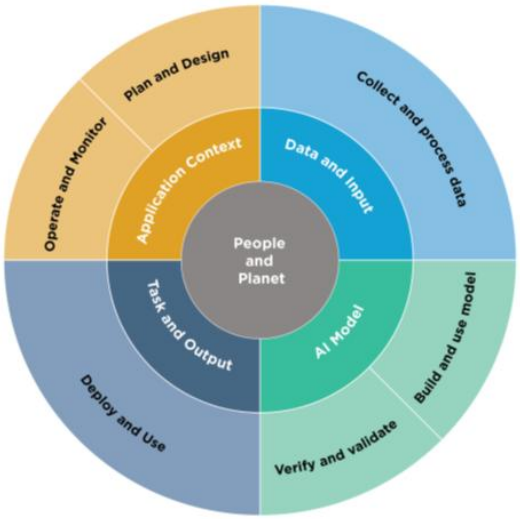


Figure 4. Lifecycle and Key Dimensions of an AI System (OECD, 2024)

The Concept of Risk in AI Systems

In the context of AI, the concept of risk differs significantly from the traditional understanding found in information systems. Classical approaches typically address risks through technical failures, system vulnerabilities, or external attacks; however, in AI systems, risk also encompasses elements such as data quality, model behavior, contextual misalignment, and unforeseen outcomes. This indicates that risk is not limited to probability and impact calculations but is systemic and dynamic in nature (Çela et al., 2025).

The learning capacity of AI systems causes the risk to evolve over time. A system initially deemed safe may exhibit unexpected behaviors due to changes in its usage context or encounters with new types of data. Therefore, risk should not be considered a one-time evaluation but rather an ongoing phenomenon that must be continuously monitored and reassessed throughout the system's lifecycle (Moghadasi et al., 2024).

The categories and levels of risk associated with the development and/or use of AI are classified as follows (EDPS, 2025):

- **Low or No Risk**: AI systems that pose little or no risk, such as spam filters, do not require restrictions, though it is recommended that these systems comply with ethical guidelines. These systems generally have minimal impact on privacy and security, ensuring limited ethical concerns.
- **Limited Risk**: AI systems that generate limited risk, such as technical programs related to functionality, development, and performance, are subject to the application of AI ethical principles outlined in this document. These systems may still raise issues of fairness and transparency, requiring careful monitoring for potential harms.
- **High Risk**: AI systems that create "high risk" to fundamental rights must undergo pre- and post-deployment suitability assessments, with compliance to ethical rules as well as relevant legal requirements being considered. Such systems often involve sensitive personal data and decisions that can significantly impact individuals' lives and freedoms.
- **Unacceptable Risk**: AI systems that pose "unacceptable risk" to human safety, livelihoods, and rights are prohibited. These include systems involved in social profiling, child abuse, or behavioral manipulation, which have severe consequences for personal autonomy and societal well-being.

Risk management must be integrated with artificial intelligence initiatives to ensure that controls are implemented throughout the entire development process, addressing risks such as data, algorithmic, compliance, and operational risks. By incorporating risk management into the artificial intelligence lifecycle, organizations, companies, or startups can proactively address potential issues and ensure reliability. Key subcomponents such as model interpretability, bias detection, and performance monitoring are vital for continuous oversight and transparency. Standards and controls applied from design to post-deployment ensure compliance with ethical and legal requirements. This comprehensive approach promotes responsible AI development and accountability by helping to mitigate risks and increase trust in AI technologies (Curzon et al., 2021).

Transition from Security to Trustworthiness in AI

An important conceptual shift in discussions about safe artificial intelligence is occurring from the concept of "security" to that of "reliability." In this context, the concept of security refers to a system's protection from external interference, while the concept of reliability relates to the system's consistent performance, predictable behavior, and the production of expected results. Within the framework of artificial intelligence systems, these two concepts are inextricably intertwined (Korobenko et al., 2024).

Trustworthiness is not only linked to technical robustness but also to the understandability and auditability of the system's decision-making logic. Users and institutions find it difficult to trust systems whose functioning is not understood or whose outcomes cannot be predicted. Therefore, secure AI extends beyond technical security measures and becomes a field of assessment that also includes human-machine interactions (Wang et al., 2020).

Secure AI should not be merely viewed as a "technical feature," but rather as a "design and governance principle." The design of AI systems should not only ensure that algorithms work correctly but also consider the societal impacts of these systems. A secure AI system, after undergoing technical validation and testing, should be designed in accordance with ethical, social, and legal responsibilities. This approach ensures the secure development and sustainability of AI technologies while protecting individual rights, social welfare, and the public interest through integrated systems (Rosak-Szyrocka & Wolniak, 2025).

Trust-Building Process

The concept of secure AI is directly related not only to technical security measures but also to trust-building mechanisms such as transparency, accountability, and auditing systems. For AI systems to gain societal trust, clear and certain information about how these systems operate must be provided. Transparency allows users and society to comprehend how the system works, what data it processes, and how decisions are made. This transparency is one of the fundamental building blocks of the trust-building process (Olawade et al., 2024).

Additionally, secure AI is tied to the principle of accountability. The decisions made by an AI system must be auditable by both developers and users. If a system makes an incorrect decision, the reasons for that decision must be explained, and the source of the error must be identified. Accountability not only increases the reliability of AI systems but also reinforces the trust placed in these systems (Alakuş & Yılmaz, 2025).

Audit mechanisms are crucial for ensuring secure AI, addressing not only technical issues but also ethical and societal concerns. Effective auditing helps prevent the misuse, abuse, or violation of societal responsibilities by AI systems and should be implemented throughout the entire lifecycle of the technology. Secure AI is closely tied to building societal trust and fulfilling ethical responsibilities, requiring transparency, accountability, and continuous evaluation. This holistic approach ensures individual security while fostering broader societal trust, laying the foundation for the fair and sustainable development of AI (Galiana et al., 2024).

Boundaries, Responsibilities, and Projections

When theoretical discussions on data ethics, privacy, and secure artificial intelligence are considered together, it is clear that artificial intelligence systems cannot be approached solely from a technical advancement perspective. In this context, the conceptual framework outlined supports the process by enabling a comprehensive assessment of the limits within which artificial intelligence applications can be developed, how responsibility should be distributed in these processes, and what future trends will be (Gupta et al., 2025).

The rapid development of AI technologies brings not only technical skills but also ethical, legal, and societal responsibilities. While AI has the capacity to transform human life, it also entails significant ethical and societal risks. Thus, it is necessary to establish specific boundaries both at the technical and societal levels when designing and using AI systems. These boundaries ensure that AI serves the common good while protecting individual rights and promoting social well-being (Farina et al., 2025).

Ethical, Legal, and Societal Boundaries: Legitimacy and Acceptability

The boundaries of AI systems are defined by the balance between technical feasibility and ethical acceptability, ensuring technological progress aligns with societal values. Principles such as justice, non-discrimination, autonomy, and dignity shape these boundaries. While crucial for societal acceptance, they should guide responsible and sustainable development without hindering innovation. Ignoring these boundaries can erode trust, exacerbate inequalities, and undermine public confidence in AI systems (Alonso & Siracuse, 2023).

The OECD has worked on identifying both real and potential risks associated with AI systems, including productive AI in work environments. Some of these risks include (OECD, 2024):

- Large-scale and widespread dissemination of false and misleading information through AI-generated content, especially content that people perceive as real;
- AI models generating convincingly wrong or fictitious responses;
- Increasing harmful bias and discrimination at scale;
- Privacy and data management risks at the level of training data, model development, data and model intersections, or human-AI interaction;
- Challenges of transparency and explainability due to the opacity and complexity of large models;
- Inability to appeal AI model outcomes;
- Breach of privacy through the leakage or inference of personal data.

To manage AI risks, policymakers must create strategies that enhance data transparency, and oversight to reduce bias and discrimination. Strict adherence to privacy standards and secure data management practices are essential to protect users' personal information. Additionally, providing the right to appeal AI outcomes audit processes will foster societal trust and encourage responsible use of AI systems (Darı & Koçyiğit, 2024).

Developers, Institutions, and Policymakers: The Multilayered Nature of Responsibilities

Responsibility in AI systems is not limited to a single actor, but requires contributions from multiple stakeholders, ranging from data providers and developers to user institutions and regulators. This multilayered structure reveals specific areas of responsibility at every stage of the AI lifecycle and underscores that ethical practices must be maintained throughout the process. The ethical evaluation of AI systems should consider not only their outputs but also their design decisions, data selection, and usage contexts. Accountability is a critical principle in this process and requires transparent oversight not only of the results but also of the decision-making and implementation processes (Jain et al., 2025).

Developers are responsible for ensuring that AI systems operate in a secure, fair, and ethical manner, but this responsibility is not merely technical—it also carries societal and ethical dimensions. Institutions and regulatory bodies should share this responsibility, ensuring that AI applications remain within legal and ethical boundaries. Policymakers must develop the necessary laws for the responsible use of AI and raise awareness in society. These shared responsibilities are vital to ensuring that artificial intelligence technologies are implemented in a safe, transparent, and fair manner within society (Safdar et al., 2020).

In a comprehensive framework, the components of the multilayered structure that strategy developers should consider are as follows (Villegas-Ch, & García-Ortiz, 2023):

- **Data Protection Policies and Practices:** To ensure the security of data used in AI systems, robust data protection policies and practices must be developed to prevent unauthorized access and protect data privacy.
- **Data Anonymization Techniques:** Anonymizing or de-identifying data while preserving individual identities is a critical technique to ensure that data usage does not lead to privacy violations.
- **Data Encryption:** Encryption techniques should be applied to secure data during transmission and storage, ensuring that data cannot be accessed or altered by unauthorized individuals.
- **Access Monitoring and Audits:** Continuous monitoring and audit mechanisms should be established to track AI system access and detect unauthorized access attempts.
- **Privacy Assessments:** Regular privacy assessments should be conducted to ensure compliance with applicable privacy regulations during the processing of personal data.
- **Security Assessments:** Comprehensive security assessments should be performed to identify potential security vulnerabilities and threats in AI systems.

These approaches and measures ensure that security and privacy are fully considered within the framework and that clear boundaries are established between the two. Such measures are thought to provide strong and reliable protection during the use process of AI systems. By maintaining these boundaries, AI systems can operate securely without compromising personal data. Ultimately, integrating security and privacy from the outset is critical to the sustainable and ethical deployment of AI technologies (Wachter & Mittelstadt, 2019).

The Future of AI Technologies: Projections and Trends

The future of AI technologies is shaped by both current developments and societal trends. These technologies will continue to transform human life, but the manner in which this transformation occurs will largely depend on ethical and societal responsibilities. The progression of AI technologies will be shaped not only by technical innovations but also by societal values and ethical principles. As AI continues to evolve, its integration into everyday life will require careful consideration of its impacts on individuals and communities. Ultimately, the future of AI must balance innovation with respect for human rights and societal welfare (Brasioli et al., 2025).

AI research is increasingly focused on developing more complex, autonomous systems that can make human-like decisions. This trend will facilitate the broader application of AI across various domains but will also introduce new ethical, security, and privacy challenges. This necessitates questioning not only technological advancement but also how these advancements can be harmonized with societal responsibility. As AI systems gain more autonomy, establishing clear ethical guidelines will be essential to mitigate potential risks. This will ensure that AI development aligns with human values and supports the greater good (Kaur & Singh, 2025).

Some of the primary trends in AI development are (Baker & Caton, 2025):

- **Explainable AI (XAI):** One primary trend is research into explainable AI. Explainability ensures that users understand how and why AI systems make decisions. This will increase the safety and societal acceptance of AI systems. Explainable AI is crucial for auditing and accountability. The transparency of these technologies will help gain societal trust. As AI systems become more complex, the need for explainability will grow, requiring clearer insights into their decision-making.
- **Ethical AI**: A major trend focuses on the development of ethical AI. Ethical AI ensures that AI systems are designed and operate according to societal values, cultures, and ethical norms. This trend is critical for ensuring AI is not only efficient but also fair, equitable, and secure. As AI technologies impact various aspects of life, establishing ethical guidelines will be key to maintaining their alignment with human rights and social justice.
- **Societal Impacts and Sustainability**: The societal impacts and sustainability of AI will become a key part of future AI work. AI's effects on human life will be felt socially, culturally, and environmentally, as well as economically and technically. Therefore, the design of AI systems should offer opportunities for a more just and sustainable society, rather than deepening inequalities. Incorporating sustainability into AI design will ensure that technology contributes to long-term well-being.

Ultimately, the future of AI technologies will be shaped not only by technical innovations but also by the fulfillment of ethical, legal, and social responsibilities. Collaboration between developers, institutions, and policymakers will ensure adaptation to AI-based systems through technical solutions as well as ethical compliance. Therefore, designing future AI applications to be sensitive to social values, individual rights, and environmental sustainability will also be decisive in the development processes of new technologies (Çela et al., 2025).

Acknowledgements or Notes

Disclosure of AI usage

During the preparation of this work, the author(s) used [ChatGPT] in order to generate text, check grammar, analyze data, or assist with literature search. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published work.

References

Abolaji, E. O., & Akinwande, O. T. (2024). AI Powered Privacy Protection: A Survey of Current State and Future Directions. *World Journal of Advanced Research and Reviews*, *23*(3), 2687-2696.

Ağdeniz, Ş. (2024). Güvenilir Yapay Zeka. ve İç Denetim. *Denetişim* (29), 112-126. https://doi.org/10.58348/denetisim.1384391 .

Alakuş, H., & Yılmaz, K. (2025). Bilimsel Araştırmalarda Yapay Zekâ Kullanımı Etik İlkeleri. *MANAS Sosyal Araştırmalar Dergisi*, *14* (Eğitim Bilimleri Ek Sayısı), 193-217.

Alkali, Y., Routray, I., & Whig, P. (2022). Strategy for Reliable, Efficient and Secure IoT Using Artificial Intelligence. *IUP Journal of Computer Sciences*, *16*(2).

Alonso, A., & Siracuse, J. J. (2023). Protecting Patient Safety and Privacy in The Era of Artificial Intelligence. in *Seminars in Vascular Surgery* (Vol. 36, No. 3, Pp. 426-429). WB Saunders.

Baker, G., & Caton, L. (Eds.). (2025). *AI-Powered Pedagogy and Curriculum Design: Practical Insights for Educators*. Routledge. https://doi.org/10.4324/9781003544920

Bandyopadhyay, A., Sardar, T.H., Mallik, S., Amin Hazarika, R., & Gourisaria, M.K. (Eds.). (2025). *AI and Data Engineering for Healthcare: Real-World Applications and Case Studies* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003600947

Boppiniti, S. T. (2023). Data Ethics in AI: Addressing Challenges in Machine Learning and Data Governance for Responsible Data Science. *International Scientific Journal for Research*, *5*(5), 1-29.

Brasioli, D., Guercio, L., Landini, G. G., & de Giorgio, A. (Eds.). (2025). *The Routledge Handbook of Artificial Intelligence and International Relations*. Routledge. https://doi.org/10.4324/9781003518495

Çela, E., Vajjhala, N. R., & Aslani, B. (Eds.). (2025). *Artificial Intelligence in Legal Systems: Bridging Law and Technology through AI*. CRC Press. https://doi.org/10.1201/9781003541899

Correia, P.M.A.R., Pedro, R.L.D., & Videira, S. (2025). Artificial Intelligence in Healthcare: Balancing Innovation, Ethics, and Human Rights Protection. *Journal of Digital Technologies and Law*, *3*(1), 143-180.

Curzon, J., Kosa, T. A., Akalu, R., & El-Khatib, K. (2021). Privacy and Artificial Intelligence. *IEEE Transactions on Artificial Intelligence,* 2(2), 96-108.

Darı, A. B., & Koçyiğit, A. (2024). Yapay Zekâ ve Etik: Yeni Medyanın Dönüşümünde Sorumluluk ve Sınırlar. *İletişim ve Toplum Araştırmaları Dergisi*, *4*(2), 246-261.

EDPS (European Data Protection Supervisor), (2025). *AI Risks Management Guidance*. European Data Protection Supervisor. https://www.edps.europa.eu/system/files/2025-11/2025-11-11_ai_risks_management_guidance_en.pdf#page=6.08.

Farina, M., Yu, X., & Chen, J. (Eds.). (2025). *Digital Development: Technology, Ethics and Governance* (1st ed.). Routledge. https://doi.org/10.4324/9781003567622

Galiana, L. I., Gudino, L. C., & González, P. M. (2024). Ethics and Artificial Intelligence. *Revista Clínica Española (English Edition)*, 224(3), 178-186.

Gupta, A., Amarnani, M., Soanki, S., & Kishore, J. (2025, February). AI and Data Privacy in Business. In *2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT)* (pp. 109-114). IEEE.

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, 30(1), 99-120.

Hermansyah, M., Najib, A., Farida, A., Sacipto, R., & Rintyarna, B. S. (2023). Artificial Intelligence and Ethics: Building An Artificial Intelligence System That Ensures Privacy and Social Justice. *International Journal of Science and Society*, 5(1), 154-168.

Huang, C., Zhang, Z., Mao, B., & Yao, X. (2022). An Overview of Artificial Intelligence Ethics. *IEEE Transactions on Artificial Intelligence*, 4(4), 799-819.

Jain, D. K., Bhatele, K. R., Garg, D., & Dhiman, G. (Eds.). (2025). *A Compendium of Responsible Artificial Intelligence*. CRC Press. https://doi.org/10.1201/9781003514497

Jobin, A., Ienca, M., & Vayena, E. (2019). The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence,* 1(9), 389-399.

Kashefi, P., Kashefi, Y., & Ghafouri Mirsaraei, A. (2024). Shaping The Future of AI: Balancing Innovation and Ethics in Global Regulation. *Uniform Law Review*, *29*(3), 524-548.

Kaur, D., & Singh, R. (2025). And Artificial Intelligence. *Next-Generation Computational Intelligence: Trends and Technologies*, 3.

Korobenko, D., Nikiforova, A., & Sharma, R. (2024). Towards A Privacy and Security-Aware Framework for Ethical AI: Guiding The Development And Assessment of AI Systems. In *Proceedings of The 25th Annual International Conference on Digital Government Research* (pp. 740-753).

Moghadasi, N., Valdez, R. S., Piran, M., Moghaddasi, N., Linkov, I., Polmateer, T. L., ... & Lambert, J. H. (2024). Risk Analysis of Artificial Intelligence in Medicine with A Multilayer Concept of System Order. *Systems*, *12*(2), 47.

NIST (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). https://doi.org/10.6028/nist.ai.100-1

OECD (2024). Data Governance and Privacy: Synergies and Areas of International Co-Operation. OECD Artıfıcıal Intellıgence Papers, OECD Publishing June 2024 No. 22 https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/ai-data-governance-and-privacy_2ac13a42/2476b1a4-en.pdf

Olawade, D. B., Wada, O. Z., Odetayo, A., David-Olawade, A. C., Asaolu, F., & Eberhardt, J. (2024). Enhancing Mental Health with Artificial Intelligence: Current Trends and Future Prospects. *Journal of Medicine, Surgery, And Public Health*, *3*, 100099.

Paliszkiewicz, J., & Gołuchowski, J. (Eds.). (2024). *Trust and Artificial Intelligence: Development and Application of AI Technology* (1st ed.). Routledge. https://doi.org/10.4324/9781032627236

Ravindran, B., & Singh, A. (Eds.). (2025). *Advancing Responsible AI in Public Sector Application: GPAI Edition* (1st ed.). CRC Press. https://doi.org/10.1201/9781003663577

Rosak-Szyrocka, J., & Wolniak, R. (2025). *AI-Driven Sustainability: The Future of Human Resources Management* (1st ed.). CRC Press. https://doi.org/10.1201/9781003668282

Safdar, N. M., Banja, J. D., & Meltzer, C. C. (2020). Ethical Considerations in Artificial Intelligence. *European Journal of Radiology,* 122, 108768.

Stahl, B. C., Brooks, L., Hatzakis, T., Santiago, N., & Wright, D. (2023). Exploring Ethics and Human Rights in Artificial Intelligence–A Delphi Study. *Technological Forecasting and Social Change*, *191*, 122502.

Szostak, M. (Ed.). (2025). *Organizations, Humanism and AI: An Aesthetics Approach* (1st ed.). Routledge. https://doi.org/10.4324/9781003662723

Tang, A. (2023). *Privacy in Practice: Establish and Operationalize a Holistic Data Privacy Program* (1st ed.). CRC Press. https://doi.org/10.1201/9781003225089

UNESCO. (2022). Recommendation on The Ethics Of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Villegas-Ch, W., & García-Ortiz, J. (2023). Toward A Comprehensive Framework for Ensuring Security and Privacy in Artificial Intelligence. *Electronics,* 12(18), 3786.

Wachter, S., & Mittelstadt, B. (2019). A Right to Reasonable Inferences: Re-Thinking Data Protection Law in The Age of Big Data and AI. *Colum. Bus. L. Rev.*, 494.

Wang, Y., Xiong, M., & Olya, H. (2020). Toward An Understanding of Responsible Artificial Intelligence Practices. In Proceedings of The 53rd Hawaii International Conference On System Sciences (pp. 4962-4971). Hawaii International Conference on System Sciences (HICSS).

Zhang, Q. (2022). Transformation of Modes of Production in The Age Of Artificial Intelli-Gence: An Analysis Based On The Phenom-Enon of "Human-Like Machines." *Journal of Socialist Theory Guide,* 2022(9):72-79.

Author Information

Hasan Küçükoğlu

https://orcid.org/0000-0002-7738-2567

Atatürk University

Department of Information Systems and Technologies

Türkiye

Contact: hasan.kucukoglu@atauni.edu.tr

Citation

Küçükoğlu, H. (2025). Data Ethics, Privacy, and Secure Artificial Intelligence. In A. Kaban, & T. Demirel (Eds.), *Interdisciplinary Applications of Artificial Intelligence - 1* (pp. 40-58). ISTES.

Chapter 4- Machine Learning

Hamza Polat

Chapter Highlights

- Machine learning is fundamentally a data-driven approach in which systems learn patterns from examples rather than relying on explicitly programmed rules.
- The primary objective of machine learning is generalization, meaning that models should perform well on previously unseen data rather than merely fitting the training set.
- Different machine learning paradigms are distinguished by the type of learning signal they use, including labeled data, unlabeled structure, weak or partial supervision, and reward-based interaction with an environment.
- Successful machine learning requires an end-to-end workflow that integrates data preprocessing, feature or representation design, model training, hyperparameter tuning, and transparent reporting.
- Rigorous evaluation practices, such as proper data splitting and cross-validation, are essential to obtain reliable estimates of real-world model performance and to avoid data leakage.
- Well-established algorithms such as regularized linear models, decision tree ensembles, and support vector machines remain highly effective when carefully tuned and appropriately validated.

Introduction

Machine learning (ML) is a discipline dedicated to the design and development of computational systems capable of autonomously improving their performance through exposure to data, rather than relying on static, explicitly programmed instructions (Jordan & Mitchell, 2015; Domingos, 2012). This approach addresses a central limitation of traditional rule-based programming: many complex, pattern-based tasks that are intuitive for humans—such as visual recognition, natural language understanding, and anomaly detection—are extraordinarily difficult to formalize into exhaustive logical statements. ML bridges this gap by employing algorithms that induce decision rules from empirical examples, positioning it as a foundational methodology for modern, data-centric scientific and industrial applications (LeCun et al., 2015; Bengio et al., 2013). At its most essential, ML provides a systematic framework for answering a critical question: How can a computational system leverage historical observations to enhance its predictive or decision-making capabilities for future instances? (Jordan & Mitchell, 2015).

A fundamental principle that distinguishes ML from mere curve-fitting or memorization is its core objective of generalization. This refers to a model's capacity to maintain high performance on novel, previously unseen data drawn from the same underlying distribution as the training set (Geman et al., 1992; Stone, 1974). A common and consequential misconception in early learning is to equate high accuracy on the training dataset with successful learning. A model may overfit by memorizing training samples or capturing spurious noise unique to that dataset, thereby achieving deceptively high training performance while failing catastrophically in real-world deployment (Geman et al., 1992; Dietterich, 1998). Consequently, mastering ML necessitates an understanding of it as an iterative, experimental practice. This practice is governed by a rigorous workflow that prioritizes robust evaluation, proper validation techniques, and transparent documentation of methods and results (Stone, 1974).

This chapter is structured to provide an accessible yet conceptually rigorous introduction to machine learning for undergraduate students. It begins by defining the field and its primary categorical paradigms, establishing the crucial link between learning, model complexity, and the pursuit of generalization. Following this conceptual groundwork, the chapter outlines the standard ML project lifecycle—encompassing data preprocessing, model training, validation, and evaluation. It then surveys major families of learning algorithms, explaining their operational principles and typical use cases at a high level (Domingos, 2012).

Machine Learning: A Data-Driven Paradigm

Machine learning (ML), a critical sub-discipline of artificial intelligence, refers to computational methods that allow systems to autonomously identify patterns within data and thereby enhance their performance on specified tasks, without the need for manually programmed, explicit rules (Janiesch et al., 2021; Bernardes, 2024; Puppala, 2025). This represents a fundamental shift from traditional programming. In conventional software development, engineers must anticipate all possible scenarios and encode precise logical instructions to govern system behavior. In contrast, ML transfers this burden from the programmer to the dataset itself. Within this paradigm, practitioners

provide representative data, define a quantifiable objective—such as the minimization of prediction error—and implement algorithms that automatically induce models capable of generalizing beyond the specific examples presented during training (Simeone, 2017; Kubát, 2017).

This inductive, data-centric methodology is uniquely suited to addressing complex, high-dimensional, and evolving problems where formulating exhaustive rules is impractical or impossible. Once the training phase is complete, the resultant model operates as an inference engine, enabling predictions, classifications, or automated decisions on novel instances. Practical applications are vast, encompassing spam filtering, image recognition, financial forecasting, and beyond. Consequently, ML has transitioned from a research specialty to a foundational technology powering contemporary innovation, from personalized recommendation engines to advanced autonomous systems.

The Nature of "Learning" in Machine Learning

In machine learning, the concept of "learning" is defined operationally as the systematic improvement of a system's performance on a task through exposure to data, which constitutes its experience. Formally, the process involves providing an algorithm with a dataset and a well-defined task objective. The system then engages in an iterative optimization procedure, methodically adjusting its internal parameters to enhance a predefined performance metric, such as classification accuracy or the reduction of prediction error (Edwards et al., 2020; Karthick, 2024). Thus, learning in this context is not a singular event but a continuous search for the optimal parameter configuration that maximizes performance given the constraints and information present in the training data.

A defining characteristic of this process is its reliance on learning from examples rather than from pre-coded rules. Instead of executing deterministic logical statements, ML systems perform inductive inference, deriving general principles from a multitude of specific data points. While this bears a loose analogy to human experiential learning, a key distinction lies in data efficiency. Human cognition can often achieve reliable generalization from limited examples by leveraging prior knowledge and contextual reasoning. Most current ML models, however, typically require large-scale, high-quality datasets to attain similar levels of robustness and reliability, as their performance is directly correlated with the breadth and representativeness of their training experience (Cohen, 2021).

A core capability enabled by this example-based learning is pattern discovery. Through sophisticated statistical and computational techniques, ML algorithms can uncover latent structures, correlations, and relationships within data that are not readily apparent through manual analysis. For instance, when applied to large-scale transactional data, a model might reveal subtle, non-linear factors influencing consumer purchasing behavior, thereby providing organizations with actionable insights for data-informed decision-making (Alpaydin, 2021). This capacity to extract and leverage meaningful patterns from complexity is a primary driver for ML adoption across diverse fields such as finance, healthcare, and educational technology.

Model Building and the Imperative of Generalization

The culmination of the learning process is the construction of a model. This model is a formal mathematical abstraction—whether a function, a set of learned parameters, or a probabilistic graph—that encapsulates the discovered knowledge. It serves as a mapping from input features to desired outputs, such as labels, predictions, or actions (Janiesch et al., 2021; , 2021). This abstraction is pivotal, as it represents a distilled understanding of the training data's underlying patterns, moving beyond simple memorization.

The ultimate quality and utility of any machine learning system are predominantly determined by this model's ability to generalize. Generalization refers to the model's capacity to maintain accurate and reliable performance when applied to new, previously unseen data that originates from the same underlying distribution as the training set. The central challenge in ML is to avoid overfitting, where a model learns the noise and specific idiosyncrasies of the training data at the expense of the general signal. Therefore, the success of an ML project is measured not by its performance on familiar training data, but by its effective and accurate operation in novel, real-world scenarios. This principle places generalization at the heart of both evaluating and designing machine learning systems.

Main Paradigms of Machine Learning

Machine learning is not a monolithic field but rather a collection of distinct learning paradigms, each shaped by the nature of the problem at hand and the type of data available. These paradigms are primarily differentiated by the source and structure of the learning signal provided to the algorithm (Hsu, 2022; Wang, 2025). The fundamental distinction lies in whether the data is labeled and whether the system learns through interaction with an environment. The following table provides a comparative summary of the four main paradigms, outlining their core principles, data requirements, and typical application domains.

Table 1. Key Machine Learning Paradigms and Their Purposes

Paradigm	Core Principle	Data Requirement	Typical Tasks
Supervised learning	Learn from labeled input – output pairs	Labeled dataset (features & target variables).	Classification, regression
Unsupervised learning	Find structure in unlabeled data	Unlabeled dataset (features only).	Clustering, dimensionality reduction
Semi-supervised learning	Combine few labels with many unlabeled points	Small labeled dataset + large unlabeled dataset.	Low-label settings
Reinforcement learning	Learn by trial and error via rewards	An interactive environment.	Control, robotics, games

The comparison in Table clarifies the operational framework of each paradigm. Supervised learning requires an explicit target variable for predicting future outcomes from historical data (Hsu, 2022; Wang, 2025; Badillo et al., 2020), while unsupervised learning analyzes unlabeled data for exploration and preprocessing (Hsu, 2022; Wang,

2025; Janiesch et al., 2021). Semi-supervised learning offers a practical hybrid, merging these approaches for resource efficiency in real-world scenarios where labeling is costly (Lu et al., 2024). Reinforcement learning is fundamentally distinct, requiring continuous interaction with a dynamic environment rather than a static dataset and focusing on optimizing a long-term objective (Alpaydin, 2021; Janiesch et al., 2021; Du et al., 2025). Therefore, the choice of paradigm is dictated by the answers to the questions: "What do we know?" (labels), "What are we trying to understand?" (structure), and "How does the system receive feedback?" (learning signal).

Categories of Machine Learning

Supervised Learning

Supervised learning is a central paradigm in contemporary machine learning, characterized by the use of labeled data to train algorithms that predict outcomes or classify observations. In supervised settings, each training instance consists of an input vector and a corresponding target output (label), and the overarching objective is to learn a function that maps inputs to outputs with minimal error on new, unseen data (Shetty et al., 2022; Singh, Thakur, & Sharma, 2016). Conceptually, this process resembles learning under guidance: the algorithm is repeatedly exposed to examples of "questions" (inputs) paired with "correct answers" (labels) and incrementally adjusts its internal parameters so that it can respond accurately when presented with novel questions (Baladram, Koike, & Yamada, 2020).

Supervised learning problems are commonly divided into classification and regression, depending on the nature of the target variable. In classification tasks, the target variable is categorical (e.g., disease vs. no disease, spam vs. not spam), and the model aims to assign each instance to one of a finite set of predefined classes (Sen, Hajra, & Ghosh, 2019). In regression tasks, by contrast, the target variable is continuous (e.g., blood pressure levels or house prices), and the goal is to predict numerical values that approximate the true outcomes as closely as possible (Shetty et al., 2022). A wide range of algorithms can be applied to both types of problems, including linear and logistic regression, decision trees, random forests, k-nearest neighbors, support vector machines, and neural networks. Each of these methods embodies distinct inductive biases and trade-offs, making them suitable for different data characteristics and application contexts (Baladram et al., 2020; Sen et al., 2019).

The supervised learning workflow typically involves several interrelated stages. First, labeled data are collected and partitioned into training, validation, and test sets to enable reliable estimation of generalization performance (Syed & Lokhande, 2024; Burkart & Huber, 2020). Second, data preprocessing and feature engineering are carried out to address missing values, reduce dimensionality, and construct informative predictors that capture relevant aspects of the underlying phenomenon (Jiang, Gradus, & Rosellini, 2020; Singh et al., 2016). Third, a suitable model class is selected and trained by minimizing a loss function that quantifies prediction error; this optimization is commonly performed using gradient-based or related numerical methods (Baladram et al., 2020). Fourth, model performance is assessed using resampling techniques such as k-fold cross-validation and task-appropriate evaluation metrics (e.g., accuracy, AUC, or mean squared error), which help balance the bias–variance trade-off and mitigate overfitting (Pargent, Schoedel, & Stachl, 2023; Rainio, Teuho, & Klén, 2024). Finally, the selected

model is interpreted, examined for robustness and fairness where relevant, and deployed within real-world application settings (Pargent et al., 2023; Burkart & Huber, 2020).

Supervised learning has become pervasive across a wide range of domains, including healthcare, psychology, finance, and drug discovery. In clinical and psychological research, supervised models are used to predict symptom trajectories, treatment responses, and risks of adverse outcomes from high-dimensional and multimodal data sources such as self-reports, neuroimaging, and physiological signals (Jiang et al., 2020; Pargent et al., 2023). In medicine more broadly, supervised algorithms underpin diagnostic classifiers, prognostic models, and clinical decision support systems, often operating on complex data types such as medical images or genomic profiles (Lo Vercio et al., 2020; Pienaar et al., 2025). In drug discovery, supervised learning has been applied to activity prediction, toxicity assessment, and molecular property optimization, thereby accelerating early-stage screening and design processes (Obaido et al., 2024). Beyond the biomedical domain, applications include spam filtering, recommendation systems, credit scoring, fraud detection, and autonomous driving, underscoring the versatility and broad impact of supervised approaches (Shetty et al., 2022).

Despite its effectiveness, supervised learning presents important methodological and ethical challenges. Acquiring high-quality labeled data is often costly and time-consuming, and label noise can substantially degrade model performance (Obaido et al., 2024; Singh et al., 2016). In addition, models may overfit idiosyncratic patterns in the training data, leading to poor generalization, particularly in high-dimensional settings with limited sample sizes (Jiang et al., 2020; Rainio et al., 2024). Many high-performing models, such as deep neural networks and kernel-based methods, are also criticized for their lack of transparency, motivating growing interest in explainable and interpretable machine learning, especially in high-stakes domains like healthcare and finance (Burkart & Huber, 2020; Lo Vercio et al., 2020). Furthermore, when training data reflect historical or societal biases, supervised models risk perpetuating or amplifying inequities, prompting increased attention to fairness, accountability, and responsible deployment (Burkart & Huber, 2020; Pargent et al., 2023).
In summary, supervised learning provides a rigorous, data-driven framework for learning predictive mappings from labeled examples and constitutes a cornerstone of modern artificial intelligence. Ongoing advances in algorithms, evaluation methodologies, interpretability, and ethical governance—along with integration with semi-supervised and unsupervised approaches—are central to realizing the full potential of supervised learning in complex, real-world settings (Jiang et al., 2020; Obaido et al., 2024; Rahaman, 2024).

Unsupervised Learning

Unsupervised learning is a core paradigm of machine learning in which algorithms analyze unlabeled data to uncover patterns, structures, or relationships without access to explicit target outputs or external reward signals (Simeone, 2018). Unlike supervised learning, where models are trained to approximate a known input – output mapping, unsupervised methods operate solely on observed inputs and seek to model properties of the underlying data distribution or its latent structure (Hinton & Sejnowski, 2018; Simeone, 2018). This paradigm is particularly valuable in modern data-intensive settings, where acquiring reliable labels is expensive, slow, or infeasible, yet vast amounts of raw data are readily available (Naeem et al., 2023; Chen et al., 2022; Li, 2024).

In contrast to supervised learning's predictive orientation, unsupervised learning primarily supports exploratory data analysis. Its goals include identifying groups of similar instances, discovering latent dimensions, detecting anomalies, and uncovering associations among variables without predefined notions of correctness (Hodeghatta & Nayak, 2017; Simeone, 2018). Typical tasks encompass clustering, in which objects are grouped so that members within the same cluster share similar characteristics; association rule mining, which reveals frequently co-occurring items or events; dimensionality reduction, which compresses high-dimensional observations into a lower-dimensional representation; and anomaly detection, which highlights data points that deviate markedly from normal patterns (Naeem et al., 2023; Sharma, 2020; Hodeghatta & Nayak, 2017; Usmani et al., 2022). From a statistical viewpoint, many unsupervised techniques can be interpreted as learning representations that reflect the intrinsic structure of the input space rather than optimizing prediction accuracy (Hinton & Sejnowski, 2018; Simeone, 2018).

A broad spectrum of algorithms operationalizes these ideas (see Table 2). Clustering methods, such as k-means, hierarchical clustering, DBSCAN, and Gaussian mixture models, are widely used to discover hidden groupings or latent subpopulations within data (Eckhardt et al., 2022; Naeem et al., 2023; Wong, 2021). Dimensionality reduction techniques, including principal component analysis (PCA), factor analysis, and multidimensional scaling, aim to identify low-dimensional manifolds that capture most of the variance in high-dimensional datasets, thereby facilitating analysis, visualization, and noise reduction (Eckhardt et al., 2022; Naeem et al., 2023; Wong, 2021; Simeone, 2018). Association rule learning algorithms, such as Apriori, ECLAT, and FP-growth, focus on extracting frequent patterns and co-occurrence rules, particularly in transactional and market-basket data (Naeem et al., 2023; Hodeghatta & Nayak, 2017). More recently, deep unsupervised and self-supervised approaches, including autoencoders, contrastive learning, and representation learning frameworks, have become central to learning transferable features from raw data in vision, language, and multimodal contexts (Chen et al., 2022; Gui et al., 2023; Zhou & Xu, 2025).

Unsupervised learning underpins applications across a wide range of domains. In healthcare, clustering and dimensionality reduction techniques are used to analyze high-dimensional clinical and imaging data, identify latent patient subgroups, and support personalized or precision medicine initiatives (Eckhardt et al., 2022; Pantanowitz et al., 2024). In communications and networking, unsupervised models characterize traffic patterns, detect anomalies, and inform data-driven communication strategies when labeled outcomes are unavailable (Simeone, 2018). In industrial systems and cybersecurity, unsupervised anomaly detection frameworks are employed to identify unusual behaviors that may signal faults, failures, or intrusions (Eppa, 2025; Usmani et al., 2022). More broadly, unsupervised learning plays a key role in market segmentation, recommendation systems, fraud detection, and other data-rich settings where meaningful structure must be inferred directly from observations (Naeem et al., 2023; Eppa, 2025; Hodeghatta & Nayak, 2017).

Despite its strengths, unsupervised learning poses distinct methodological challenges. In the absence of labeled ground truth, evaluation and model selection are inherently difficult: different algorithms or parameter choices can yield divergent clusterings or representations, with no definitive criterion for determining which solution is

"correct" (Watson, 2023). Critical design decisions—such as the number of clusters, choice of similarity measures, or dimensionality of latent spaces—can substantially influence results, and many algorithms are sensitive to noise and outliers (Watson, 2023; Sharma, 2020). Interpretability constitutes another major concern, as the discovered structures do not necessarily correspond to meaningful or actionable domain concepts, raising epistemic and ethical questions about how such patterns should inform decision-making processes (Watson, 2023; Pantanowitz et al., 2024).

Table 2. Core Types and Examples of Unsupervised Learning

Type of task	Main goal	Typical algorithms / examples
Clustering	Group similar instances into clusters	K-means, hierarchical, DBSCAN, Gaussian mixtures
Dimensionality reduction	Compress data, reveal low-dimensional structure	PCA, factor analysis, multidimensional scaling
Association rule mining	Find frequent co-occurring items/patterns	Apriori, ECLAT, FP-growth
Anomaly/outlier detection	Detect unusual or rare behaviors	Clustering-based, reconstruction-based, density models
Deep representation learning	Learn rich features from unlabeled data	Autoencoders, contrastive/self-supervised methods

Nevertheless, the capacity to leverage vast amounts of unlabeled data has made unsupervised learning increasingly central to modern artificial intelligence. It provides principled mechanisms for organizing and compressing data, discovering latent structure, and learning general-purpose representations that can be reused in downstream supervised tasks (Naeem et al., 2023; Chen et al., 2022; Gui et al., 2023). As research progresses in self-supervised and hybrid learning paradigms, unsupervised methods are poised to play a foundational role in building more data-efficient, adaptable, and scalable intelligent systems (Watson, 2023; Chen et al., 2022; Gui et al., 2023).

Semi-Supervised and Weakly Supervised Learning

Semi-supervised learning (SSL) and weakly supervised learning (WSL) occupy an intermediate position between classical supervised and unsupervised paradigms, directly addressing the growing mismatch between the abundance of raw data and the scarcity or imperfection of labels. In many real-world applications—such as medical imaging, industrial monitoring, natural language processing, and web-scale computer vision—obtaining accurate, instance-level annotations is costly, time-consuming, or even infeasible, whereas unlabeled or weakly labeled data are readily available at scale (Van Engelen & Hoos, 2019; Zhou, 2018; Ren et al., 2023; Martínez-Heredia & Ventura, 2025; Qi & Luo, 2019). These learning regimes aim to exploit imperfect supervision in order to construct predictive models that are more accurate and data-efficient than purely unsupervised approaches, while substantially reducing the annotation burden associated with fully supervised learning.

Semi-Supervised Learning

Semi-supervised learning leverages a small set of labeled examples together with a typically much larger pool of unlabeled data to improve performance on tasks such as classification, regression, and clustering (Van Engelen & Hoos, 2019; Nartey et al., 2020; Zhu & Goldberg, 2009). Conceptually, SSL lies between supervised learning, where all training instances are labeled, and unsupervised learning, where no labels are available (Van Engelen & Hoos, 2019; Song et al., 2021; Zhu & Goldberg, 2009). The core intuition behind SSL is that unlabeled data contain valuable information about the structure of the input space—such as cluster organization or low-dimensional manifolds—that can be exploited to regularize decision boundaries and learn more robust representations (Van Engelen & Hoos, 2019; Song et al., 2021; Yang et al., 2021; Qi & Luo, 2019).

In typical SSL scenarios, labeled data are expensive or difficult to obtain (e.g., expert-annotated medical images, industrial quality labels, or linguistic annotations), while unlabeled data are inexpensive and abundant (Van Engelen & Hoos, 2019; Cacciarelli & Kulahci, 2025; Li et al., 2024). To exploit this asymmetry, SSL methods rely on assumptions linking the data distribution to labels, including the cluster assumption (points in the same cluster tend to share a label), the low-density separation assumption (decision boundaries should pass through regions of low data density), and various smoothness or manifold assumptions (Van Engelen & Hoos, 2019; Song et al., 2021; Mey & Loog, 2022; Qi & Luo, 2019). When these assumptions hold, unlabeled data can significantly improve generalization beyond what is achievable with labeled data alone. However, when assumptions are violated, incorporating unlabeled data may degrade performance, highlighting the need for careful modeling and validation (Van Engelen & Hoos, 2019; Mey & Loog, 2022).

Algorithmically, SSL encompasses a diverse family of approaches, including generative models that estimate joint or conditional data distributions, graph-based methods that propagate labels across similarity graphs, margin-based methods adapted to semi-supervised settings, and self-training or pseudo-labeling schemes that iteratively assign labels to confident unlabeled examples (Van Engelen & Hoos, 2019; Song et al., 2021; Yang et al., 2021; Zhu & Goldberg, 2009). In the context of deep learning, semi-supervised techniques such as consistency regularization, semi-supervised generative modeling, and pseudo-labeling have become prominent, often formulated as the combination of a supervised loss on labeled samples and an unsupervised or consistency loss on unlabeled data (Yang et al., 2021; Qi & Luo, 2019). These methods have demonstrated strong empirical performance in domains including medical image analysis, industrial predictive modeling, and fine-grained visual recognition, where large unlabeled datasets coexist with limited high-quality labels (Ren et al., 2023; Cacciarelli & Kulahci, 2025; Nartey et al., 2020; Li et al., 2024; Qi & Luo, 2019).

Weakly Supervised Learning

Weakly supervised learning generalizes the idea of limited supervision by relaxing the requirement for strong, fully accurate labels and allowing multiple forms of imperfect supervision. Rather than assuming precise instance-level ground truth, WSL encompasses settings in which labels are incomplete, coarse-grained, or noisy, yet still informative enough to train effective models (Zhou, 2018; Ren et al., 2023; Basu, 2019; Martínez-Heredia &

Ventura, 2025). A widely adopted taxonomy distinguishes three primary forms of weak supervision: incomplete supervision, where only a subset of instances is labeled; inexact supervision, where labels are coarse (e.g., bag-level labels in multi-instance learning or case-level labels instead of pixel-level annotations); and inaccurate supervision, where labels are corrupted or noisy and may not always correspond to the true class (Zhou, 2018; Li et al., 2021; Ren et al., 2023; Basu, 2019; Martínez-Heredia & Ventura, 2025; Yang et al., 2021; Qi & Luo, 2019).

From this perspective, semi-supervised learning can be viewed as a prototypical instance of weak supervision corresponding to incomplete supervision, since only a subset of data points carries labels while the remainder remains unlabeled (Zhou, 2018; Li et al., 2021; Basu, 2019; Martínez-Heredia & Ventura, 2025; Yang et al., 2021). Other WSL paradigms include learning with noisy labels, partial or ambiguous labeling, and multi-instance learning, all of which trade label precision or granularity for increased scalability and reduced annotation costs (Zhou, 2018; Li et al., 2021; Basu, 2019; Martínez-Heredia & Ventura, 2025; Sun et al., 2020; Zhou et al., 2025). The appeal of WSL is particularly pronounced in medical imaging and radiology, where fine-grained annotations are prohibitively expensive, but weaker signals—such as image-level diagnoses, clinical reports, or automatically mined labels—can still support effective deep learning models (Ren et al., 2023; Misera et al., 2024).

A central challenge in weakly supervised learning concerns safety and robustness. Unlike fully supervised learning, incorporating additional weakly supervised data does not always guarantee improved performance and may sometimes lead to degradation (Li et al., 2021; Mey & Loog, 2022). This risk has motivated research on safe weakly supervised learning, ensemble strategies, and robust optimization objectives that aim to ensure non-decreasing or worst-case-bounded performance gains when combining multiple weak supervision sources (Li et al., 2021; Mey & Loog, 2022). Recent work also explores principled frameworks for integrating heterogeneous weak signals—such as noisy annotators, heuristic rules, or similarity-based cues—and for correcting or denoising labels while controlling error propagation (Li et al., 202; Tate et al., 2025; Zhou et al., 2025).

Taken together, semi-supervised and weakly supervised learning provide a conceptual and practical bridge between idealized supervised learning and the imperfect supervision that dominates real-world data. By explicitly modeling and exploiting unlabeled and weakly labeled data, these paradigms aim to reduce annotation costs, unlock previously underutilized data sources, and enable scalable learning in data-rich yet label-scarce environments (Van Engelen & Hoos, 2019; Zhou, 2018; Ren et al., 2023; Martínez-Heredia & Ventura, 2025; Yang et al., 2021; Qi & Luo, 2019).

Reinforcement Learning

Reinforcement learning (RL) is a major paradigm in machine learning focused on sequential decision-making under uncertainty. Instead of learning from labeled examples, an RL agent learns by interacting with an environment, choosing actions, and receiving scalar feedback in the form of rewards or punishments over time (Shakya et al., 2023; François-Lavet et al., 2018; Kaelbling et al., 1996; Sutton & Barto, 2005). The objective is to learn a policy – a mapping from states to actions—that maximizes the cumulative (often discounted) reward,

making RL well suited to tasks where actions have long-term consequences rather than immediate, one-shot outcomes (François-Lavet et al., 2018; Kaelbling et al., 1996; Yu et al., 2019; Sutton & Barto, 2005).

Conceptually, RL is often formalized using Markov decision processes (MDPs), in which an agent observes a state, selects an action, transitions to a new state, and receives a reward, repeating this cycle over discrete time steps (Naeem et al., 2020; Kaelbling et al., 1996; Yu et al., 2019; Sutton & Barto, 2005). This framework captures key challenges not addressed by standard supervised or unsupervised learning, including delayed rewards, stochastic transitions, and the need to balance exploration (trying new actions to gain information) and exploitation (using current knowledge to obtain high reward) (Naeem et al., 2020; François-Lavet et al., 2018; Kaelbling et al., 1996; Yu et al., 2019). Classical dynamic programming methods can solve small MDPs with known dynamics, but RL extends these ideas to large or unknown environments by learning value functions and policies directly from experience using techniques such as temporal-difference learning, Q-learning, and policy gradient methods (François-Lavet et al., 2018; Kaelbling et al., 1996; Yu et al., 2019; Sutton & Barto, 2005; Gosavi, 2009).

Recent advances in deep reinforcement learning (DRL), which combine RL with deep neural networks, have dramatically extended the range of solvable problems. Deep networks serve as powerful function approximators for value functions and policies, enabling agents to operate in high-dimensional state spaces such as raw images or sensor streams (Shakya et al., 2023; Naeem et al., 2020; Sewak, 2020; Nian et al., 2020). DRL methods have achieved superhuman performance in complex games, robust control in robotics, and adaptive strategies in domains like healthcare, energy management, finance, and industrial process control (Shakya et al., 2023; Naeem et al., 2020; Sewak, 2020; Littman, 2015; Wörgötter & Porr, 2019; Nian et al., 2020; Sivamayil et al., 2023; Al-Hamadani et al., 2024). These successes have made RL a central tool for designing autonomous systems that must learn to act effectively in dynamic, uncertain, and partially known environments.

Despite this progress, RL remains challenging: sample inefficiency, stability and safety of learning, reward design, and deployment in real-world systems with constraints and partial observability are active research topics (Shakya et al., 2023; Sewak, 2020; Littman, 2015; Wörgötter & Porr, 2019; Nian et al., 2020; Sivamayil et al., 2023; Al-Hamadani et al., 2024). Nevertheless, the ability to learn from evaluative feedback and experience positions reinforcement learning as a foundational component of modern artificial intelligence and control.

Main Characteristics of Reinforcement Learning

While supervised, unsupervised, and semi-supervised learning paradigms primarily focus on learning from static datasets, reinforcement learning (RL) addresses a fundamentally different class of problems in which learning arises through sequential interaction with an environment. Rather than relying on labeled examples or weak supervision, reinforcement learning agents learn by taking actions, observing their consequences, and receiving feedback in the form of scalar reward signals. This interaction-driven perspective makes reinforcement learning

particularly well suited to dynamic, decision-making tasks where outcomes depend on long-term consequences rather than immediate correctness.

Table 3. Core Elements and Properties of Reinforcement Learning.

Aspect	Brief description
Learning signal	Scalar rewards, often delayed and stochastic
Problem formulation	Markov decision processes, sequential decisions
Core challenge	Exploration–exploitation trade-off
Key algorithms	TD learning, Q-learning, policy gradients, actor–critic
Modern extension	Deep reinforcement learning with neural networks

Table X summarizes the main characteristics of reinforcement learning, highlighting how RL differs from other machine learning paradigms in terms of learning signals, problem formulation, and algorithmic challenges. First, the learning signal in RL consists of scalar rewards, which are often delayed and stochastic, meaning that the quality of an action may only become apparent after many subsequent steps (Shakya et al., 2023; François-Lavet et al., 2018; Kaelbling et al., 1996; Yu et al., 2019; Sutton & Barto, 2005). This contrasts sharply with supervised learning, where immediate and explicit error feedback is available for each training instance.

Second, reinforcement learning problems are typically formalized using Markov decision processes (MDPs), which provide a mathematical framework for modeling sequential decision-making under uncertainty. Within this framework, an agent observes states, selects actions according to a policy, transitions to new states, and accumulates rewards over time (Naeem et al., 2020; Kaelbling et al., 1996; Yu et al., 2019; Sutton & Barto, 2005). This sequential structure introduces dependencies across time steps and requires the agent to reason about long-term returns rather than isolated predictions.

A central challenge in reinforcement learning is the exploration–exploitation trade-off. Agents must balance exploiting actions that are known to yield high rewards with exploring uncertain actions that may lead to better outcomes in the future (Naeem et al., 2020; François-Lavet et al., 2018; Kaelbling et al., 1996). Effective learning depends critically on managing this trade-off, as insufficient exploration can trap the agent in suboptimal behavior, while excessive exploration can hinder convergence.

The table also outlines core algorithmic families in reinforcement learning, including temporal-difference (TD) learning, Q-learning, policy gradient methods, and actor–critic architectures. These approaches differ in how they represent and optimize policies or value functions, but all aim to estimate expected long-term rewards from experience rather than from labeled examples (François-Lavet et al., 2018; Kaelbling et al., 1996; Yu et al., 2019; Sutton & Barto, 2005; Gosavi, 2009).

Finally, Table X highlights deep reinforcement learning as a modern extension that integrates reinforcement learning with deep neural networks. By using neural networks as function approximators for policies or value functions, deep reinforcement learning enables RL agents to operate in high-dimensional state spaces such as images, sensor streams, or raw signals (Shakya et al., 2023; Naeem et al., 2020; Sewak, 2020; Nian et al., 2020).

This development has significantly expanded the applicability of reinforcement learning to complex real-world problems, including robotics, autonomous systems, and large-scale control tasks.

The Machine Learning Workflow

Modern machine learning (ML) projects are best understood as end-to-end workflows rather than isolated model-building steps. Across application domains such as medicine, environmental science, and construction, empirical evidence consistently shows that predictive performance, robustness, and real-world usefulness depend not only on the choice of learning algorithm but also—often more critically—on how data are prepared, features are constructed, and models are evaluated (Golazad et al., 2024; Pfob et al., 2022). Consequently, contemporary ML practice emphasizes systematic pipelines that integrate multiple stages, from raw data handling to transparent reporting.

Table 4. Core Stages and Roles in the Machine Learning Workflow

Stage	Core role in pipeline
Data preprocessing	Clean, transform, and structure raw data
Feature engineering & representation	Construct or learn informative, often interpretable features
Model training & hyperparameter tuning	Fit models and optimize configuration for target tasks
Reproducibility & reporting	Ensure transparent, traceable, and reusable ML analyses

As seen Table 4, a typical machine learning workflow comprises data preprocessing, feature engineering and representation learning, model training and hyperparameter tuning, and reproducible reporting and documentation (Chai et al., 2023; Mumuni & Mumuni, 2024). These stages are tightly coupled: decisions made early in the pipeline constrain later modeling choices and directly influence the credibility of reported results.

Data Preprocessing

Data preprocessing focuses on transforming raw, heterogeneous, and often noisy data into a form suitable for analysis. Common preprocessing steps include data cleaning, handling missing values, outlier detection and treatment, normalization or standardization, and encoding of categorical variables (Rao et al., 2023; Yasodha, 2025). Systematic reviews across multiple domains demonstrate that poorly designed or ad hoc preprocessing can introduce bias, amplify spurious correlations, and reduce generalizability, whereas carefully selected preprocessing strategies substantially improve model accuracy and stability (Golazad et al., 2024; Shahidi et al., 2025). To address these challenges, automated and standardized preprocessing pipelines are increasingly proposed, particularly to support non-expert practitioners and to reduce variability in analytical practice (Mumuni & Mumuni, 2024).

Feature Engineering and Representation Learning

Beyond basic data cleaning, feature engineering and representation learning determine how domain information is presented to learning algorithms. In many structured-data applications, handcrafted feature construction, selection, and transformation remain essential for embedding domain expertise into models (Mumuni & Mumuni, 2024; Xiao et al., 2023). In contrast, for high-dimensional or multimodal data, learned representations—obtained through matrix factorization, deep neural networks, or fuzzy and information-theoretic approaches—play a central role (Yan et al., 2025).

Recent studies emphasize that effective feature design requires balancing predictive performance, interpretability, and computational cost. Moreover, incorporating domain structure—such as physical symmetries, spatial relationships, or chemistry–structure coupling—can significantly enhance generalization and robustness (Alam et al., 2025; Xiao et al., 2025). Thus, feature engineering is not merely a technical step but a key site where methodological rigor and domain understanding intersect.

Model Training and Hyperparameter Tuning

Once features are defined, model training and hyperparameter tuning convert prepared data into predictive models. Comparative studies in medical and environmental ML underscore the importance of principled data splitting, strict avoidance of data leakage, and rigorous cross-validation to obtain trustworthy estimates of generalization performance (Pfob et al., 2022). Hyperparameters—such as regularization coefficients, learning rates, or tree depths—control model complexity and learning dynamics and must therefore be tuned systematically rather than adjusted informally.

To this end, grid search, random search, and increasingly AutoML frameworks are widely adopted to explore model and hyperparameter spaces efficiently and reproducibly (Chai et al., 2023; Eldeeb & Elshawi, 2025; Ko & Kang, 2025). These approaches help mitigate subjective bias in model selection and improve comparability across studies.

Reproducibility and Reporting

Finally, reproducibility and transparent reporting have emerged as central concerns in machine learning research and practice. Surveys of applied ML highlight the need for detailed documentation of datasets, preprocessing choices, feature representations, model configurations, hyperparameter tuning procedures, and evaluation metrics, alongside sharing code and, where feasible, data (Chai et al., 2023; Pfob et al., 2022). In response, automated pipelines and agent-based systems have been proposed to enforce consistent, traceable workflows and to generate standardized reports, thereby reducing human error and enhancing repeatability (Ko & Kang, 2025; Raghhul, 2025).

Overview of Popular Algorithms

Modern machine learning practice relies on a relatively small set of foundational algorithms that have proven effective across a wide range of application domains. These algorithms differ in how they balance flexibility, interpretability, computational efficiency, and robustness to noise. In supervised learning settings, linear models with regularization, decision trees, tree-based ensemble methods (such as random forests and boosting), and support vector machines (SVMs) are widely regarded as core techniques for both classification and regression tasks in fields ranging from engineering and finance to medicine and the social sciences (Faouzi & Colliot, 2023; Ghildiyal et al., 2024; Kumar, 2024).

Rather than being replaced by newer methods, these "classic" algorithms continue to form the backbone of many successful ML pipelines. Empirical comparisons consistently show that, when properly tuned and combined with sound preprocessing and evaluation practices, they can achieve performance comparable to or even exceeding that of more complex models, particularly on structured and tabular data (Fernández-Delgado et al., 2019).

Linear Models with Regularization

Linear models provide a simple and highly interpretable baseline by assuming a linear relationship between predictors and the target variable. Despite their apparent simplicity, linear models remain powerful, especially when augmented with regularization techniques such as ridge, lasso, or elastic net. Regularization controls model complexity, mitigates multicollinearity, and improves generalization in high-dimensional settings (Abdulhafedh, 2022). Large-scale benchmark studies demonstrate that well-tuned regularized linear models often remain competitive with more flexible nonlinear approaches, particularly when relationships in the data are approximately linear or when interpretability is a priority (Fernández-Delgado et al., 2019).

Decision Trees and Ensemble Methods

Decision trees model data using a sequence of hierarchical if–then rules, producing models that are easy to interpret and capable of capturing nonlinear relationships and feature interactions without explicit transformations (Mienye & Jere, 2023; Abdulhafedh, 2022). However, individual trees are known to be unstable and prone to overfitting, as small changes in the training data can lead to substantially different tree structures.

To address these limitations, ensemble methods combine multiple base learners to improve predictive performance and robustness (González et al., 2020; Mienye & Sun, 2022). Among these, random forests aggregate predictions from many randomized trees, substantially reducing variance while retaining some interpretability through variable importance measures (Abdulhafedh, 2022). Boosting methods, such as AdaBoost and gradient boosting, build models sequentially, with each new learner focusing on correcting the errors of previous ones. These approaches frequently achieve state-of-the-art performance on real-world problems, albeit at the cost of

increased model complexity and reduced transparency (Fernández-Delgado et al., 2019; González et al., 2020; Mienye & Sun, 2022).

Support Vector Machines

Support vector machines (SVMs) adopt a fundamentally different perspective by framing learning as a margin maximization problem. For classification, SVMs seek decision boundaries that maximize the separation between classes; for regression, they aim to fit functions within a specified tolerance margin. Through the use of kernel functions, SVMs can implicitly map data into high-dimensional feature spaces, enabling them to model complex nonlinear relationships (Faouzi & Colliot, 2023). Comparative reviews consistently identify SVMs—alongside random forests and boosted trees—as among the most accurate classical ML algorithms, particularly when hyperparameters and kernel choices are carefully tuned (Fernández-Delgado et al., 2019; Ghildiyal et al., 2024; Kumar, 2024).

Conclusion

This chapter introduced machine learning as a data-driven approach for building systems that improve with experience, emphasizing that the real measure of success is generalization—reliable performance on new, unseen data—rather than high accuracy on the training set. We contrasted the major learning paradigms (supervised, unsupervised, semi-/weakly supervised, and reinforcement learning) by focusing on the type of learning signal available (labels, structure, weak labels, or rewards) and the kinds of questions each paradigm is designed to answer. We also presented machine learning as an end-to-end workflow—preprocessing, feature/representation design, model training and hyperparameter tuning, and reproducible reporting—where methodological choices at each stage directly shape validity, robustness, and trustworthiness. Finally, by surveying widely used algorithm families (regularized linear models, decision trees and ensembles, and SVMs), we highlighted how practical ML balances interpretability, flexibility, and computational cost. With these foundations in place, the next step is to apply the workflow to real datasets: formulating clear problem statements, selecting appropriate paradigms and metrics, rigorously validating models, and reporting results transparently so that findings can be reproduced and responsibly deployed.

Acknowledgements or Notes

Generative artificial intelligence tools were employed to support the conceptualization of this book chapter, assist in the literature review, contribute to the drafting process, and facilitate language editing.

References

Abdulhafedh, A. (2022). *Machine learning models and regularization techniques*. Springer

Al-Hamadani, M., Fadhel, M., Alzubaidi, L., & Harangi, B. (2024). Reinforcement learning algorithms and

applications in healthcare and robotics: A comprehensive and systematic review. *Sensors*, *24*(8). https://doi.org/10.3390/s24082461

Alpaydin, E. (2021). *Machine learning* (4th ed.). MIT Press. https://doi.org/10.7551/mitpress/13811.001.0001

Badillo, S., Bánfai, B., Birzele, F., Davydov, I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., & Zhang, J. (2020). An introduction to machine learning. *Clinical Pharmacology & Therapeutics*, *107*(4), 871–885. https://doi.org/10.1002/cpt.1796

Baladram, M., Koike, A., & Yamada, K. (2020). Introduction to Supervised Machine Learning for Data Science. *Interdisciplinary Information Sciences*. https://doi.org/10.4036/iis.2020.a.03

Basu, S. (2019). Semi-Supervised Learning, 2613-2615. https://doi.org/10.1007/978-0-387-39940-9_609

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798–1828. https://doi.org/10.1109/TPAMI.2013.50

Bernardes, R. (2024). Machine learning: Basic principles. *Acta Ophthalmologica*. https://doi.org/10.1111/aos.16281

Burkart, N., & Huber, M. (2020). A Survey on the Explainability of Supervised Machine Learning. *ArXiv*, *abs/2011.07876*. https://doi.org/10.1613/jair.1.12228

Cacciarelli, D., & Kulahci, M. (2025). Semi-supervised learning for predictive modeling in industrial applications. *Quality Engineering*, 37, 339 - 346. https://doi.org/10.1080/08982112.2024.2440371

Chai, C., Wang, J., Luo, Y., Niu, Z., & Li, G. (2023). Data Management for Machine Learning: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 35, 4646-4667. https://doi.org/10.1109/tkde.2022.3148237

Chen, Y., Mancini, M., Zhu, X., & Akata, Z. (2022). Semi-Supervised and Unsupervised Deep Visual Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46, 1327-1347. https://doi.org/10.1109/tpami.2022.3201576

Cohen, S. (2021). The basics of machine learning: strategies and techniques. *Artificial Intelligence and Deep Learning in Pathology*. https://doi.org/10.1016/b978-0-323-67538-3.00002-6

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, *10*(7), 1895–1923.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78–87. https://doi.org/10.1145/2347736.2347755

Du, K.-L., Zhang, R., Jiang, B., Zeng, J., & Lu, J. (2025). Understanding Machine Learning Principles: Learning, Inference, Generalization, and Computational Learning Theory. *Mathematics*. https://doi.org/10.3390/math13030451

Eckhardt, C., Madjarova, S., Williams, R., Ollivier, M., Karlsson, J., Pareek, A., & Nwachukwu, B. (2022). Unsupervised machine learning methods and emerging applications in healthcare. *Knee Surgery, Sports Traumatology, Arthroscopy*, 31, 376-381. https://doi.org/10.1007/s00167-022-07233-7

Edwards, A., Kaplan, B., & Jie, T. (2020). A Primer on Machine Learning. *Transplantation*. https://doi.org/10.1097/tp.0000000000003316

Eldeeb, H., & Elshawi, R. (2025). Empowering Machine Learning With Scalable Feature Engineering and Interpretable AutoML. *IEEE Transactions on Artificial Intelligence*, 6, 432-447. https://doi.org/10.1109/tai.2024.3400752

Eppa, A. (2025). Machine Learning Algorithms: A Comparative Study on Efficiency, Effectiveness, and Practical Applications Using the GRA Method. *Journal of Artificial intelligence and Machine Learning*. https://doi.org/10.55124/jaim.v3i1.260

Faouzi, J., & Colliot, O. (2023). Classical machine learning algorithms revisited. *Pattern Recognition Letters*, *164*, 25–34.

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2019). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, *15*, 3133–3181.

François-Lavet, V., Henderson, P., Islam, R., Bellemare, M., & Pineau, J. (2018). An Introduction to Deep Reinforcement Learning. *Found. Trends Mach. Learn.*, 11, 219-354. https://doi.org/10.1561/2200000071

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, *4*(1), 1–58.

Ghildiyal, S., et al. (2024). Comparative evaluation of classical machine learning algorithms. *Applied Artificial Intelligence*.

Golazad, S., Mohammadi, A., Rashidi, A., & Ilbeigi, M. (2024). From raw to refined: Data preprocessing for construction machine learning (ML), deep learning (DL), and reinforcement learning (RL) models. *Automation in Construction*. https://doi.org/10.1016/j.autcon.2024.105844

González, S., et al. (2020). Ensemble learning: A survey. *Information Fusion*, *61*, 1–15.

Gosavi, A. (2009). Reinforcement Learning: A Tutorial Survey and Recent Advances. *INFORMS J. Comput.*, 21, 178-192. https://doi.org/10.1287/ijoc.1080.0305

Gui, J., Chen, T., Zhang, J., Cao, Q., Sun, Z., Luo, H., & Tao, D. (2023). A Survey on Self-Supervised Learning: Algorithms, Applications, and Future Trends. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*, 9052–9071. https://doi.org/10.1109/tpami.2024.3415112

Hinton, G., & Sejnowski, T. (2018). *Unsupervised Learning*. 1009. https://doi.org/10.1007/978-3-319-17885-1_101437

Hodeghatta, U., & Nayak, U. (2017). Unsupervised Machine Learning, 161-186. https://doi.org/10.1007/978-1-4842-2514-1_7

Hsu, D. (2022). Overview of Machine Learning. *International Journal of Advanced Research in Science, Communication and Technology*. https://doi.org/10.48175/ijarsct-4844

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*, 685–695. https://doi.org/10.1007/s12525-021-00475-2

Jiang, T., Gradus, J., & Rosellini, A. (2020). Supervised Machine Learning: A Brief Primer. *Behavior Therapy*,

51 5, 675–687. https://doi.org/10.1016/j.beth.2020.05.002

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, *4*, 237–285. https://doi.org/10.1613/jair.301

Karthick, K. (2024). Comprehensive Overview of Optimization Techniques in Machine Learning Training. *Control Systems and Optimization Letters*. https://doi.org/10.59247/csol.v2i1.69

Ko, E., & Kang, P. (2025). Evaluating Coding Proficiency of Large Language Models: An Investigation Through Machine Learning Problems. *IEEE Access*, 13, 52925-52938. https://doi.org/10.1109/access.2025.3553870

Kubát, M. (2017). *An Introduction to Machine Learning*. 1–348. https://doi.org/10.1007/978-3-319-63913-0

Kumar, A. (2024). Machine learning algorithms: A comparative study. *Journal of Artificial Intelligence Research*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444. https://doi.org/10.1038/nature14539

Li, H. (2024). Machine Learning Methods. *Machine Learning Methods*. https://doi.org/10.1007/978-981-99-3917-6

Li, J. (2019). Regression and Classification in Supervised Learning. *Proceedings of the 2nd International Conference on Computing and Big Data*. https://doi.org/10.1145/3366650.3366675

Li, S., Kou, P., , M., Yang, H., Huang, S., & Yang, Z. (2024). Application of Semi-Supervised Learning in Image Classification: Research on Fusion of Labeled and Unlabeled Data. *IEEE Access*, 12, 27331-27343. https://doi.org/10.1109/access.2024.3367772

Li, Y., Guo, L., & Zhou, Z. (2021). Towards Safe Weakly Supervised Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43, 334-346. https://doi.org/10.1109/tpami.2019.2922396

Littman, M. (2015). Reinforcement learning improves behaviour from evaluative feedback. *Nature*, 521, 445-451. https://doi.org/10.1038/nature14540

Lo Vercio, L., Amador, K., Bannister, J., Crites, S., Gutierrez, A., MacDonald, M. E., Moore, J., Mouches, P., Rajasheka, D., Schimert, S., Subbanna, N., Tuladhar, A., Wang, N., Wilms, M., Winder, A., & Forkert, N. (2020). Supervised machine learning tools: a tutorial for clinicians. *Journal of Neural Engineering*, *17*. https://doi.org/10.1088/1741-2552/abbff2

Lu, J., Guangzhi, & Zhang, G. (2024). Fuzzy Machine Learning: A Comprehensive Framework and Systematic Review. *IEEE Transactions on Fuzzy Systems*, *32*, 3861–3878. https://doi.org/10.1109/tfuzz.2024.3387429

Martínez-Heredia, A., & Ventura, S. (2025). Weak Supervision: A Survey on Predictive Maintenance. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 15. https://doi.org/10.1002/widm.70022

Mey, A., & Loog, M. (2022). Improved Generalization in Semi-Supervised Learning: A Survey of Theoretical Results. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45, 4747-4767. https://doi.org/10.1109/tpami.2022.3198175

Mienye, I. D., & Jere, S. (2023). Decision trees and ensemble learning methods. *Applied Soft Computing*.

Mienye, I., & Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access*, 10, 99129-99149. https://doi.org/10.1109/access.2022.3207287

Misera, L., Müller-Franzes, G., Truhn, D., & Kather, J. (2024). Weakly Supervised Deep Learning in Radiology.. *Radiology*, 312 1, e232085. https://doi.org/10.1148/radiol.232085

Mumuni, A., & Mumuni, F. (2024). Automated data processing and feature engineering for deep learning and big data applications: a survey. *ArXiv*, abs/2403.11395. https://doi.org/10.1016/j.jiixd.2024.01.002

Naeem, M., Rizvi, S., & Coronato, A. (2020). A Gentle Introduction to Reinforcement Learning and its Application in Different Fields. *IEEE Access*, 8, 209320-209344. https://doi.org/10.1109/access.2020.3038605

Naeem, S., Ali, A., Anam, S., & Ahmed, M. (2023). An Unsupervised Machine Learning Algorithms: Comprehensive Review. *International Journal of Computing and Digital Systems*. https://doi.org/10.12785/ijcds/130172

Nartey, O., Yang, G., Wu, J., & Asare, S. (2020). Semi-Supervised Learning for Fine-Grained Classification With Self-Training. *IEEE Access*, 8, 2109-2121. https://doi.org/10.1109/access.2019.2962258

Nian, R., Liu, J., & Huang, B. (2020). A review On reinforcement learning: Introduction and applications in industrial process control. *Comput. Chem. Eng.*, 139, 106886. https://doi.org/10.1016/j.compchemeng.2020.106886

Obaido, G., Mienye, I. D., Egbelowo, O., Emmanuel, I. D., Ogunleye, A., Ogbuokiri, B., Mienye, P., & Aruleba, K. (2024). Supervised machine learning in drug discovery and development: Algorithms, applications, challenges, and prospects. *Machine Learning with Applications*. https://doi.org/10.1016/j.mlwa.2024.100576

Pantanowitz, L., Pearce, T., Abukhiran, I., Hanna, M., Wheeler, S., Soong, T., Tafti, A., Pantanowitz, J., Lu, M., Mahmood, F., Gu, Q., & Rashidi, H. (2024). Non-Generative Artificial Intelligence (AI) in Medicine: Advancements and Applications in Supervised and Unsupervised Machine Learning.. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 100680. https://doi.org/10.1016/j.modpat.2024.100680

Pargent, F., Schoedel, R., & Stachl, C. (2023). Best Practices in Supervised Machine Learning: A Tutorial for Psychologists. *Advances in Methods and Practices in Psychological Science*, *6*. https://doi.org/10.1177/25152459231162559

Pfob, A., Lu, S., & Sidey-Gibbons, C. (2022). Machine learning in medicine: a practical introduction to techniques for data pre-processing, hyperparameter tuning, and model comparison. *BMC Medical Research Methodology*, 22. https://doi.org/10.1186/s12874-022-01758-8

Pienaar, M., Paed, S. C. C., Naidoo, P., & BCh, D. Fcp. S. M. (2025). Classification and predictive models using

supervised machine learning: A conceptual review. *Southern African Journal of Critical Care*, *41*. https://doi.org/10.7196/sajcc.2025.v411.2937

Puppala, A. (2025). A Comprehensive Overview of Machine Learning Models: Techniques, Applications, and Future Directions. *International Journal For Multidisciplinary Research*. https://doi.org/10.36948/ijfmr.2025.v07i02.40603

Qi, G., & Luo, J. (2019). Small Data Challenges in Big Data Era: A Survey of Recent Progress on Unsupervised and Semi-Supervised Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44, 2168-2187. https://doi.org/10.1109/tpami.2020.3031898

Raghhul, O. (2025). Data Insights to Machine Learning Model. *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2025.70343

Rahaman, M. J. (2024). A Comprehensive Review to Understand the Definitions, Advantages, Disadvantages and Applications of Machine Learning Algorithms. *International Journal of Computer Applications*. https://doi.org/10.5120/ijca2024923868

Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, *14*. https://doi.org/10.1038/s41598-024-56706-x

Rao, M., Saikrishna, G., & Supriya, K. (2023). Data preprocessing techniques: emergence and selection towards machine learning models - a practical review using HPA dataset. *Multimedia Tools and Applications*, 1-20. https://doi.org/10.1007/s11042-023-15087-5

Ren, Z., Wang, S., & Zhang, Y. (2023). Weakly supervised machine learning. *CAAI Trans. Intell. Technol.*, 8, 549-580. https://doi.org/10.1049/cit2.12216

Sen, P., Hajra, M., & Ghosh, M. (2019). Supervised Classification Algorithms in Machine Learning: A Survey and Review. *Advances in Intelligent Systems and Computing*. https://doi.org/10.1007/978-981-13-7403-6_11

Sewak, M. (2020). Introduction to Reinforcement Learning. *Deep Reinforcement Learning with Python*. https://doi.org/10.1007/978-981-13-8285-7_1

Shahidi, S., Samadzai, A., & Shahbazi, H. (2025). Effective Data Preprocessing in Data Science: From Method Selection to Domain-Specific Optimization. *Journal of Advanced Computer Knowledge and Algorithms*. https://doi.org/10.29103/jacka.v2i4.22886 .

Shakya, A., Pillai, G., & Chakrabarty, S. (2023). Reinforcement learning algorithms: A brief survey. *Expert Syst. Appl.*, 231, 120495. https://doi.org/10.1016/j.eswa.2023.120495

Sharma, R. (2020). Study of Supervised Learning and Unsupervised Learning. *International Journal for Research in Applied Science and Engineering Technology*, 8, 588-593. https://doi.org/10.22214/ijraset.2020.6095

Shetty, S., Shetty, S., Singh, C., & Rao, A. S. (2022). Supervised Machine Learning: Algorithms and Applications. *Fundamentals and Methods of Machine and Deep Learning*. https://doi.org/10.1002/9781119821908.ch1

Simeone, O. (2017). A Brief Introduction to Machine Learning for Engineers. *ArXiv*, *abs/1709.0*. https://doi.org/10.1561/2000000102

Simeone, O. (2018). A Very Brief Introduction to Machine Learning With Applications to Communication Systems. *IEEE Transactions on Cognitive Communications and Networking*, 4, 648-664. https://doi.org/10.1109/tccn.2018.2881442

Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 1310–1315. https://consensus.app/papers/a-review-of-supervised-machine-learning-algorithms-singh-thakur/75420cba25b7554da6cac855ec64a985/

Sivamayil, K., Rajasekar, E., Aljafari, B., Nikolovski, S., Vairavasundaram, S., & Vairavasundaram, I. (2023). A Systematic Study on Reinforcement Learning Based Applications. *Energies*. https://doi.org/10.3390/en16031512

Song, Z., Yang, X., Xu, Z., & King, I. (2021). Graph-Based Semi-Supervised Learning: A Comprehensive Review. *IEEE Transactions on Neural Networks and Learning Systems*, 34, 8174-8194. https://doi.org/10.1109/tnnls.2022.3155478

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*, *36*(2), 111–147.

Sun, L., Ye, P., Lyu, G., Feng, S., Dai, G., & Zhang, H. (2020). Weakly-supervised multi-label learning with noisy features and incomplete labels. *Neurocomputing*, 413, 61-71. https://doi.org/10.1016/j.neucom.2020.06.101

Sutton, R., & Barto, A. (2005). Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*, 16, 285-286. https://doi.org/10.1109/tnn.1998.712192

Syed, I., & Lokhande, V. (2024). An overview of the supervised machine learning. *International Research Journal of Modernization in Engineering Technology and Science*. https://doi.org/10.56726/irjmets51366

Tate, T., Sugiyama, K., & Uchida, M. (2025). Learning from Similarity-Confidence and Confidence-Difference. *ArXiv*, abs/2508.05108. https://doi.org/10.48550/arxiv.2508.05108

Usmani, U. A., Happonen, A., & Watada, J. (2022). *A Review of Unsupervised Machine Learning Frameworks for Anomaly Detection in Industrial Applications*. 158–189. https://doi.org/10.1007/978-3-031-10464-0_11

Van Engelen, J., & Hoos, H. (2019). A survey on semi-supervised learning. *Machine Learning*, 109, 373-440. https://doi.org/10.1007/s10994-019-05855-6

Wang, S. (2025). A Comprehensive Review of Machine Learning and Data Science Applications: From Fundamentals to Advanced Techniques. *Science and Technology of Engineering, Chemistry and Environmental Protection*. https://doi.org/10.61173/bhqfg820

Watson, D. (2023). On the Philosophy of Unsupervised Learning. *Philosophy & Technology*, 36, 1-26. https://doi.org/10.1007/s13347-023-00635-6

Wong, P. (2021). Unsupervised Machine Learning. *Methodology of Educational Measurement and Assessment*. https://doi.org/10.1007/978-3-030-74394-9_10

Wörgötter, F., & Porr, B. (2019). Reinforcement learning. *Astron. Comput.*, 48, 100833. https://doi.org/10.4249/scholarpedia.1448

Xiao, B., Tang, Y., & Liu, Y. (2025). Integrating Materials Representations Into Feature Engineering in Machine Learning for Crystalline Materials: From Local to Global Chemistry-Structure Information Coupling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 15. https://doi.org/10.1002/wcms.70044

Xiao, M., Wang, D., Wu, M., Liu, K., Xiong, H., Zhou, Y., & Fu, Y. (2023). Traceable Group-Wise Self-Optimizing Feature Transformation Learning: A Dual Optimization Perspective. *ACM Transactions on Knowledge Discovery from Data*, 18, 1 - 22. https://doi.org/10.1145/3638059

Yang, X., Song, Z., King, I., & Xu, Z. (2021). A Survey on Deep Semi-Supervised Learning. *IEEE Transactions on Knowledge and Data Engineering*, 35, 8934-8954. https://doi.org/10.1109/tkde.2022.3220219

Yasodha, P. (2025). Data Preprocessing Methods for Machine Learning: An Empirical Comparison. *International Journal For Multidisciplinary Research*. https://doi.org/10.36948/ijfmr.2025.v07i03.48569

Yu, C., Liu, J., & Nemati, S. (2019). Reinforcement Learning in Healthcare: A Survey. *ACM Computing Surveys (CSUR)*, 55, 1 - 36. https://doi.org/10.1145/3477600

Zhou, M.-L., & Xu, H. (2025). Learning rich features without labels: contrastive approaches in multimodal artificial intelligence systems. *International Journal of Advanced Artificial Intelligence Research*. https://doi.org/10.55640/ijaair-v02i04-02

Zhou, Z.-H. (2018). A brief introduction to weakly supervised learning. *National Science Review*, *5*, 44–53. https://doi.org/10.1093/nsr/nwx106

Zhou, Z., Wang, J., Wei, T., & Zhang, M. (2025). Weakly-Supervised Contrastive Learning for Imprecise Class Labels. *ArXiv*, abs/2505.22028. https://doi.org/10.48550/arxiv.2505.22028

Zhu, X., & Goldberg, A. (2009). Introduction to Semi-Supervised Learning. https://doi.org/10.1007/978-3-031-01548-9

Author Information

Hamza Polat

https://orcid.org/0000-0002-9646-7507

Atatürk University, Erzurum, Türkiye

Contact e-mail: *hamzapolat@atauni.edu.tr*

Citation

Polat, H., (2025). Machine Learning. In A. Kaban, & T. Demirel (Eds.), *Interdisciplinary Applications of Artificial Intelligence - 1* (pp. 59-82). ISTES.

Chapter 5- Artificial Intelligence in Research Processes

Turgay Demirel

Chapter Highlights

- Generative Artificial Intelligence (GenAI) serves as a "cognitive partner" in research, augmenting human capabilities across the research lifecycle.
- Artificial Intelligence (AI) supports idea generation, literature reviews, data analysis, academic writing, and publication strategies.
- Its key capabilities include text generation, summarization, cross-language support, and complex reasoning over documents.
- Responsible AI adoption necessitates transparency, human oversight, and rigorous verification to mitigate risks such as hallucinations and algorithmic biases.
- Researchers must remain the primary decision-makers and interpreters, treating AI as a co-pilot rather than an independent author.
- Institutional guidelines and AI literacy training are essential for the ethical and effective integration of AI into research workflows.
- Future directions include multimodal models, agentic workflows, discipline-specific fine-tuned models, and evolving best practices.

1. Introduction

1.1. AI and the Changing Nature of Research

Artificial intelligence is an interdisciplinary field that develops computational systems capable of emulating aspects of human intelligence across tasks such as learning, problem-solving, perception, speech recognition, and decision-making (Luckin & Holmes, 2016; Zawacki-Richter et al., 2019). It encompasses key technologies, including machine learning, natural language processing, and artificial neural networks (Zawacki-Richter et al., 2019).

Generative AI, a pivotal subset of AI, utilizes deep learning models, often transformers, to generate human-like content, including text, images, audio, or video based on user prompts (Jovanovic & Campbell, 2022; Mutelo, 2025). In contrast to traditional AI, which relies on fixed rules or patterns, GenAI produces novel and creative outputs drawn from learned data distributions (Polat, 2023), positioning it as a major transformative force across diverse fields (Panda & Kaur, 2024; Polat, 2023).

At the core of modern generative AI are large language models built on transformer architectures and trained on massive text datasets (Carchiolo & Malgeri, 2025; Minaee et al., 2024). These models excel in processing natural language for tasks such as text completion, question answering, translation, and summarization (Carchiolo & Malgeri, 2025; Ooi et al., 2023). Within academic research, LLMs have emerged as "cognitive partners", amplifying human creativity, refining problem formulation, and deepening interpretation in areas such as literature reviews, data analysis, and writing (Katsamakas et al., 2024).

Artificial intelligence is transforming scholarly work across research stages as a cognitive partner that augments human capabilities rather than merely automating tasks. This shift is evident in literature reviews, data analyses, and academic writing, providing increased research efficiency and new research opportunities (Polat, 2023; Wu et al., 2024).

In the literature review, AI streamlines processes by generating precise keywords for efficient search and retrieval, while also facilitating cross-language research (Pisica et al., 2023). It enhances screening, summarization, and synthesis by enabling interactive document analysis and automated extraction of key scholarly information from large bodies of literature (Ngwenyama & Rowe, 2024). AI can also map literature and identify trends; however, maintaining critical distance is crucial to avoid shallow summaries or hallucinated citations, thereby necessitating careful verification (Ngwenyama & Rowe, 2024).

In data analysis, AI transforms the handling of large datasets by enabling more efficient and scalable analytical processes (Ekundayo et al., 2024). For qualitative data, AI supports coding and thematic analysis by suggesting initial codes and structures (Perkins & Roe, 2024), while researchers retain primary interpretive responsibility (Ngwenyama & Rowe, 2024). In quantitative analysis, AI supports data cleaning, exploratory analysis, regression, and forecasting (George & Wooden, 2023). Large language models may serve as "statistical assistants" (Ke &

Ribeiro, 2025), but researchers must be aware of the underlying assumptions that may not be immediately obvious (Ekundayo et al., 2024).

AI tools provide comprehensive support for academic writing. AI tools assist in outlining, structuring, and drafting specific sections, and also facilitate research question formulation and hypothesis generation (Katsamakas et al., 2024; Xu, 2025). Furthermore, AI enhances language and style, providing cross-language support that particularly benefits non-native English speakers (Shimray & Subaveerapandiyan, 2025). Although AI assists with citation management, cross-checking references remains essential because of potential fabrication (Katsamakas et al., 2024). Furthermore, it aids in plagiarism and AI detection (Shimray & Subaveerapandiyan, 2025). However, researchers must be aware of its limitations and prioritize originality and transparency in AI usage. AI can also simulate reviewer feedback, thereby enhancing writing proficiency (Lo et al., 2025; Luttges et al., 2025). Overall, AI functions as a valuable assistant, augmenting research efficiency and effectiveness rather than replacing human expertise and creativity; researchers must remain the primary decision-makers and interpreters (George & Wooden, 2023).

1.2. Purpose and Scope of the Chapter

This chapter provides a comprehensive exploration of the role of artificial intelligence in academic research, tracing its impact and potential applications across the entire research lifecycle, from initial idea generation to the final publication stages. This study aims to clarify how AI tools can serve as responsible digital support systems that augment human capabilities at each critical juncture. The central research question is: How can AI tools provide responsible digital support throughout the research process?

1.3. Structure of the Chapter

This chapter begins by examining how AI, particularly generative AI and large language models, fundamentally reshapes the nature of research. The chapter distinguishes between traditional AI automation and the more nuanced concept of generative AI as a cognitive partner, emphasizing the capacity of generative AI to facilitate creativity, problem framing, and interpretation. Subsequent sections delve into specific phases of research, starting with idea generation and the formulation of well-defined research questions, followed by AI-supported literature review workflows. The chapter then moves into the design of methodologies and data collection, demonstrating how AI can assist in developing instruments and gathering data. A significant portion is dedicated to AI's role in data analysis and visualization, including qualitative coding, quantitative analysis, and advanced modeling, while highlighting the importance of identifying potential "hidden assumptions" within the AI outputs. The narrative progresses to AI-supported academic writing and quality control, covering outlining, drafting, language refinement, citation management, and crucial aspects of plagiarism detection and originality issues. Finally, the chapter addresses AI in publication strategy and post-publication dissemination, before concluding with a thorough discussion of the ethical, legal, and integrity considerations inherent in using AI in research and the necessity of developing AI literacy and institutional guidelines for its responsible adoption by researchers.

2. Generative AI as a Cognitive Partner in Research

2.1. From Automation to Augmentation

Generative AI represents a significant shift from traditional AI automation (Polat, 2023), by moving beyond basic task execution to function as a cognitive partner in research. Unlike conventional AI tools, such as statistical packages that automate specific predetermined processes, generative AI offers advanced support. This advanced AI paradigm focuses on augmenting human capabilities (Fecher et al., 2023), particularly by fostering creativity, assisting in problem framing, and enhancing interpretation (Wu et al., 2024). Through collaboration with researchers, generative AI can stimulate innovative thinking, support the conceptualization of complex problems, and deepen analytical insight by processing and synthesizing large-scale information (Minaee et al., 2024; Ngwenyama & Rowe, 2024), thereby enhancing the intellectual contributions of human researchers (Perkins & Roe, 2024).

2.2. Key Capabilities Relevant to Researchers

Generative AI offers researchers a suite of key capabilities that significantly enhance various stages of academic research. At its core, it excels in advanced text generation, enabling the drafting of specific manuscript sections, refining language, and supporting overall structuring (Katsamakas et al., 2024; Xu, 2025). Furthermore, AI tools provide superior summarization and synthesis, efficiently distilling of vast amounts of literature and extraction of key information for rapid insight generation. Cross-language support, including seamless translation, further democratizes access, particularly for non-native English speakers (Pisica et al., 2023). In data handling, AI delivers coding support for qualitative analysis by suggesting initial codes and thematic structures (Perkins & Roe, 2024) and extending to sophisticated information retrieval, such as generating precise keywords for targeted searches (Pisica et al., 2023) and complex reasoning over documents, which aids in problem framing and interpretation. This multifaceted support is powered by a burgeoning ecosystem of specialized platforms, including ChatGPT, SciSpace, Avidnote, Jenni, and Paperpal (Bolaños et al., 2024; Zhou et al., 2025).

2.3. Opportunities and Risks

Generative AI serves as a cognitive partner in research, augmenting human capabilities by automating routine tasks to boost efficiency, broadening access to scientific resources for wider participation, and enabling interdisciplinary synthesis through novel cross-domain connections (Boyko et al., 2023; Fecher et al., 2023). These capabilities allow researchers to dedicate more cognitive resources to creative synthesis, contextualization, and innovation, thereby accelerating the scholarly workflow (Fecher et al., 2023; Katsamakas et al., 2024).

However, this partnership also entails substantial risks, including hallucinations that fabricate plausible yet erroneous content, algorithmic biases perpetuated from training data, overreliance that erodes critical skills, and data privacy vulnerabilities from cloud-based processing (Bjelobaba et al., 2024; Ooi et al., 2023). Therefore, responsible use requires thorough human verification of outputs, vigilance in detecting flaws, and sustained

critical thinking to maintain academic integrity. This positions GenAI as a complement to, rather than a substitute for human judgment (Binz et al., 2023; Perkins & Roe, 2024). Ultimately, this balanced approach aligns with contemporary debates on AI-assisted epistemology, ensuring advances in research ethics and reliability (Boyko et al., 2023; Fecher et al., 2023).

3. From Research Idea to Well-Formulated Questions

3.1. Topic Exploration and Idea Generation

Generative artificial intelligence supports researchers during the early ideation phase by facilitating the exploration of broad thematic landscapes within a discipline, identifying emerging trends, and mapping conceptual gaps through the interactive prompting of large language models (Katsamakas et al., 2024; Nigam et al., 2024). For instance, by prompting LLMs to suggest high-impact, novel, or underexplored research directions—such as “generate topic ideas relating to AI ethics while highlighting underexplored gaps”—researchers can systematically brainstorm options and iteratively refine them into focused and feasible topics (Katsamakas et al., 2024; Lund et al., 2024). Tools such as SciSpace's "Find Topics" feature exemplify such exploration by using clustering to generate interactive topic maps that visualize global and specific themes from the literature, thereby aiding the detection of patterns and underexplored areas (Bolaños et al., 2024). These methods leverage LLMs to detect patterns across vast literature, suggest hypotheses, and reframe problems from identified gaps, thus accelerating the transition from broad exploration to precise questioning (Nigam et al., 2024). However, the success of this process hinges on human judgment and expertise; researchers must critically assess AI outputs for relevance, feasibility, and originality, refining suggestions by validating them against scholarly standards and personal knowledge to mitigate risks such as superficiality or hallucinations (Ngwenyama & Rowe, 2024; Perkins & Roe, 2024). Ultimately, this human-AI collaboration ensures rigorous idea development, aligning with the current emphasis on augmented rather than automated scholarship (Doshi et al., 2025).

3.2. Identifying Research Gaps and Contradictions

Generative AI scans recent academic literature to identify research gaps, underexplored populations, neglected methodologies, and unresolved debates. It achieves this by processing vast text collections using advanced natural language processing and topic-modelling techniques (Bolaños et al., 2024; Schryen et al., 2025). Tools such as SciSpace enable researchers to extract stated limitations, future research directions, and methodological suggestions from multiple documents, automating the aggregation of fragmented evidence into interpretable summaries that highlight underexplored areas (Ge et al., 2024). Similarly, Consensus employs features such as the Consensus Meter to quantify agreement or disagreement across studies, revealing contradictions, lack of consensus, or inconsistent findings that signal opportunities for original scholarly contributions (Zhou et al., 2025). By grouping topics and highlighting inconsistencies, such as divergent results across specific populations or methods, AI-driven analyses accelerate literature comprehension by converting raw textual data into actionable insights for research agendas (Ngwenyama & Rowe, 2024; Nigam et al., 2024). Nonetheless, while AI accelerates the process, human expertise remains essential for critically interpreting outputs, validating them against domain

knowledge, and formulating strong research questions, thereby avoiding risks such as oversimplification or missed nuances (Ngwenyama & Rowe, 2024; Schryen et al., 2025). This hybrid approach aligns with ongoing debates on AI-enhanced systematic literature reviews, balancing efficiency and epistemic rigor (Bolaños et al., 2024; Ge et al., 2024).

3.3. Extracting Limitations and Future Research Directions

Generative artificial intelligence facilitates multi-document analysis of PDF articles by employing natural language processing techniques to simultaneously process multiple sources, extracting stated limitations, methodological constraints, and future research suggestions through automated information retrieval and synthesis (Bolaños et al., 2024). This approach identifies recurring themes across studies, such as common limitations or proposed directions, enabling researchers to construct systematic research agendas grounded in aggregated evidence (Bolaños et al., 2024; Schryen et al., 2025). For instance, SciSpace supports the automatic extraction of predefined elements, such as limitations from papers, allowing queries for specific content across documents (Bolaños et al., 2024). Avidnote complements this by facilitating chats with uploaded documents to derive insights on limitations and future work while also analyzing reference lists to propose new study ideas, titles, and abstracts, thus building long-term publishable pipelines (Shimray & Subaveerapandiyan, 2025). Thesify.ai functions as a digital supervisor for pre-submission drafts, detecting internal gaps and weaknesses while suggesting concrete next steps. Similarly, tools such as Jenni.ai enable the iterative refinement of writing outputs by incorporating extracted insights, while Paperpal aids in manuscript checks to highlight weaknesses and ensure compliance during the drafting of the Discussion sections (Pinzolits, 2023). These capabilities aid strategic planning by prioritizing original contributions, anticipating critiques, and promoting robust scholarship (Nigam et al., 2024; Schryen et al., 2025). Nonetheless, while AI accelerates extraction, human expertise remains essential for validating outputs against domain knowledge, interpreting context-specific nuances, and avoiding risks such as oversimplification or hallucinations (Ngwenyama & Rowe, 2024; Schryen et al., 2025), thus ensuring a balanced human-AI collaboration that upholds scholarly rigor.

3.4. Generating Tailored Research Questions and Hypotheses

Generative artificial intelligence supports researchers in formulating precise, theory-aligned research questions and testable hypotheses. It achieves this by synthesizing vast literature patterns, surfacing conceptual linkages, and generating novel, grounded propositions through interactive prompting and knowledge aggregation (Ngwenyama & Rowe, 2024; Zhou et al., 2025). Specialized tools further enhance these processes. For example, SciSpace's Find Topics feature clusters thematic inputs to propose methodological and conceptual directions, transforming broad themes into focused inquiries amenable to hypothesis testing (Bolaños et al., 2024; Jain et al., 2024). Similarly, Consensus uses its Consensus Meter to highlight scientific disagreements and evidential fragmentation, enabling the framing of original questions around unresolved tensions for novel contributions (Zhou et al., 2025). Avidnote facilitates tailored questions and hypothesis generation from user-specified study aims via document chat, serving as a brainstorming adjunct (Shimray & Subaveerapandiyan, 2025). In addition, general-purpose tools such as ChatGPT, when prompted to act as a mentor with domain-specific constraints,

iteratively refine broad ideas into rigorous, publishable formats (Rahman & Watanobe, 2023; Rahman et al., 2023). Crucially, effective AI use is iterative; researchers must refine outputs using domain expertise, theory checks, and feasibility tests to ensure quality and avoid superficiality (Ngwenyama & Rowe, 2024; Schryen et al., 2025). This symbiotic paradigm upholds human epistemic oversight in the face of accelerating ideation (Delios et al., 2024; Katsamakas et al., 2024).

4. AI-Supported Literature Review Workflows

4.1. Efficient Literature Search and Retrieval

Generative artificial intelligence, particularly large language models, supports the systematic generation of precise keywords, Boolean search strings, and synonymous or related terms by leveraging natural language processing to semantically expand the queries. This enables comprehensive and reproducible literature search strategies that surpass rigid keyword limitations (Bolaños et al., 2024; Wang et al., 2023). For instance, these models facilitate hybrid keyword expansion through topic modeling and co-occurrence analysis, refining initial terms into structured Boolean combinations for targeted retrieval (Feng et al., 2023). Academic-focused search environments further enhance this efficiency. Perplexity AI's Academic Mode filters results to scholarly sources, processing complex question-based queries with structured, citation-linked responses (Bolaños et al., 2024; Xu & Peng, 2025). Leading platforms exemplify this capability: SciSpace permits full-sentence querying and multi-document chat for literature analysis and gap identification (Bolaños et al., 2024; Jain et al., 2024), whereas Consensus synthesizes evidence using its Consensus Meter, gauging agreement or disagreement across studies (Bolaños et al., 2024). Semantic Scholar enables natural language searches over millions of papers with field-specific filtering and citation tracking (Bolaños et al., 2024; Kinney et al., 2023), complemented by OpenAlex for trend analysis and open-access discovery through resource description framework (RDF) knowledge graphs (Färber et al., 2023a, 2023b). Finally, literature mapping tools, such as Research Rabbit, LitMaps, and Connected Papers, visualize citation graphs and conceptual interconnections from seed papers, accelerating evidence synthesis (Cole & Boutet, 2023; Pinzolits, 2023).

4.2. Screening, Summarization, and Synthesis

Generative artificial intelligence has revolutionized literature reviews by enabling interactive document analysis through "chat with PDF" workflows, wherein researchers query single or multiple uploaded documents to extract targeted insights via natural language interfaces (Bolaños et al., 2024; Jain et al., 2024). SciSpace exemplifies this capability with its multi-document chat feature and column-based filtering, allowing simultaneous examination of papers for key constructs, methodologies, principal findings, and limitations, thus streamlining screening without requiring exhaustive manual reading (Bolaños et al., 2024; Jain et al., 2024). Avidnote complements this by supporting interactive PDF chats to derive structured data, study contributions and qualitative thematic synthesis (Bolaños et al., 2024; Shimray & Subaveerapandiyan, 2025). These systems automate document extraction using retrieval-augmented generation and embeddings, accelerating comparative evidence aggregation (Bolaños et al., 2024; Ngwenyama & Rowe, 2024).

For initial screening and filtering, several AI tools offer distinct capabilities. Perplexity AI's Academic Mode restricts outputs to scholarly sources via complex queries. Consensus AI applies rigorous filters by journal impact, citations, dates, and methods such as RCTs (Bolaños et al., 2024; Musleh & AlRyalat, 2025). OpenAlex refines title/abstract keywords, trends, and open-access status (Bolaños et al., 2024). In summarization and synthesis, multi-document querying reveals consensus via Consensus Meter assessments, uncovers contradictions, and pinpoints gaps, fostering rigorous workflows (Bolaños et al., 2024; Schryen et al., 2025).

4.3. Mapping the Literature and Identifying Trends

Generative artificial intelligence supports literature mapping, topic clustering, and research trend identification by leveraging citation networks, semantic embeddings, and graph-based visualizations. These techniques cluster topics and delineate thematic relationships across scholarly corpora (Bolaños et al., 2024; Pinzolits, 2023). Tools such as ResearchRabbit produce interactive maps that distinguish established sources from suggested expansions by incorporating author networks and chronological evolution. Meanwhile, LitMaps and Connected Papers employ co-citation and bibliographic coupling to reveal conceptual structures, influential prior works, and derivative studies, thereby elucidating domain evolution and gaps (Cole & Boutet, 2023; Pinzolits, 2023). SciSpace augments this through multi-document querying and tabular extractions, facilitating thematic clustering through a comparative analysis of the methods and findings (Bolaños et al., 2024; Jain et al., 2024).

These platforms also identify temporal trends and hot topics through citation growth rates and evidence analysis. For example, OpenAlex offers detailed data on scientific publications and annual citation trends to facilitate trend analysis (Gu & Krenn, 2025), Consensus's Consensus Meter highlights scientific debates via agreement/disagreement metrics (Bolaños et al., 2024), and Elicit reveals conflicting evidence through a semantic search (Bolaños et al., 2024). Complementing these capabilities, SciSpace and Avidnote enable the automated extraction and comparison of methods and datasets, thereby identifying methodological innovations such as novel paradigms in recent studies (Jain et al., 2024; Shimray & Subaveerapandiyan, 2025).

4.4. Maintaining Critical Distance

Maintaining a critical distance when integrating generative artificial intelligence into research processes is essential to preserve academic integrity and epistemic rigor amid the risks of hallucinations, fabricated citations, and superficial outputs that undermine scholarly reliability (Helmy et al., 2025; Sallam, 2023; Vicent et al., 2025). Generative models frequently produce plausible yet erroneous references, such as nonexistent sources, comprising up to 55% of ChatGPT-3.5-generated bibliographies and 18% of ChatGPT-4-generated bibliographies (Walters & Wilder, 2023). Overreliance on these outputs fosters cognitive disengagement, eroding researchers' critical judgment and interpretive agency (Mezzadri, 2025; Sarkar et al., 2024). This necessitates a paradigm shift from viewing AI as a ghostwriter to a co-pilot or cognitive tutor, wherein human epistemic responsibility remains paramount through iterative oversight and validation (Dgebuadze & Kenkebashvili, 2025; Eacersall et al., 2024). Basic strategies include cross-checking claims against primary sources, triangulating outputs across tools, and

direct verification via Consensus's Consensus Meter for evidential agreement, Perplexity AI's Academic Mode for traceable citations, and SciSpace or Avidnote's document chats to confirm methods and datasets (Christou, 2023; Fardim et al., 2023). Pre-submission assessments using AI detection and plagiarism screeners, such as Turnitin, alongside supervisory feedback tools, anticipate reviewer scrutiny and bolster the originality (Cordero et al., 2025; Teremetskyi et al., 2024). In essence, researcher supervision guarantees that AI enhances rather than replaces scholarly judgment (Eacersall et al., 2024; Ngwenyama & Rowe, 2024).

5. Designing Methodology and Data Collection with AI Support

5.1. Methodological Frameworks and Research Designs

Generative artificial intelligence serves as an intelligent methodological brainstorming assistant. It semantically analyzes research questions and study aims to propose suitable research designs, such as experimental, mixed-methods, or case studies, drawing on patterns from literature gaps and evidential structures. This enhances alignment and innovation without supplanting human decision-making (Delios et al., 2024; Katsamakas et al., 2024; Yuan & Harun, 2024). For instance, Avidnote's Suggest Methods feature generates tailored methodologies (e.g., qualitative coding, grounded theory, or thematic analysis) and theoretical frameworks with explanatory applications based on inputted objectives (Shimray & Subaveerapandiyan, 2025). Similarly, SciSpace identifies methodological limitations and underexplored areas across uploaded papers to inform novel designs (Jain et al., 2024). The alignment between methodology and theoretical frameworks is further advanced through large language model prompts configured as mentors or role assignments. This suggests candidate theories, validates prior applications in the literature, and iteratively refines coherence (Fecher et al., 2023; Walter, 2024). Finally, constraint-based prompting, iterative refinement, and structured outline generation ensure rigor by focusing on outputs in the methodological sections and mitigating superficiality (Walter, 2024).

5.2. Instrument/Protocol Development and Data Collection

Generative artificial intelligence supports the development of research instruments and data collection protocols by functioning as a methodological brainstorming assistant, auto-generating tailored interview guides, survey items, and observation protocols from study descriptions and objectives (Dam & Glomann, 2025; Parker et al., 2023; Salah et al., 2024). For instance, Avidnote auto-generates topic-specific survey questions, interview guides, follow-up prompts, and justifications for methodological adaptations aligned with study aims across qualitative, quantitative, or mixed designs (Shimray & Subaveerapandiyan, 2025), whereas SciSpace extracts and adapts methodologies from field-specific PDFs via chat interfaces for contextual customization (Jain et al., 2024). General-purpose language models, such as ChatGPT, configured with mentor or editor roles, further draft structured protocols incorporating ethical and procedural elements through role-assigned prompts (Dam & Glomann, 2025; Yuan & Harun, 2024). To ensure validity and reliability, AI proposes parallel or alternative items, checks clarity via iterative meta-prompting, and suggests response formats such as Likert scales, thereby enhancing measurement robustness (Behrend & Landers, 2025; Hoffmann et al., 2024; Salah et al., 2024). Meanwhile, Paperpal refines the academic tone and flow of these items (Shimray & Subaveerapandiyan, 2025).

Concurrently, AI facilitates the collection and structuring of emerging data sources, such as learning analytics, log data, digital traces, and unstructured or semi-structured information through automated multimodal processing (Huber & Bannert, 2023; Winter et al., 2024; Yan et al., 2024).

6. AI in Data Analysis and Visualization

6.1. Qualitative Data Coding and Thematic Analysis

Generative artificial intelligence serves as a pre-analysis assistant in qualitative data coding and thematic analysis. By processing unstructured texts such as interview transcripts or field notes, AI proposes initial codes aligned with established frameworks—including thematic analysis, grounded theory, and narrative approaches—thereby expediting the familiarization and code generation phases (Christou, 2024; Nguyen-Trung, 2024). Through semantic pattern recognition and associative clustering, AI facilitates the grouping of these codes into higher-order themes by identifying correlations, collating supporting evidence across multiple documents or transcripts, and generating hierarchical maps that highlight emerging patterns and evidential linkages (Dunivin, 2025; Paoli, 2023). Furthermore, AI tools propose thematic structures and analytical roadmaps, such as structured outlines, subthemes, and logical section breaks for qualitative reports or theses, enabling efficient narrative organization (Katz et al., 2024; Nguyen-Trung, 2024). Avidnote offers dedicated modules for qualitative coding, pattern recognition, and theme extraction throughout the workflow, whereas ChatGPT and analogous large language models provide brainstorming for subtopics and iterative structural refinement via guided prompting (Amani et al., 2025; Lee et al., 2024). SciSpace aids literature-based analyses by extracting structured qualitative insights from documents into tables for thematic identification (Perkins & Roe, 2024). Notwithstanding these efficiencies, human interpretive authority is indispensable, demanding rigorous validation of outputs for contextual nuances, reflexivity, and legitimacy to mitigate biases and preserve academic standards (Christou, 2024; Davison et al., 2024).

6.2. Quantitative Analysis, Pattern Detection, and Modeling

Generative artificial intelligence bolsters quantitative data analysis, pattern detection, and statistical modeling through multiple mechanisms. First, it automates data cleaning via code generation for outlier removal, variable transformation, and filtering. Second, it expedites exploratory data analysis through histograms, summaries, and characterizations. Third, it enables regression, classification, clustering, and forecasting with tailored Python or R implementations, yielding efficiency gains and scalability for large datasets (Lund et al., 2024; Perkins & Roe, 2024). Large language models function as statistical consultants or co-pilots, recommending appropriate tests based on data properties, generating analysis code, and aiding output interpretation with contextual insights. However, the final decisions remain with the researchers (Shen et al., 2023; Lund et al., 2024). Several AI tools exemplify these capabilities. ChatGPT excels in advanced reasoning for handling messy data and developing structured plans (Yuan & Harun, 2024). Julius AI supports code generation, visualizations, and self-correction for novices (Perkins & Roe, 2024). Vizly enables interactive exploratory graphs (Perkins & Roe, 2024). Avidnote facilitates quantitative workflows and meta-analysis extraction (Shimray & Subaveerapandiyan, 2025). SciSpace structures data extracted from PDFs (Jain et al., 2024). These tools enhance accessibility across disciplines

(Perkins & Roe, 2024; Sun et al., 2024). Notwithstanding these advances, limitations persist, including hallucinated outputs, implicit assumptions such as unverified normality, absence of conceptual understanding beyond pattern matching, and variability from vague prompting or stochasticity, which risk spurious correlations or irreproducibility (Perkins & Roe, 2024; Röttger et al., 2025). Human statistical judgment, rigorous verification, and ethical oversight are thus imperative for ensuring validity, transparency, and reproducibility, particularly in Q1-indexed contexts, thereby affirming epistemic authority (Brodeur et al., 2025; Shen et al., 2023).

6.3. Visualization and Communication of Results

Generative artificial intelligence enhances the visualization and communication of research results by automating the selection of appropriate charts, interactive dashboards, and graphics, while generating contextual narrative explanations that elucidate quantitative trends, qualitative patterns, and their implications (Aguado Tevar et al., 2025; Ye et al., 2024). Tools such as Vizly facilitate adaptive, interactive visualizations through features such as zooming and hovering on datasets, whereas ChatGPT and Julius AI generate code-assisted visualizations, such as real-time graphs and Python-based summaries, accompanied by plain-language interpretations of insights (Aguado Tevar et al., 2025; Perkins & Roe, 2024). Similarly, ResearchRabbit and LitMaps render graph-based depictions of literature networks, citation evolution, and scholarly interconnections to reveal trends and gaps (Cole & Boutet, 2023; Prashar & Chander, 2024). These tools further adapt communication for diverse audiences via role-based prompts, tailoring narratives with formal technical language for experts or simple analogies for non-experts, thereby improving understanding without oversimplification. Additionally, AI supports synthesis through structured tables and concise summaries tailored for dissemination, including post-publication formats (Bolaños et al., 2024; Perkins & Roe, 2024). Notwithstanding these advances, limitations such as the potential for misleading visualizations from spurious correlations, "black box" processes, and stochastic outputs necessitate human oversight to ensure interpretative accuracy, transparency, and scholarly integrity (Perkins & Roe, 2024).

7. AI-Supported Academic Writing and Quality Control

7.1. Outlining, Structuring, and Drafting

Generative artificial intelligence supports the outlining, structuring, and drafting of academic texts by generating detailed interactive templates that visualize logical flow and prevent fragmentation in journal articles, theses, and dissertations (Kacena et al., 2024; Park, 2025). Several tools support this process. Jenni AI enables collaborative real-time outlines with section prompts and integrated references (Bolaños et al., 2024; Pinzolits, 2023). SciSpace AI Writer produces literature-based structures for introductions and conclusions from multiple documents (Bolaños et al., 2024). Paperpal offers quick manuscript outlines via its Word plug-in (Park, 2025; Pinzolits, 2023). Meanwhile, ChatGPT excels at structuring reviews and chapters from data visuals or notes (Kacena et al., 2024; Xu, 2025), whereas NotebookLM customizes outlines from personal libraries. These tools also transform researcher notes or bullet points into formal prose tailored to specific sections, such as methods as reproducible protocols, discussions through result-comparison frameworks, limitations that preempt critiques, implications aligned with theory and unified conclusions (Kacena et al., 2024; Salvagno et al., 2023). Beyond drafting, AI

facilitates comparative synthesis by situating results against literature patterns and articulating evidence-based linkages without claiming autonomous authorship (Lund et al., 2024; Schryen et al., 2025). Ultimately, authorial control is preserved by treating AI as a co-pilot, with human judgment being essential for verification, manual integration, and cross-checking against primary sources to avoid hallucinations and superficiality, thereby ensuring intellectual ownership (Mann et al., 2024; Park, 2025).

7.2. Language, Style, and Translation

Generative artificial intelligence enhances clarity, conciseness, and academic tone in scholarly research by serving as an editorial assistant, with tools such as Paperpal providing real-time proofreading, paraphrasing, synonym suggestions, and "make it more academic" functions tailored for Q1 journal standards (Park, 2025; Pinzolits, 2023). When prompted as journal editors, large language models restructure arguments, eliminate redundancy or "waffling," and incorporate hedging phrases like "the data suggests" to align with scholarly conventions, transforming informal drafts into precise prose (Lin, 2025; Zhao et al., 2024). For cross-language support, particularly aiding non-native English speakers, AI facilitates source translation via tools such as DeepL or Google Translate, audio/video transcription and translation in Avidnote, and workflow drafting in native languages before polishing them into English (Giglio & Costa, 2023; Hu et al., 2025). The voice-to-text features in ChatGPT enable "brain dump" ideation from spoken thoughts, followed by iterative refinements (Shimray & Subaveerapandiyan, 2025; Zhao et al., 2024). Manual incorporation of suggestions preserves learning and authorship, whereas human control verifies outputs to prevent over-editing, homogenization, or loss of the authorial voice, ensuring intellectual ownership (Cheng et al., 2025; Park, 2025).

7.3. Citation Management and Reference Checking

Generative artificial intelligence supports citation management and reference checking in academic research by automating the generation and formatting of citations in styles such as APA, MLA, and Chicago. It also ensures consistency through gap detection and produces linked outputs that integrate seamlessly with platforms such as Microsoft Word, Google Docs, Overleaf and LaTeX (Bolaños et al., 2024; Fricke, 2018). Tools such as Paperpal generate cited content as a plug-in across these platforms and identify citation gaps to meet publishers' standards. SciSpace supports multipaper queries with extractable linked citations via its AI manager (Bolaños et al., 2024), whereas Perplexity provides traceable academic references and Consensus offers summaries with embedded citations (Bolaños et al., 2024). These AI workflows complement traditional reference managers, such as Zotero, by accelerating discovery and enabling verified, long-term organization (Fricke, 2018; Prashar & Chander, 2024).

However, general-purpose language models frequently hallucinate fabricated references, with studies documenting rates typically in the 47–69% range of non-existent citations in generated bibliographies that appear plausible (Walters & Wilder, 2023). Such errors necessitate systematic cross-checking against authoritative databases, including Scopus, Web of Science, Semantic Scholar, OpenAlex, and PubMed, to confirm their accuracy and traceability (Bolaños et al., 2024; Fricke, 2018).

7.4. Plagiarism, AI Detection, and Originality

AI-supported tools enhance plagiarism control and similarity assessment in academic writing through integrations such as Paperpal's embedding of Turnitin, which scans manuscripts against extensive databases to detect overlaps and align with university and publisher standards (Shimray & Subaveerapandiyan, 2025). AI detection mechanisms, such as SciSpace's AI Detector for batch analysis of up to 1,500 words and Thesify.ai's pre-submission argumentative evaluations, identify potentially generated content prior to journal submission (Barrot & Aranda, 2025; Shimray & Subaveerapandiyan, 2025). These technologies, leveraging machine learning and natural language processing, report high accuracy rates, with Turnitin distinguishing human and AI texts at 92–100% (Shimray & Subaveerapandiyan, 2025). However, limitations persist: false positives disproportionately affect non-native English speakers and unique styles, alongside poor detection of nuanced human writing and vulnerabilities to advanced models such as GPT-4 (Giray, 2024; Liang et al., 2023). Deliberate evasion via "humanizing" text is ethically unacceptable, in contrast to responsible light-touch, explainable editing (Cheng et al., 2025; Mann et al., 2024). True originality stems from research designs targeting overlooked gaps, cross-disciplinary insights, and author control over synthesis (Bolaños et al., 2024; Shimray & Subaveerapandiyan, 2025).

7.5. Feedback, Revision, and “Supervisor-like” Support

Generative artificial intelligence simulates supervisor or peer-review feedback in academic writing. Tools such as Thesify.ai offer pre-submission checks on argument structure, strengths, and needed revisions (Chamoun et al., 2024). Paperpal enhances text flow, clarity, and tone through targeted suggestions (Lin, 2025; Pinzolits, 2023), whereas Jenni AI facilitates collaborative editing and argument refinement (Bolaños et al., 2024). Prompting large language models to act as peer reviewers generates structured major and minor comments (Chamoun et al., 2024; Liang et al., 2024). This feedback bolsters argumentation by pinpointing weak claims, insufficient justification, methodological gaps, or biases prior to submission, often aligning closely with the critiques of human reviewers (Chamoun et al., 2024; Liang et al., 2024). It also improves coherence and narrative flow via suggested structural revisions, logical section breaks, and concise reformulations that remove redundancy without altering meaning (Lin, 2025; Pividori & Greene, 2023). Furthermore, AI clarifies research contributions by evaluating their originality, theoretical significance and implications (Ngwenyama & Rowe, 2024; Park, 2025). Nonetheless, researchers must interpret, select, and manually integrate this feedback, maintaining AI as a supportive co-pilot under manual verification to uphold academic standards (Liang et al., 2024; Ngwenyama & Rowe, 2024).

8. AI in Publication Strategy and Post-Publication Dissemination

8.1. Journal Selection and Pre-Submission Strategy

Generative artificial intelligence aids journal selection and pre-submission strategies in academic publishing by analyzing manuscript abstracts, keywords, and scopes to recommend suitable venues, including high-impact Q1 journals, along with their impact factors, scope alignments, and response times (Farber, 2025; Kousha & Thelwall,

2022). Platforms such as Avidnote's Publish section offer detailed recommendations from full-text inputs, prioritizing fast-review options. Paperpal identifies citation gaps to gauge readiness against target standards. Thesify.ai assesses scope-specific "match factors," including conference suggestions. These tools also align manuscripts with journal aims by scrutinizing exemplar articles to infer expected section orders, rhetorical patterns, and disciplinary norms, using platforms such as SciSpace and Paperpal. Finally, AI produces submission-ready materials, including tailored titles, abstracts, keywords, and cover letters, while enforcing journal-specific formatting and length constraints (Deveci et al., 2023).

8.2. Submission Materials and Peer Review Simulation

Generative artificial intelligence supports the preparation of submission materials in academic publishing by generating titles, abstracts, cover letters, and response-to-reviewer letters that are aligned with journal-specific conventions and editorial expectations (Deveci et al., 2023). For instance, Paperpal produces titles, abstracts, and journal-tailored cover letters directly from full manuscripts, whereas SciSpace generates introductory abstracts from multi-document analyses (Bolaños et al., 2024). Similarly, large language models configured with mentor-style prompts draft response-to-reviewer letters that systematically address the reviewers' comments (Liang et al., 2024). Thesify.ai complements these capabilities through comprehensive pre-submission assessments that evaluate the robustness of the arguments (Chamoun et al., 2024). Beyond preparation, AI simulates peer review by producing structured major and minor comments—pinpointing methodological weaknesses, insufficient novelty, unclear contributions, or theoretical gaps prior to submission—with substantial overlap with human feedback (Chamoun et al., 2024; Liang et al., 2024). This diagnostic feedback enables the refinement of argumentation, clarification of contributions, preemption of counterarguments, and explicit addressing of limitations, thereby mitigating the risk of desk rejection (Carabantes-Alarcón et al., 2023; Liang et al., 2024).

8.3. Post-Publication Promotion and Knowledge Mobilization

Generative artificial intelligence facilitates post-publication promotion and knowledge mobilization in academic research by repurposing scholarly outputs into diverse formats, including social media posts, LinkedIn or X threads, lay summaries, and blogs. This expands the reach of the findings beyond traditional journal readerships (Hendriks et al., 2025; Kessler et al., 2025). For instance, platforms such as Avidnote generate structured social media threads from research abstracts to enhance researcher visibility. These tools also transform full papers into visual and multimedia assets, such as slides, infographics, graphical abstracts, and explainer videos, as exemplified by SciSpace's PDF-to-video generator, slide creation features, and image generation models for conceptual diagrams (Bolaños et al., 2024; Yang et al., 2025). Ultimately, AI enables strategic dissemination practices that sustain long-term research impact (Fecher et al., 2023).

9. Ethical, Legal, and Integrity Considerations

9.1. Data Protection, Privacy, and Anonymity

Cloud-based artificial intelligence tools pose significant privacy risks in academic research when handling sensitive and unpublished data. These include persistent storage in user histories linked to profiles, inadvertent incorporation into proprietary model training datasets, intellectual property exposure through third-party access, and institutional bans or restrictions due to data security vulnerabilities (Hsu & Ching, 2023; King et al., 2025). Such threats are particularly amplified for human subject data or proprietary manuscripts, potentially enabling breaches or unauthorized inference attacks (Gallastegui & Forradellas, 2021; Wong et al., 2023). To mitigate these risks, privacy-preserving strategies such as temporary chats that exclude data from history, memory, and training; incognito modes for untracked queries; and opt-out configurations in data controls or custom GPTs to prevent model improvement from user inputs are essential (Khowaja et al., 2024; King et al., 2025). Additionally, pre-upload anonymization practices, including the manual removal of personal identifiers, pseudonymization, and de-identification of qualitative transcripts or quantitative records per HIPAA standards, play a crucial role (Campbell et al., 2023; U.S. Department of Health and Human Services, 2012). These measures align with the core GDPR principles, such as purpose limitation, data minimization, informed consent, confidentiality, and accountability, necessitating the sharing of ethical protocols with supervisors and institutional review boards (Sınacı et al., 2020; Toma et al., 2024). Ultimately, compliance with institutional and journal policies requires researchers to assume responsibility for verification, documentation, and oversight, thereby safeguarding participant privacy and research integrity (Liu & Jagadish, 2024; Song & Yu, 2025).

9.2. Bias, Fairness, and Transparency in AI-Assisted Research

Generative artificial intelligence exhibits algorithmic bias, which is rooted in training data associations and pattern-based inference rather than conceptual comprehension. This bias permeates literature recommendations by favoring highly cited papers through heightened citation preferences, hallucinated non-existent references, and the inclusion of non-academic sources that skew evidence toward mainstream narratives (Algaba et al., 2024; Barolo et al., 2025). Similarly, in qualitative coding and thematic analysis, language models introduce non-random errors influenced by demographic correlations in unstructured data, yielding inconsistent themes and misinterpretations of contextual nuances (Dengel et al., 2023; Perkins & Roe, 2024). In quantitative analysis and modeling, generative AI produces superficial outputs through uncritical metadata reuse and spurious correlations arising from unverified assumptions in the generated code, thereby compromising the analytical depth and reproducibility (Ooi et al., 2023; Vaithilingam et al., 2022). To counter these challenges, publisher guidelines from Springer, Nature, and institutions mandate transparency by documenting AI tools, prompts, purposes, and verification in the Methods section to ensure reproducibility and avert desk rejections (Lin, 2024; Walter, 2024). Ultimately, researchers should treat AI as a co-pilot, not an independent author, thereby keeping humans in control of the interpretation and conclusions. Manual citation checks, primary source triangulation, and method reporting further reduce bias and ensure ethical accountability (Eacersall et al., 2024; Henkel, 2025).

9.3. Authorship, Attribution, and Academic Integrity

Artificial intelligence cannot be considered a co-author in academic research because it lacks agency, accountability, and the capacity to assume responsibility for scholarly claims or ethical compliance (Schmidt & Meir, 2023). Major publishers, including Elsevier, Nature, Springer Nature, and the International Committee of Medical Journal Editors, explicitly prohibit listing AI tools in bylines. Instead, they conceptualize AI as assistants, co-pilots, editors, or tutors that aid tasks such as data cleaning, coding, drafting support, or language refinement—without replacing human-led critical thinking, synthesis, interpretation, and conclusions (Bozkurt, 2024; Cheng et al., 2025). Proper attribution requires transparent disclosure—specifying tools, versions, prompts, purposes, and verification—in the Methods section for substantive contributions, Acknowledgements for ancillary roles, or cover letters, aligning with the Committee on Publication Ethics guidelines and institutional policies (Committee on Publication Ethics, 2023; Rughiniș et al., 2025). Journals and universities permit limited AI use, such as for proofreading or brainstorming, but prohibit authorship or undisclosed substantive involvement and mandate researcher supervision (Schmidt & Meir, 2023; Walter, 2024). Verification with plagiarism and AI-detection tools, such as Turnitin, is essential, rendering the concealment of AI use ethically unacceptable while affirming human authorship, accountability, and epistemic responsibility (Cheng et al., 2025; Weber-Wulff et al., 2023).

9.4. Maintaining Critical Thinking and Intellectual Contribution

Researchers must remain the primary decision-makers and interpreters in AI-assisted academic research. Generative artificial intelligence relies on pattern recognition and data correlations, but lacks conceptual understanding, lived experience, and epistemic responsibility—qualities essential for scholarly authorship, theoretical synthesis, and contextual judgment (Messeri & Crockett, 2024; Ngwenyama & Rowe, 2024). Publisher guidelines and ethical frameworks position AI strictly as a co-pilot or assistant for augmentation, incapable of assuming accountability for interpreting data or drawing conclusions, thereby necessitating researcher supervision to preserve disciplinary insight and integrity (Bozkurt, 2024; Fecher et al., 2023). Overreliance on AI manifests as superficial analyses lacking novelty, fabricated citations in bibliographies, opaque methodologies yielding desk rejections, and erosion of critical skills through cognitive disengagement (Fecher et al., 2023; Ngwenyama & Rowe, 2024; Walters & Wilder, 2023). Practical strategies include developing contrarian ideas and gaps prior to AI engagement, reserving AI for routine tasks such as data cleaning, employing brain-dump workflows followed by polishing, and treating outputs as unverified drafts requiring manual primary triangulation and refinement (Mezzadri, 2025; Perkins & Roe, 2024). Maintaining a critical distance through human-in-the-loop verification ensures methodological transparency, originality, and ethical accountability, upholding intellectual ownership (Eacersall et al., 2024; Henkel, 2025).

10. Developing AI Literacy and Institutional Guidelines

10.1. Core AI Literacy for Researchers and Students

Core AI literacy for researchers and students encompasses an understanding of both the strengths and limitations of generative AI. Strengths include pattern recognition, data synthesis, summarization, and automating routine

tasks such as citation formatting. Limitations include the absence of true conceptual understanding, hallucinated citations, superficial reasoning akin to undergraduate arguments, and challenges with unstructured data. It also involves grasping the basics, such as temperature settings and the distinctions between general-purpose large language models and academic tools prioritizing peer-reviewed sources (Tzirides et al., 2024; Walter, 2024). This literacy is essential for responsible research, as it prevents an uncritical overreliance that undermines critical judgment and fosters cognitive disengagement, thereby safeguarding research integrity and ethics (Fecher et al., 2023; Ngwenyama & Rowe, 2024). Key skills include effective prompt engineering, rigorous output verification, and workflow strategies that position AI as a co-pilot for routine tasks, while reserving human judgment for interpretation, synthesis, and innovation (Lo, 2023; Walter, 2024). Ultimately, AI literacy frees cognitive resources for deep work and upholds human epistemic authority and ethical accountability (Fecher et al., 2023; Ngwenyama & Rowe, 2024).

10.2. Best Practices for Integrating AI into Research Workflows

Best practices for integrating generative artificial intelligence into academic research workflows center on core operational principles that prioritize human-in-the-loop decision-making, whereby researchers maintain authority over interpretation, theoretical synthesis, and scholarly judgments while positioning AI strictly as a co-pilot for augmentation (Park, 2025; Perkins & Roe, 2024). Staged verification protocols necessitate rigorous manual fact-checking of AI outputs—particularly citations prone to hallucinations— by cross-referencing them against primary sources using tools such as Consensus or SciSpace, thereby mitigating errors and biases (Bolaños et al., 2024; Ngwenyama & Rowe, 2024). Systematic documentation of AI involvement—encompassing tools, prompts, purposes, and verification steps—must be reported in the Methods sections or cover letters, alongside version control for iterative drafts, in accordance with publisher guidelines from Springer Nature emphasizing transparency for replicability (Lin, 2024; Rughiniș et al., 2025). These principles are exemplified through practical workflows. For instance, in AI-assisted literature reviews, Research Rabbit can be used to map, SciSpace to filter, Avidnote to extract data, Consensus Meter to check consensus, and Jenni AI to generate outlines, all of which culminate in human validation. Similarly, brain dump-to-manuscript workflows involve voice-to-text for ideas, visual structuring, co-pilot drafting in Paperpal, and pre-submission checks with Thesify and Turnitin to protect the intellectual property (Bolaños et al., 2024; Shimray & Subaveerapandiyan, 2025). These approaches ensure academic rigor, methodological transparency, and epistemic responsibility (Cheng et al., 2025; Lin, 2024).

10.3. Institutional and Disciplinary Guidelines

Emerging policies from universities, funding agencies, and academic publishers universally recognize generative artificial intelligence as a research support tool rather than an author. Key stakeholders—including the National Institutes of Health, the European Research Council, Elsevier, and Springer Nature—emphasize that AI cannot assume accountability, ethical responsibility, or legal ownership for scholarly outputs (Rughiniș et al., 2025; Schmidt & Meir, 2023). Elsevier mandates the disclosure of AI tools in manuscripts to promote transparency and compliance with the terms of use, while Springer Nature explicitly prohibits AI attribution on bylines and requires

documentation of its role in the methods or acknowledgements sections, aligning with the Committee on Publication Ethics recommendations (Helmy et al., 2025; Schmidt & Meir, 2023). These requirements for mandatory reporting in cover letters, methodology, or acknowledgements address transparency, reproducibility, and integrity concerns, thereby preventing ghostwriting and fabricated content (AlFayyad et al., 2025; Yoo, 2025). The heightened adoption of AI detection mechanisms by Q1 journals, such as MDPI, reflects apprehensions over diminished novelty, analytical depth, and methodological rigor in undisclosed AI-generated submissions (Ganjavi et al., 2023; Picazo-Sanchez & Ortiz-Martin, 2024). Effective institutional guidelines should incorporate transparency through declarations of AI roles, data privacy protections via opt-outs and incognito modes, researcher training in prompt engineering and verification, and support structures, including ethical AI mentors such as Thesify.ai and standardized operating procedures (Chan, 2023; Walter, 2024).

11. Overview of Key Platforms and Use Cases

11.1. Comparative Snapshot of Selected Tools

Generative AI tools such as SciSpace, Avidnote, Paperpal, ChatGPT, Perplexity, Consensus, Research Rabbit, LitMaps, Jenni.ai, and Julius AI offer specialized features tailored to various research stages, significantly boosting efficiency while requiring rigorous human verification (Bolaños et al., 2024; Shimray & Subaveerapandiyan, 2025).

For idea generation and topic selection, tools such as SciSpace scan PDFs to identify literature gaps, ChatGPT refines broad themes into precise research questions using PICO frameworks, and Consensus's Meter highlights areas of scholarly disagreement to inspire novel directions (Bolaños et al., 2024; Zhou et al., 2025).

In literature reviews and mapping, Research Rabbit generates chronological visualizations and citation networks, Perplexity's academic mode provides scholarly searches with structured summaries, and LitMaps reveals conceptual connections and overlooked papers (Bolaños et al., 2024; Pinzolits, 2023).

The theoretical frameworks and methodology design are enhanced by Avidnote's suggestions for qualitative-quantitative alignment and Consensus's validation of previous empirical applications (Bolaños et al., 2024; Shimray & Subaveerapandiyan, 2025).

Data collection and analysis benefit from Avidnote's automated survey and interview generation, ChatGPT's code-based reasoning for unstructured data, and Julius AI's self-correcting statistical models (Shimray & Subaveerapandiyan, 2025; Zhou et al., 2025).

Academic writing is supported by Paperpal's Word-integrated proofreading and tone adjustments, Jenni.ai's detailed outlines, and Grammarly's foundational checks (Bolaños et al., 2024; Shimray & Subaveerapandiyan, 2025).

Finally, for publication and promotion, Thesify.ai offers pre-submission reviews, Paperpal generates cover letters and abstracts, SciSpace provides AI detection and video creation tools, and Avidnote manages social media threads (Bolaños et al., 2024; Shimray & Subaveerapandiyan, 2025).

Across all stages, these platforms demand strict human control to uphold academic standards (Bolaños et al., 2024; Shimray & Subaveerapandiyan, 2025). Table 1 presents the categorization of AI tools by the research stage.

Table 1. Categorization of AI Tools by Research Stages

Research Stages	AI Tools	Functionalities
1. Idea Generation and Finding Topics	**SciSpace**, **Consensus**, **Avidnote**, **ChatGPT (FastTrack Mentor)**	Identifying gaps in the literature, finding points of scientific consensus or disagreement, narrowing research questions, and structuring them according to the PICO model.
2. Literature Review and Mapping	**Research Rabbit**, **LitMaps**, **Perplexity**, **Semantic Scholar**, **SciSpace**	Mapping visual connections between academic works, conducting in-depth searches focused on academic sources, extracting data through simultaneous interaction with multiple PDFs.
3. Methodology and Theoretical Framework	**Avidnote**, **Consensus**, **ChatGPT**	Proposing appropriate theoretical frameworks and research designs (qualitative, quantitative, mixed methods), verifying the use of the proposed frameworks in previous studies, designing research protocols.
4. Data Collection and Analysis	**Julius AI**, **Vizly**, **ChatGPT 4.0**, **Avidnote**	Statistical analysis and self-correction of code errors, creating interactive charts and visualizations, analyzing unstructured scientific data, generating interview questions, and assisting with qualitative coding.
5. Academic Writing and Drafting	**Paperpal**, **Jenni AI**, **Thesify**, **ThesisAI**	Real-time academic language editing within Microsoft Word, creating comprehensive thesis and article draft texts, producing long-form content, improving academic tone and flow.
6. Post-Publication Promotion and Knowledge Mobilization	**Thesify**, **Paperpal**, **SciSpace**, **Avidnote**	Simulating pre-submission evaluation and peer review comments, creating cover letters and abstracts, transforming the article into videos or social media (X/LinkedIn) threads.

12. Conclusion

12.1. Synthesizing the Role of Generative AI Across the Research Lifecycle

Generative artificial intelligence contributes substantially across the research lifecycle by augmenting human capabilities at each stage while necessitating oversight to ensure rigor (Bolaños et al., 2024; Zhou et al., 2025). In idea generation and topic selection, it facilitates brainstorming, hypothesis formulation, and gap identification through pattern synthesis from literature, enabling rapid exploration of emerging directions using tools such as Consensus and SciSpace (Ngwenyama & Rowe, 2024; Ooi et al., 2023). Literature review workflows benefit from

efficient search, screening, summarization, synthesis, and mapping, with AI automating keyword expansion, multi-document querying, and trend visualization on platforms such as Research Rabbit and Perplexity to accelerate evidence aggregation (Bolaños et al., 2024; Ngwenyama & Rowe, 2024). Methodological design leverages AI to propose frameworks, research designs, and instrument development, including survey generation and theoretical alignment via Avidnote and Consensus (Bolaños et al., 2024; Shimray & Subaveerapandiyan, 2025). Data analysis and visualization encompass qualitative coding, quantitative modeling, pattern detection, and predictive tasks, where tools such as ChatGPT and Julius AI handle cleaning, code generation, and interactive graphics for unstructured datasets (Ngwenyama & Rowe, 2024; Shimray & Subaveerapandiyan, 2025). Academic writing support is provided for outlining, drafting, language refinement, citation management, and feedback simulation through Paperpal and Jenni AI, enhancing the structure and tone (Ooi et al., 2023; Shimray & Subaveerapandiyan, 2025). Publication strategies involve journal matching, submission materials, peer review simulations, and dissemination via Thesify.ai and SciSpace, which streamline pre-submission and post-publication mobilization (Bolaños et al., 2024; Zhou et al., 2025). These integrations require human validation to mitigate limitations such as hallucinations and preserve epistemic authority (Bolaños et al., 2024; Ngwenyama & Rowe, 2024).

12.2. Towards Responsible and Reflective Adoption

The responsible and reflective adoption of generative artificial intelligence in academic research requires transparent and critical integration. This expands researchers' capabilities in ideation, literature synthesis, data analysis, and dissemination, enabling the efficient augmentation of human workflows while upholding epistemic authority through proactive risk mitigation and human oversight (Ganguly et al., 2025; Walter, 2024). This approach leverages AI's pattern recognition and automation strengths to foster innovation and interdisciplinary synthesis, provided that outputs are rigorously verified against primary sources to counter hallucinations, biases, and superficiality that could erode research integrity (Fecher et al., 2023; Lin, 2024). Ethical imperatives, encompassing data privacy protections via opt-outs and incognito modes, mandatory disclosure of AI roles in methods or acknowledgements, and accountability for accuracy, must guide implementation. These align with the publisher guidelines of Elsevier and Springer Nature, which prohibit AI authorship and demand reproducibility (Ganguly et al., 2025). Ultimately, maintaining scholarly judgment demands sustained critical thinking, ethical decision-making, and institutional training in AI literacy to ensure that AI serves as a cognitive partner subordinate to human interpretation, synthesis, and moral responsibility (Fecher et al., 2023; Walter, 2024).

12.3. Future Directions

The future directions of generative artificial intelligence in academic research point toward several transformative advancements. First, multimodal models capable of integrating text, images, audio, and video will enable richer data analysis, visualization, and literature synthesis, accommodating diverse scholarly modalities and enhancing interdisciplinary workflows (Pulk & Koris, 2025; Nguyen et al., 2024). Concurrently, agentic workflows powered by autonomous agents for multistep reasoning, planning, tool orchestration, and iterative execution will automate complex research pipelines from ideation to empirical validation and dissemination, as seen in economic modeling

and full lifecycle support systems (Dawid et al., 2025). In parallel, discipline-specific fine-tuned large language models, such as AstroSage-Llama for astronomy and CNS-Obsidian for neurosurgery, are proliferating across fields such as physics, biology, and social sciences, offering tailored precision for domain-unique tasks while overcoming generalist limitations (Alyakin et al., 2025; Haan et al., 2024). Finally, researchers and institutions must foster evolving, context-sensitive practices via adaptive guidelines that emphasize ongoing AI literacy training, transparent disclosure protocols, and human-in-the-loop oversight, aligning with ethical imperatives, institutional policies, and disciplinary nuances in this era of rapid technological change (Ganguly et al., 2025; Thompson et al., 2023).

Acknowledgements or Notes

Disclosure of AI usage

During the preparation of this work, the author(s) used Jenni.ai, ChatGPT, and NotebookLM in order to generate text, check grammar, and assist with literature search. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published work.

References

Aguado Tevar, O., Díaz-Marcos, L., Corral de la Mata, D. A., & García de Blanes Sebastián, M. (2025). Artificial intelligence in data visualization: Enhancing research through generative graphs. In J. R. Saura (Ed.), *Data-driven governance through AI, digital marketing, and the privacy interplay* (pp. 239–274). IGI Global. https://doi.org/10.4018/979-8-3693-6945-6.ch009

AlFayyad, I., Zeegers, M. P., Bouter, L. M., Macdonald, H., & Schroter, S. (2025). Authors self-disclosed use of artificial intelligence in research submissions to 49 biomedical journals: A cross-sectional study. *medRxiv*. https://doi.org/10.1101/2025.10.24.25338574

Algaba, A., Mazijn, C., Holst, V., Tori, F., Wenmackers, S., & Ginis, V. (2024). Large Language Models Reflect Human Citation Patterns with a Heightened Citation Bias. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2405.15739

Alyakin, A., Stryker, J., Alber, D. A., Lee, J. V., Sangwon, K. L., Duderstadt, B., Save, A., Kurland, D., Frome, S., Singh, S., Zhang, J., Yang, E., Park, K. Y., Orillac, C., Valliani, A. A., Neifert, S., Liu, A., Patel, A., Livia, C., … Oermann, E. K. (2025). *CNS-Obsidian: A Neurosurgical Vision-Language Model Built From Scientific Publications*. https://doi.org/10.48550/ARXIV.2502.19546

Amani, S., White, L., Balart, T., Watson, K., & Shryock, K. (2025). *Applying Generative AI for Thematic Analysis: A Model for Researchers Exploring Qualitative Methods*. https://doi.org/10.2139/ssrn.5128905

Committee on Publication Ethics. (2023). Authorship and AI tools. COPE. https://doi.org/10.24318/ccvrzbms

Barolo, D., Valentin, C., Karimi, F., Méndez, G. G., Méndez, G. G., & Espín-Noboa, L. (2025). Whose Name Comes Up? Auditing LLM-Based Scholar Recommendations. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2506.00074

Barrot, J. S., & Aranda, Ma. R. R. (2025). Efficacy of AI-Text Detection Tools in Distinguishing Student-Produced, AI-Edited, and AI-Generated Essays. *Technology Knowledge and Learning*. https://doi.org/10.1007/s10758-025-09884-0

Behrend, T. S., & Landers, R. N. (2025). Participant Interactions with Artificial Intelligence: Using Large Language Models to Generate Research Materials for Surveys and Experiments. *Journal of Business and Psychology*, *40*(6), 1275. https://doi.org/10.1007/s10869-025-10035-6

Binz, M., Alaniz, S., Roskies, A. L., Aczél, B., Bergstrom, C. T., Allen, C., Schad, D. J., Wulff, D. U., West, J. D., Zhang, Q., Shiffrin, R. M., Gershman, S. J., Попов, В. В., Bender, E. M., Marelli, M., Botvinick, M., Akata, Z., & Schulz, E. (2023). How should the advent of large language models affect the practice of science? *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2312.03759

Bjelobaba, S., Waddington, L., Perkins, M., Foltýnek, T., Bhattacharyya, S., & Weber-Wulff, D. (2024). *Research Integrity and GenAI: A Systematic Analysis of Ethical Challenges Across Research Phases*. https://doi.org/10.48550/ARXIV.2412.10134

Bolaños, F. J., Salatino, A. A., Osborne, F., & Motta, E. (2024). Artificial intelligence for literature reviews: opportunities and challenges. *Artificial Intelligence Review*, *57*(10), 268. https://doi.org/10.1007/s10462-024-10902-3

Boyko, J., Cohen, J., Fox, N. A., Veiga, M. H., Li, J. I.-H., Liu, J., Modenesi, B., Rauch, A. H., Reid, K. N., Tribedi, S., Visheratina, A., & Xie, X. (2023). An Interdisciplinary Outlook on Large Language Models for Scientific Research. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2311.04929

Bozkurt, A. (2024). GenAI et al.: Cocreation, Authorship, Ownership, Academic Ethics and Integrity in a Time of Generative AI. *Open Praxis*, *16*(1), 1. https://doi.org/10.55982/openpraxis.16.1.654

Brodeur, A., Valenta, D. T., Marcoci, A., Aparicio, J., Mikola, D., Barbarioli, B., Alexander, R., Deer, L., & Stafford, T. (2025). Comparing Human-Only, AI-Assisted, and AI-Led Teams on Assessing Research Reproducibility in Quantitative Social Science. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5118632

Campbell, R., Javorka, M., Engleton, J., Fishwick, K., Gregory, K., & Goodman-Williams, R. (2023). Open-Science Guidance for Qualitative Research: An Empirically Validated Approach for De-Identifying Sensitive Narrative Data. *Advances in Methods and Practices in Psychological Science*, *6*(4). https://doi.org/10.1177/25152459231205832

Carabantes-Alarcón, D., González-Geraldo, J. L., & Olmeda, G. J. (2023). ChatGPT could be the reviewer of your next scientific paper. Evidence on the limits of AI-assisted academic reviews. *El Profesional de La Informacion*. https://doi.org/10.3145/epi.2023.sep.16

Carchiolo, V., & Malgeri, M. (2025). Trends, Challenges, and Applications of Large Language Models in Healthcare: A Bibliometric and Scoping Review. *Future Internet*, *17*(2), 76. https://doi.org/10.3390/fi17020076

Chamoun, E., Schlichktrull, M., & Vlachos, A. (2024). Automated Focused Feedback Generation for Scientific Writing Assistance. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2405.20477

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, *20*(1), 38. https://doi.org/10.1186/s41239-023-00408-3

Cheng, A., Calhoun, A. W., & Reedy, G. (2025, April 18). Artificial intelligence-assisted academic writing: recommendations for ethical use. In *Advances in Simulation* (Vol. 10, Issue 1). BioMed Central. https://doi.org/10.1186/s41077-025-00350-6

Christou, P. (2023). How to use artificial intelligence (AI) as a resource, Methodological and Analysis Tool in Qualitative Research? *The Qualitative Report*. https://doi.org/10.46743/2160-3715/2023.6406

Christou, P. (2024). Thematic analysis through artificial intelligence (AI). *The Qualitative Report*. https://doi.org/10.46743/2160-3715/2024.7046

Cole, V., & Boutet, M. (2023). ResearchRabbit (product review). *Journal of the Canadian Health Libraries Association / Journal de l Association de Bilbiothèques de La Santé Du Canada*, *44*(2), 43. https://doi.org/10.29173/jchla29699

Cordero, J., Torres-Zambrano, J., & Cordero-Castillo, A. (2025). Integration of generative artificial intelligence in higher education: Best practices. *Education Sciences*, 15(1), 32. https://doi.org/10.3390/educsci15010032

Dam, T. V., & Glomann, L. (2025). Evaluating AI-generated Research Plans: Expert Insights from a Blind Authorship Study. *AHFE International*, *160*. https://doi.org/10.54941/ahfe1005843

Davison, R. M., Chughtai, H., Nielsen, P., Marabelli, M., Iannacci, F., van Offenbeek, M., Tarafdar, M., Trenz, M., Techatassanasoontorn, A. A., Andrade, A. D., & Panteli, N. (2024). The ethics of using generative AI for qualitative data analysis. *Information Systems Journal*, *34*(5), 1433. https://doi.org/10.1111/isj.12504

Dawid, H., Harting, P., Wang, H., Wang, Z., & Yi, J. S. (2025). Agentic Workflows for Economic Research: Design and Implementation. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2504.09736

Delios, A., Tung, R. L., & Witteloostuijn, A. van. (2024). How to intelligently embrace generative AI: the first guardrails for the use of GenAI in IB research. *Journal of International Business Studies*. https://doi.org/10.1057/s41267-024-00736-0

Dengel, A., Gehrlein, R., Fernes, D., Görlich, S., Maurer, J., Pham, H. H., Großmann, G., & Eisermann, N. D. genannt. (2023). Qualitative Research Methods for Large Language Models: Conducting Semi-Structured Interviews with ChatGPT and BARD on Computer Science Education. *Informatics*, *10*, 78. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85180688010&doi=10.3390%2finformatics10040078&partnerID=40&md5=feb2c6d888ea2dc885618ca2b0db64c9

Deveci, C. D., Baker, J. J., Sikander, B., & Rosenberg, J. (2023). A comparison of cover letters written by ChatGPT-4 or humans. *Danish Medical Journal*, *70*(12), A06230412. https://pubmed.ncbi.nlm.nih.gov/38018708

Dgebuadze, M., & Kenkebashvili, K. (2025). Integrating artificial intelligence in academic writing and research methods courses for master students: A teaching model for responsible use. *DergiPark (Istanbul University)*. https://dergipark.org.tr/en/pub/ijaedu/issue/91344/1642487

Doshi, A. R., Chai, S., & Troebinger, M. (2025). How Experience Moderates the Impact of Generative AI Ideas on the Research Process. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5013086

Dunivin, Z. O. (2025). Scaling hermeneutics: a guide to qualitative coding with LLMs for reflexive content analysis. *EPJ Data Science*, *14*(1). https://doi.org/10.1140/epjds/s13688-025-00548-8

Eacersall, D., Pretorius, L., Смирнов, И. Е., Spray, E., Illingworth, S., Chugh, R., Strydom, S., Stratton-Maher, D., Simmons, J., Jennings, I., Roux, R., Kamrowski, R., Downie, A., Thong, C. L., & Howell, K. A. (2024). Navigating Ethical Challenges in Generative AI-Enhanced Research: The ETHICAL Framework for Responsible Generative AI Use. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2501.09021

Ekundayo, T., Khan, Z., & Nuzhat, S. (2024). Evaluating the Influence of Artificial Intelligence on Scholarly Research: A Study Focused on Academics. *Human Behavior and Emerging Technologies*, *2024*, 1. https://doi.org/10.1155/2024/8713718

Färber, M., Lamprecht, D., Krause, J., Aung, L., & Haase, P. (2023a). SemOpenAlex: The Scientific Landscape in 26 Billion RDF Triples. In *Lecture notes in computer science* (p. 94). Springer Science+Business Media. https://doi.org/10.1007/978-3-031-47243-5_6

Färber, M., Lamprecht, D., Krause, J., Aung, L., & Haase, P. (2023b). SemOpenAlex: The Scientific Landscape in 26 Billion RDF Triples. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2308.03671

Farber, S. (2025). Enhancing Academic Decision-Making: A Pilot Study of AI-Supported Journal Selection in Higher Education. *Innovative Higher Education*. https://doi.org/10.1007/s10755-025-09791-3

Fardim, K. A. C., Gonçalves, S. E. de P., & Tribst, J. P. M. (2023). Scientific writing with artificial intelligence: key considerations and alerts. *Brazilian Dental Science*, *26*(3). https://doi.org/10.4322/bds.2023.e3988

Fecher, B., Hebing, M., **Laufer, M.,** Pohle, J., & Sofsky, F. (2023). Friend or foe? Exploring the implications of large language models on the science system. *AI & Society*, *40*, 447–459. https://doi.org/10.1007/s00146-023-01791-1

Feng, Y., Vanam, S., Cherukupally, M., Zheng, W., Qiu, M., & Chen, H. (2023). Investigating code generation performance of ChatGPT with crowdsourcing social data. *In 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)* (pp. 876–885). IEEE. https://doi.org/10.1109/COMPSAC57700.2023.00117

U.S. Department of Health and Human Services. (2012). *Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

Fricke, S. (2018). Semantic Scholar. *Journal of the Medical Library Association JMLA*, *106*(1), **145-147**. https://doi.org/10.5195/jmla.2018.280

Gallastegui, L. M. G., & Forradellas, R. F. R. (2021). Business Methodology for the Application in University Environments of Predictive Machine Learning Models Based on an Ethical Taxonomy of the Student's Digital Twin. *Administrative Sciences*, *11*(4), 118. https://doi.org/10.3390/admsci11040118

Ganguly, A., Johri, A., Ali, A., & McDonald, N. (2025). Generative artificial intelligence for academic research: evidence from guidance issued for researchers by higher education institutions in the United States. *AI and Ethics*. https://doi.org/10.1007/s43681-025-00688-7

Ganjavi, C., Eppler, M., Pekcan, A., Biedermann, B., Abreu, A. L., Collins, G. S., Gill, I. S., & Cacciamani, G. (2023). Bibliometric Analysis of Publisher and Journal Instructions to Authors on Generative-AI in Academic and Scientific Publishing. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2307.11918

Ge, L., Agrawal, R., Singer, M., Kannapiran, P., Molina, J. A. D. C., Teow, K. L., Yap, C. W., & Abisheganaden, J. (2024). Leveraging artificial intelligence to enhance systematic reviews in health research: advanced tools and challenges. *Systematic Reviews*, *13*(1), 269. https://doi.org/10.1186/s13643-024-02682-2

Pulk, K., & Koris, R. (Eds.). (2025). *Generative AI in higher education: The good, the bad, and the ugly*. Edward Elgar Publishing. https://doi.org/10.4337/9781035326020

George, B., & Wooden, O. S. (2023). Managing the Strategic Transformation of Higher Education through Artificial Intelligence. *Administrative Sciences*, *13*(9), 196. https://doi.org/10.3390/admsci13090196

Giglio, A. del, & Costa, M. U. P. da. (2023). The use of artificial intelligence to improve the scientific writing of non-native english speakers. *Revista Da Associação Médica Brasileira*, *69*(9). Brazilian Medical Association. https://doi.org/10.1590/1806-9282.20230560

Giray, L. (2024). The Problem with False Positives: AI Detection Unfairly Accuses Scholars of AI Plagiarism. *The Serials Librarian*, 1. https://doi.org/10.1080/0361526x.2024.2433256

Gu, X., & Krenn, M. (2025). Forecasting high-impact research topics via machine learning on evolving knowledge graphs. *Machine Learning Science and Technology*. https://doi.org/10.1088/2632-2153/add6ef

Haan, T. de, Ting, Y.-S., Ghosal, T., Nguyen, T. D., Accomazzi, A., Wells, A., Ramachandra, N., Pan, R., & Sun, Z. (2024). AstroMLab 3: Achieving GPT-4o Level Performance in Astronomy with a Specialized 8B-Parameter Large Language Model. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2411.09012

Helmy, M., Jin, L., Alhossary, A., Mansour, T., Pellegrina, D., & Selvarajoo, K. (2025). Ten simple rules for optimal and careful use of generative AI in science. *PLoS Computational Biology*, *21*(10). https://doi.org/10.1371/journal.pcbi.1013588

Hendriks, F., David, Y. B.-B., Banse, L., Fick, J., Greussing, E., Klein-Avraham, I., Rakedzon, T., Taddicken, M., & Baram-Tsabari, A. (2025). Generative AI in Science Communication: Fostering Scientists' Good Working Habits for Ethical and Effective Use. *Science Communication*. https://doi.org/10.1177/10755470251343486

Henkel, J. (2025). The mathematician's assistant: integrating AI into research practice. *Mathematische Semesterberichte*, *72*(2), 117. https://doi.org/10.1007/s00591-025-00400-0

Hoffmann, S., Lasarov, W., & Dwivedi, Y. K. (2024). AI-empowered scale development: Testing the potential of ChatGPT. *Technological Forecasting and Social Change*, *205*, 123488. https://doi.org/10.1016/j.techfore.2024.123488

Shen, H., Li, T., Li, T. J.-J., Park, J. S., & Yang, D. (2023). Shaping the emerging norms of using large language models in social computing research. arXiv. https://doi.org/10.48550/arxiv.2307.04280

Hsu, Y., & Ching, Y. (2023). Generative Artificial Intelligence in Education, Part One: the Dynamic Frontier. *TechTrends*, *67*(4), 603. https://doi.org/10.1007/s11528-023-00863-9

Hu, H., Zhou, Q., & Hashim, H. (2025). Negotiating identity in the age of ChatGPT: non-native English researchers' experiences with AI-assisted academic writing. *Humanities and Social Sciences Communications*, *12*(1). https://doi.org/10.1057/s41599-025-05351-4

Huber, K., & Bannert, M. (2023). Investigating learning processes through analysis of navigation behavior using log files. *Journal of Computing in Higher Education*, *36*(3), 683. https://doi.org/10.1007/s12528-023-09372-3

Jain, S., Kumar, A., Roy, T., Shinde, K., Vignesh, G., & Tondulkar, R. (2024). SciSpace Literature Review: Harnessing AI for Effortless Scientific Discovery. In *Lecture notes in computer science* (p. 256). Springer Science+Business Media. https://doi.org/10.1007/978-3-031-56069-9_28

Jovanovic, M., & Campbell, M. (2022). Generative Artificial Intelligence: Trends and Prospects. *Computer*, *55*(10), 107. https://doi.org/10.1109/mc.2022.3192720

Kacena, M. A., Plotkin, L. I., & Fehrenbacher, J. C. (2024). The Use of Artificial Intelligence in Writing Scientific Review Articles. *Current Osteoporosis Reports*, *22*(1), 115. Springer Science+Business Media. https://doi.org/10.1007/s11914-023-00852-0

Katsamakas, E., Pavlov, O. V., & Saklad, R. (2024). Artificial Intelligence and the Transformation of Higher Education Institutions: A Systems Approach. *Sustainability*, *16*(14), 6118. https://doi.org/10.3390/su16146118

Katz, A., Fleming, G. C., & Main, J. (2024). Thematic Analysis with Open-Source Generative AI and Machine Learning: A New Method for Inductive Qualitative Codebook Development. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2410.03721

Ke, R., & Ribeiro, R. M. (2025). Roadmap for using large language models (LLMs) to accelerate cross-disciplinary research with an example from computational biology. *arXiv*. https://doi.org/10.48550/arXiv.2507.03722

Kessler, S. H., Mahl, D., Schäfer, M. S., & Volk, S. C. (2025). Science Communication in the Age of Artificial Intelligence. *Journal of Science Communication*, *24*(2). https://doi.org/10.22323/2.24020501

Khowaja, S. A., Khuwaja, P., Dev, K., Wang, W., & Nkenyereye, L. (2024). ChatGPT Needs SPADE (Sustainability, PrivAcy, Digital divide, and Ethics) Evaluation: A Review. *Cognitive Computation*, *16*(5), 2528. Springer Science+Business Media. https://doi.org/10.1007/s12559-024-10285-1

King, J., Klyman, K., Capstick, E., Saade, T., & Hsieh, V. (2025). User Privacy and Large Language Models: An Analysis of Frontier Developers' Privacy Policies. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2509.05382

Kinney, R., Anastasiades, C., Authur, R., Beltagy, I., Bragg, J., Buraczynski, A., Cachola, I., Candra, S., Chandrasekhar, Y., Cohan, A., Crawford, M., Downey, D., Dunkelberger, J., Etzioni, O., Evans, R. L., Feldman, S., Gorney, J., Graham, D. Y., Hu, F., … Weld, D. S. (2023). The Semantic Scholar Open Data Platform. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2301.10140

Kousha, K., & Thelwall, M. (2022). Artificial intelligence technologies to support research assessment: A review. *arXiv (Cornell University)*. Cornell University. https://doi.org/10.48550/arxiv.2212.06574

Lee, V. V., Lubbe, S. C. C. van der, Goh, L. H., & Valderas, J. M. (2024). Harnessing ChatGPT for Thematic Analysis: Are We Ready? *Journal of Medical Internet Research*, *26*. https://doi.org/10.2196/54974

Liang, W., Yüksekgönül, M., Mao, Y., Wu, E. Q., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, *4*(7), 100779. https://doi.org/10.1016/j.patter.2023.100779

Liang, W., Zhang, Y., Cao, H., Wang, B., Ding, D. Y., Yang, X., Vodrahalli, K., He, S., Smith, D. S., Yin, Y., McFarland, D. A., & Zou, J. (2024). Can Large Language Models Provide Useful Feedback on

Research Papers? A Large-Scale Empirical Analysis. *NEJM AI*, *1*(8). https://doi.org/10.1056/aioa2400196

Lin, Z. (2024). Building ethical guidelines for generative AI in scientific research. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2401.15284

Lin, Z. (2025). Techniques for supercharging academic writing with generative AI. *Nature Biomedical Engineering*, *9*(4), 426. https://doi.org/10.1038/s41551-024-01185-8

Liu, J., & Jagadish, H. V. (2024). Institutional Efforts to Help Academic Researchers Implement Generative AI in Research. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.2c8e7e81

Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, *49*(4), 102720. https://doi.org/10.1016/j.acalib.2023.102720

Lo, N., Wong, A., & Chan, S. T. (2025). The impact of generative AI on essay revisions and student engagement. *Computers and Education Open*, *9*, 100249. https://doi.org/10.1016/j.caeo.2025.100249

Luckin, R., & Holmes, W. (2016). *Intelligence unleashed: An argument for AI in education*. https://oro.open.ac.uk/50104/

Lund, B., Lamba, M., & Oh, S. (2024). The Impact of AI on Academic Research and Publishing. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2406.06009

Luttges, B. L., Rogers, T., Goldstein, D. G., Ungar, L., & Duckworth, A. (2025). Coach not crutch: AI assistance can enhance rather than hinder skill development. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2502.02880

Mann, S. P., Vazirani, A. A., Aboy, M., Earp, B. D., Minssen, T., Cohen, I. G., & Savulescu, J. (2024). Guidelines for ethical use and acknowledgement of large language models in academic writing. *Nature Machine Intelligence*, *6*(11), 1272. https://doi.org/10.1038/s42256-024-00922-7

Messeri, L., & Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, *627*(8002), 49. https://doi.org/10.1038/s41586-024-07146-0

Mezzadri, D. (2025). The Paradox of Ethical AI-Assisted Research. *Journal of Academic Ethics*. https://doi.org/10.1007/s10805-025-09671-7

Minaee, S., Mikolov, T., Nikzad-Khasmakhi, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). Large Language Models: A Survey. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2402.06196

Musleh, A., & AlRyalat, S. A. (2025). Artificial Intelligence and Large Language Model Powered Literature Review Services. *High Yield Medical Reviews*, *3*(1). https://doi.org/10.59707/hymrpsey7778

Mutelo, I. (2025). Understanding the Generative Artificial Intelligence Revolution in Zambian Higher Education Research: Adoption, Challenges, and Strategies for Responsible Integration. *International Journal of Research and Innovation in Social Science*, 5731. https://doi.org/10.47772/ijriss.2025.903sedu0416

Nguyen, A., Kremantzis, M. D., Essien, A., Petrounias, I., & Hosseini, S. (2024). Editorial: Enhancing Student Engagement Through Artificial Intelligence (AI): Understanding the Basics, Opportunities, and Challenges. *Journal of University Teaching and Learning Practice*, *21*(6). https://doi.org/10.53761/caraaq92

Nguyen-Trung, K. (2024). *ChatGPT in Thematic Analysis: Can AI become a research assistant in qualitative research?* https://doi.org/10.31219/osf.io/vefwc

Ngwenyama, O., & Rowe, F. (2024). Should We Collaborate with AI to Conduct Literature Reviews? Changing Epistemic Values in a Flattening World. *Journal of the Association for Information Systems*, *25*(1), 122-136. https://doi.org/10.17705/1jais.00869

Nigam, H., Patwardhan, M., Vig, L., & Shroff, G. (2024). Acceleron: A Tool to Accelerate Research Ideation. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2403.04382

Ooi, K., Tan, G. W., Al-Emran, M., Al-Sharafi, M. A., Căpățînă, A., Chakraborty, A., Dwivedi, Y. K., Huang, T.-L., Kar, A. K., Lee, V., Loh, X.-M., Micu, A., Mikalef, P., Mogaji, E., Pandey, N., Raman, R., Rana, N. P., Sarker, P., Sharma, A., … Wong, L. (2023). The Potential of Generative Artificial Intelligence Across Disciplines: Perspectives and Future Directions. *Journal of Computer Information Systems*, 1. https://doi.org/10.1080/08874417.2023.2261010

Panda, S., & Kaur, N. (2024). Exploring the role of generative AI in academia: Opportunities and challenges. *IP Indian Journal of Library Science and Information Technology*, *9*(1), 12. https://doi.org/10.18231/j.ijlsit.2024.003

Paoli, S. D. (2023). Can Large Language Models emulate an inductive Thematic Analysis of semi-structured interviews? An exploration and provocation on the limits of the approach and the model. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2305.13014

Park, J.-B. (2025). Towards a transparent and reproducible AI-assisted research paper writing. *Genomics & Informatics*, *23*(1), 26. https://doi.org/10.1186/s44342-025-00057-0

Parker, J., Richard, V. M., & Becker, K. D. (2023). Flexibility & Iteration: Exploring the Potential of Large Language Models in Developing and Refining Interview Protocols. *The Qualitative Report*. https://doi.org/10.46743/2160-3715/2023.6695

Perkins, M., & Roe, J. (2024). Generative AI Tools in Academic Research: Applications and Implications for Qualitative and Quantitative Research Methodologies. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2408.06872

Picazo-Sanchez, P., & Ortiz-Martin, L. (2024). Analysing the impact of ChatGPT in research. *Applied Intelligence*, *54*(5), 4172. https://doi.org/10.1007/s10489-024-05298-0

Pinzolits, R. F. J. (2023). AI in academia: An overview of selected tools and their areas of application. *MAP Education and Humanities*, *4*(1), 37. https://doi.org/10.53880/2744-2373.2023.4.37

Pisica, A. I., Edu, T., Zaharia, R. M., & Zaharia, R. (2023). Implementing Artificial Intelligence in Higher Education: Pros and Cons from the Perspectives of Academics. *Societies*, *13*(5), 118. https://doi.org/10.3390/soc13050118

Pividori, M., & Greene, C. S. (2023). A publishing infrastructure for AI-assisted academic authoring. *bioRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/2023.01.21.525030

Polat, H. (2023). Transforming education with artificial intelligence: Shaping the path forward. In A. Kaban & A. Stachowicz-Stanusch (Eds.), Empowering education: Exploring the potential of artificial intelligence (pp. 3–20). ISTES Organization

Prashar, P., & Chander, H. (2024). Role of Research Management Tools for Enhancing Academic Research Efficiency: A Focus on AI-Driven Literature Mapping with Research Rabbit. *Journal of Information and Knowledge*, 313. https://doi.org/10.17821/srels/2024/v61i6/171639

Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, *13*(9), 5783. https://doi.org/10.3390/app13095783

Rahman, Md. M., Terano, H. J. R., Rahman, M. N., Salamzadeh, A., & Rahaman, Md. S. (2023). ChatGPT and Academic Research: A Review and Recommendations Based on Practical Examples. *Journal of Education Management and Development Studies*, *3*(1), 1. https://doi.org/10.52631/jemds.v3i1.175

Röttger, P., Urman, A., Wendsjö, A., & Plaza-del-Arco, F. M. (2025). Large Language Model Hacking: Quantifying the Hidden Risks of Using LLMs for Text Annotation. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2509.08825

Rughiniș, C., Vulpe, S., Țurcanu, D., & Rughiniș, R. (2025). AI at the knowledge gates: institutional policies and hybrid configurations in universities and publishers. *Frontiers in Computer Science*, *7*. https://doi.org/10.3389/fcomp.2025.1608276

Salah, M., Abdelfattah, F., Alhalbusi, H., Jassem, S., Mohammed, M., Ismail, M. M., & Washahi, M. A. (2024). Can Generative AI Craft Variable Questions? A Mixed-Method Study on AI's Capability to Adopt, Adapt, and Create New Scales. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-3924447/v1

Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare*, *11*(6), 887. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/healthcare11060887

Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Can artificial intelligence help for scientific writing?. *Critical Care*, *27*(1), 75. BioMed Central. https://doi.org/10.1186/s13054-023-04380-2

Sarkar, A., Xiaotong, Xu, Toronto, N., Drosos, I., & Poelitz, C. (2024). When Copilot Becomes Autopilot: Generative AI's Critical Risk to Knowledge Work and a Critical Solution. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2412.15030

Schmidt, P. G., & Meir, A. J. (2023). Using Generative AI for Literature Searches and Scholarly Writing: Is the Integrity of the Scientific Discourse in Jeopardy? *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2311.06981

Schryen, G., Marrone, M., & Yang, J. (2025). Exploring the scope of generative AI in literature review development. *Electronic Markets*, *35*(1). https://doi.org/10.1007/s12525-025-00754-2

Shimray, S. R., & Subaveerapandiyan, A. (2025). Artificial Intelligence in Academic Writing and Research: Adoption and Effectiveness. *Open Information Science*, *9*(1). https://doi.org/10.1515/opis-2025-0026

Sınacı, A. A., Núñez-Benjumea, F. J., Gençtürk, M., Jauer, M.-L., Deserno, T. M., Chronaki, C., Cangioli, G., Cavero-Barca, C., Rodríguez-Pérez, J. M., Pérez-Pérez, M. M., Ertürkmen, G. B. L., Hernández-Pérez, T., Méndez, E., & Parra-Calderón, C. L. (2020). From Raw Data to FAIR Data: The FAIRification Workflow for Health Research. *Methods of Information in Medicine*, *59*. https://doi.org/10.1055/s-0040-1713684

Song, R., & Yu, S. (2025). Generative AI Policies in Academic Research. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5139585

Sun, M., Han, R., Jiang, B., Qi, H., Sun, D., Yuan, Y., & Huang, J. (2024). A Survey on Large Language Model-based Agents for Statistics and Data Science. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2412.14222

Teremetskyi, V., Burylo, Y., Zozuliak, O., Koshmanov, M., Polyova, N., & Olha, P. (2024). Academic Integrity in The Age of Artificial Intelligence: World Trends and Outlook for Ukraine from The Legal Perspective. *Pakistan Journal of Life and Social Sciences (PJLSS)*, *22*(1), 2020. https://doi.org/10.57239/pjlss-2024-22.1.00147

Thompson, K., Corrin, L., & Lodge, J. M. (2023). AI in tertiary education: progress on research and practice. *Australasian Journal of Educational Technology*, *39*(5), 1. https://doi.org/10.14742/ajet.9251

Toma, M., Bönisch, C., Löhnhardt, B., Kelm, M., Bohnenberger, H., Winkelmann, S., Ströbel, P., & Kesztyüs, T. (2024). Research collaboration data platform ensuring general data protection. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-61912-8

Tzirides, A. O., Zapata, G. C., Kastania, N. P., Saini, A. K., Castro, V., Ismael, S. A., You, Y., Santos, T. A. dos, Searsmith, D., O'Brien, C., Cope, B., & Kalantzis, M. (2024). Combining human and artificial intelligence for enhanced AI literacy in higher education. *Computers and Education Open*, *6*, 100184. https://doi.org/10.1016/j.caeo.2024.100184

Vaithilingam, P., Zhang, T., & Glassman, E. L. (2022). Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–7). ACM. https://doi.org/10.1145/3491101.3519665

Vicent, M., Katushabe, C., Muhoza, G., & Rugasira, J. (2025). A systematic review of the impact of generative AI on postgraduate research: opportunities, challenges, and ethical implications. *Discover Artificial Intelligence*, *5*(1). Springer Nature. https://doi.org/10.1007/s44163-025-00495-3

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-024-00448-3

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, *13*(1), 14045. https://doi.org/10.1038/s41598-023-41032-5

Wang, S., Scells, H., Koopman, B., & Zuccon, G. (2023). Can ChatGPT write a good Boolean query for systematic review literature search? In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1426–1436). ACM. https://doi.org/10.1145/3539618.3591703

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, *19*(1), 26. https://doi.org/10.1007/s40979-023-00146-z

Winter, M., Mordel, J., Mendzheritskaya, J., Biedermann, D., Ciordas-Hertel, G.-P., Hahnel, C., Bengs, D., Wolter, I., Goldhammer, F., Drachsler, H., Artelt, C., & Horz, H. (2024). Behavioral trace data in an online learning environment as indicators of learning engagement in university students. *Frontiers in Psychology*, *15*, 1396881. https://doi.org/10.3389/fpsyg.2024.1396881

Wong, J., Racine, L., Henderson, T., & Ball, K. (2023). Online Learning as a Commons: Supporting students' data protection preferences through a collaborative digital environment. In *Journal of Intellectual Property, Information Technology and E-Commerce Law* (Vol. 14, p. 251). https://www.scopus.com/inward/record.uri?eid=2-s2.0-85167907776&partnerID=40&md5=52342223d1af3a164d57291a0df33cf7

Wu, D., Zhang, S., Ma, Z., Yue, X., & Dong, R. K. (2024). Unlocking Potential: Key Factors Shaping Undergraduate Self-Directed Learning in AI-Enhanced Educational Environments. *Systems*, *12*(9), 332. https://doi.org/10.3390/systems12090332

Xu, R., & Peng, J. (2025). A Comprehensive Survey of Deep Research: Systems, Methodologies, and Applications. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2506.12594

Xu, Z. (2025). Patterns and Purposes: A Cross-Journal Analysis of AI Tool Usage in Academic Writing. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2502.00632

Yan, L., Martínez-Maldonado, R., & Gašević, D. (2024). *Generative Artificial Intelligence in Learning Analytics: Contextualising Opportunities and Challenges through the Learning Analytics Cycle*. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (LAK '24) (pp. 101-111). ACM. https://doi.org/10.1145/3636555.3636856

Yang, Y., Jiang, W., Wang, Y., Wang, Y., & Zhang, C. (2025). Auto-Slides: An Interactive Multi-Agent System for Creating and Customizing Research Presentations. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2509.11062

Ye, Y., Hao, J., Hou, Y., Wang, Z., Xiao, S., Luo, Y., & Zeng, W. (2024). Generative AI for visualization: State of the art and future directions. *Visual Informatics*, *8*(2), 43. https://doi.org/10.1016/j.visinf.2024.04.003

Yoo, J. (2025). Defining the Boundaries of AI Use in Scientific Writing: A Comparative Review of Editorial Policies. *Journal of Korean Medical Science*, *40*(23). Korean Academy of Medical Sciences. https://doi.org/10.3346/jkms.2025.40.e187

Yuan, Y., & Harun, J. (2024). Empowering Doctoral Academic Research: Artificial Intelligence-driven Insights from Large Language Models. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-4337026/v1

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zhao, X., Cox, A., & Cai, L. (2024). ChatGPT and the digitisation of writing. *Humanities and Social Sciences Communications*, *11*(1). https://doi.org/10.1057/s41599-024-02904-x

Zhou, Z., Feng, X., Huang, L., Feng, X., Song, Z., Chen, R., Zhao, L., Ma, W., Gu, Y., Wang, B., Wu, D., Hu, G., Liu, T., & Qin, B. (2025). From Hypothesis to Publication: A Comprehensive Survey of AI-Driven Research Support Systems. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2503.01424

Author Information

Turgay DEMİREL

https://orcid.org/0000-0001-9210-8876

Ataturk University

Faculty of Applied Sciences, Erzurum

Country: Türkiye

Contact e-mail: *turgaydemirel85@gmail.com*

Citation

Demirel, T. (2025). Artificial Intelligence in Research Processes. In A. Kaban & T. Demirel (Eds.), *Interdisciplinary Applications of Artificial Intelligence - 1* (pp. 83-114). ISTES.

Chapter 6- An AI-Integrated Platform for End-to-End Meta-Analysis: Design, Implementation, and Evaluation of "tasresearch.org"

Nurullah Taş

Chapter Highlights

- A web-based, AI-assisted platform (**tasresearch.org**) enabling end-to-end systematic review and meta-analysis workflows.
- Automated full-text PDF screening and quantitative data extraction integrated with human oversight.
- Scalable, asynchronous processing architecture supporting large document collections.
- Unified implementation of core meta-analytic methods with interactive, publication-ready visualizations.
- Workflow-centric design that complements traditional statistical software and reduces manual workload.

Introduction

The increasing volume and complexity of scientific literature have made systematic reviews and meta-analyses both more essential and more demanding than ever before. While existing software provides strong support for statistical synthesis, many stages of the evidence synthesis process—particularly literature screening, data extraction, and workflow coordination—remain highly manual and fragmented. Recent advances in artificial intelligence offer new opportunities to address these challenges. In this context, **tasresearch.org** was developed as a web-based, AI-assisted platform designed to support end-to-end meta-analysis workflows in a transparent, scalable, and researcher-centered manner.

Importance of Systematic Reviews and Meta-Analyses

Systematic reviews and meta-analyses constitute core components of evidence-based practice by providing structured and transparent methods for synthesizing large and often conflicting bodies of research. In evidence hierarchies, systematic reviews of randomized controlled trials are typically positioned at the highest level, as they minimize bias and maximize the use of available data to address focused clinical or policy questions (Askie & Offringa, 2015; Zhai & Guyatt, 2023). By applying predefined and replicable protocols for literature searching, study selection, appraisal, and synthesis, systematic reviews reduce the subjectivity inherent in narrative reviews and yield more reliable summaries of the evidence base (Johnson & Hennessy, 2019; Suchmacher & Geller, 2018).

A systematic review systematically identifies and appraises relevant studies addressing a specific research question, while meta-analysis quantitatively combines their results to produce pooled estimates and confidence intervals (Ahn & Kang, 2018; Zhai & Guyatt, 2023). When appropriately conducted, meta-analysis increases statistical power, improves precision, and helps resolve uncertainty arising from underpowered or inconsistent individual studies (Gopal et al., 2025; Vetter, 2019). As a result, systematic reviews with meta-analyses often provide the most reliable estimates of intervention effects or associations, directly informing clinical decision-making and guideline development (Harris et al., 2017; Parums, 2021).

The importance of these methods spans multiple disciplines, including medicine, psychology, education, and veterinary sciences, where they support evidence-informed practice and policy by consolidating fragmented findings into actionable conclusions (Allen et al., 2023; Prott et al., 2024; Sargeant & O'Connor, 2020; Siddaway et al., 2019). In addition, systematic reviews and meta-analyses play a critical role in identifying knowledge gaps and guiding future research priorities (Møller et al., 2018) .

The growing volume of systematic reviews further highlights the need for rigorous methodological and reporting standards. Frameworks such as PRISMA, AMSTAR 2, and GRADE are widely recommended to enhance transparency, reproducibility, and the trustworthy translation of synthesized evidence into practice and policy (Harris et al., 2017; Soni, 2025). When conducted in accordance with these standards, systematic reviews and

meta-analyses remain indispensable tools for strengthening the reliability of research and advancing evidence-based decision-making.

Workflow Complexity and Bottlenecks in Meta-Analysis

Meta-analysis promises concise quantitative answers to complex research questions, yet its Meta-analysis aims to provide concise quantitative answers to complex research questions, yet its production involves a highly intricate, multi-stage workflow that is both resource-intensive and prone to bottlenecks. Standard guidelines describe sequential steps including question formulation, comprehensive searching, screening, data extraction, model selection, heterogeneity assessment, and reporting (Martínez et al., 2025; Mikolajewski et al., 2023). In practice, each stage introduces methodological decisions and operational demands that can delay completion and compromise validity. Empirical analyses show that conventional systematic reviews with meta-analysis often require more than a year to complete and involve large multidisciplinary teams, underscoring the substantial time and human resource burden of current workflows (Borah et al., 2017).

Early stages are particularly labor-intensive. Large database searches frequently yield tens of thousands of records, of which only a small proportion are ultimately included, making screening slow and error-prone (Borah et al., 2017). Heterogeneous study designs, outcomes, and measurement scales further complicate screening and data extraction, while missing or inconsistently reported statistics often require complex transformations and expert judgment, increasing the risk of error (Lorenc et al., 2016).

Substantial complexity also arises during synthesis. Selecting among fixed-effect, random-effects, meta-regression, or more advanced models entails trade-offs between interpretability, realism, and computational burden (Cipriani et al., 2013; Jung & Aloe, 2025). Additional analyses addressing heterogeneity, publication bias, and small-study effects introduce further layers of modeling and sensitivity checks, which are frequently implemented inconsistently, potentially leading to misleading conclusions (Esterhuizen & Thabane, 2016).

These challenges have motivated the development of process guides, automation tools, and web-based platforms intended to streamline specific workflow components (Abbas et al., 2024; Muka et al., 2019; Scott et al., 2023). However, evidence suggests that while certain tasks can be partially automated, critical activities such as nuanced eligibility assessment, complex data extraction, and model selection remain largely manual and represent persistent bottlenecks (Tsafnat et al., 2014). Addressing these recurring challenges is therefore essential not only for improving efficiency, but also for maintaining methodological rigor in an era of rapidly expanding evidence bases and growing reliance on meta-analytic findings for clinical and policy decision-making.

Limitations of Existing Manual and Semi-Automated Tools

Despite documented efficiency gains, existing manual and semi-automated tools for systematic reviews and meta-analyses exhibit important limitations that constrain their impact on workload reduction, accuracy, and scalability.

Chapter 6: An AI-Integrated Platform for End-to-End Meta-Analysis: Design, Implementation, and Evaluation of "tasresearch.org"

Manual workflows remain prevalent across disciplines, even as the rapid growth of scientific literature renders fully hand-driven processes increasingly unsustainable (Altena et al., 2019) . Although a growing number of automation tools are available, their adoption remains limited and largely ad hoc, often driven by peer recommendation rather than formal training or institutional support. Barriers such as licensing costs, steep learning curves, limited technical support, and poor integration with existing workflows continue to restrict widespread use (Altena et al., 2019; Scott et al., 2023).

Most semi-automated systems focus on isolated workflow stages—primarily title and abstract screening—while leaving other labor-intensive tasks, including data extraction, risk-of-bias assessment, and synthesis, largely manual (Khalil et al., 2021). While screening tools have reached relative methodological maturity, data extraction remains an active and fragmented area characterized by domain-specific prototypes with limited usability and long-term maintenance (Legate et al., 2024; Schmidt et al., 2021). Moreover, automation performance varies substantially across topics: aggressive automation strategies often reduce workload at the expense of missed relevant studies, necessitating conservative use to preserve sensitivity and undermining user trust (Gates et al., 2019; Reddy et al., 2020).

Recent machine learning and large language model–based approaches introduce further challenges. Many remain research prototypes requiring advanced technical expertise and lacking integration into stable, production-ready platforms. Reported limitations include hallucinated outputs, inconsistent predictions, reduced transparency of decision processes, and incomplete reporting of performance metrics (Chen et al., 2025; Ye et al., 2024). Consequently, most tools provide modest end-to-end efficiency gains and exhibit variable sensitivity and specificity, leaving the real-world benefits of semi-automation uncertain (Schmidt et al., 2023).

Artificial intelligence is increasingly shaping evidence synthesis workflows in response to exponential growth in the scientific literature and rising demands for timely updates, particularly in health-related fields (Blaizot et al., 2022; Ofori-Boateng et al., 2024). Techniques spanning machine learning, natural language processing, and generative AI have emerged as promising approaches to support or automate key review tasks, including searching, screening, data extraction, and risk-of-bias assessment (Jimenez et al., 2022; Ofori-Boateng et al., 2024).

Research Gap and Motivation

Despite the widespread use of systematic reviews and meta-analyses, their end-to-end workflow remains fragmented, labor-intensive, and poorly supported by integrated software solutions. Most existing tools focus on isolated stages—typically statistical synthesis or title–abstract screening—while critical tasks such as full-text data extraction, moderator coding, sensitivity analysis, and bias assessment remain largely manual. As a result, researchers rely on ad hoc combinations of spreadsheets, scripts, and standalone applications, limiting scalability, transparency, and reproducibility.

Recent advances in artificial intelligence, particularly large language models, offer opportunities to address persistent bottlenecks in evidence synthesis by automating text-intensive tasks such as PDF screening and quantitative data extraction. However, most AI-based approaches remain experimental or narrowly scoped, often embedded in research prototypes that lack robustness, validation, and integration with established meta-analytic workflows. Concerns regarding reliability, transparency, and quality control continue to hinder adoption in applied research settings.

Early AI applications primarily targeted title–abstract screening, where active learning and text classification methods have demonstrated substantial workload reductions while maintaining high recall (Jimenez et al., 2022; Qin et al., 2021). More recent tools extend AI support to topic modeling, study selection, and limited aspects of data synthesis, often within human-in-the-loop frameworks (Atkinson, 2023; Orel et al., 2023). Generative AI and LLM-based systems have further expanded possibilities, showing promising performance in controlled tasks such as semi-automated screening and data pooling. However, systematic evaluations report considerable variability in accuracy, including missed studies, misclassification, and extraction errors, underscoring the continued need for expert oversight (Clark et al., 2025).

Collectively, these findings position AI not as a replacement for expert judgment but as an augmentative technology that can reduce workload and accelerate reviews when embedded within transparent, validated, and human-supervised workflows. Addressing this gap motivates the development of integrated, production-ready platforms that combine AI-assisted processing with standard meta-analytic methods to support efficient, reproducible, and trustworthy evidence synthesis.

Related Studies

Systematic reviews and meta-analyses are central tools in evidence-based practice, providing structured and transparent methods for synthesizing large and sometimes conflicting bodies of research (Askie & Offringa, 2015; Zhai & Guyatt, 2023). By applying explicit and replicable procedures for literature searching, study selection, appraisal, and synthesis, these approaches minimize bias and yield more reliable summaries than traditional narrative reviews (Suchmacher & Geller, 2018).

Despite their importance, meta-analysis involves a complex, multistage workflow that is both time- and labor-intensive. Standard guidance describes sequential steps including question formulation, comprehensive searching, screening, data extraction and coding, model selection, heterogeneity assessment, and interpretation, each introducing methodological and operational challenges (Mikolajewicz & Komarova, 2019; Muka et al., 2019). Empirical studies indicate that conventional systematic reviews with meta-analysis often take more than a year to complete and require large multidisciplinary teams (Borah et al., 2017). Early stages are particularly burdensome, as broad searches frequently yield tens of thousands of records, of which only a small proportion are ultimately included, making screening slow and error-prone.

To mitigate these bottlenecks, prior work has proposed structured process guides and semi-automated tools, including multi-stage review roadmaps, accelerated review methodologies, and specialized software for effect-size computation and data conversion (Abbas et al., 2024; Muka et al., 2019; Scott et al., 2023). While some tasks—such as citation de-duplication and basic screening support—can be partially automated, others, including nuanced eligibility assessment, complex data extraction, and model selection, remain largely manual (Tsafnat et al., 2014).

More recently, artificial intelligence–driven approaches have been explored to support text-intensive stages of evidence synthesis, particularly screening and data extraction (Jimenez et al., 2022). However, most existing AI-based systems remain narrowly scoped or experimental and lack robust validation and integration with established meta-analytic workflows, highlighting a clear need for integrated, production-ready platforms that support end-to-end evidence synthesis.

Traditional Software for Meta-Analysis

Before the emergence of AI-assisted and web-based platforms, meta-analyses were primarily conducted using general-purpose statistical packages and specialized desktop software that focused almost exclusively on the quantitative synthesis stage. Tools such as RevMan, Stata, R (e.g., *metafor*), Comprehensive Meta-Analysis (CMA), and SPSS macros became standard solutions for computing effect sizes, fitting fixed- and random-effects models, and producing forest or funnel plots across multiple disciplines (Metelli & Chaimani, 2020). While statistically robust, these tools offered little support for upstream tasks such as searching, screening, or full-text data extraction.

Traditional software environments assume that data have already been manually curated and structured, typically in spreadsheets prepared outside the analysis platform. Consequently, labor-intensive activities—screening large numbers of records, extracting quantitative data from PDFs, harmonizing heterogeneous outcomes, and coding moderators—are handled separately using generic tools such as Excel or custom scripts (Geissbühler et al., 2021; Jeyaraj & Dwivedi, 2020). This separation between data management and statistical analysis has contributed to fragmented workflows, version-control issues, and reproducibility challenges, particularly in large or continuously updated reviews.

In addition, most traditional packages provide limited guidance for model selection, bias assessment, and sensitivity analysis. Users must independently choose among alternative models and diagnostic procedures, often without structured decision support, despite the methodological complexity involved (Cipriani et al., 2013; Esterhuizen & Thabane, 2016) . As a result, publication bias analyses and heterogeneity exploration are inconsistently applied and frequently depend on advanced statistical expertise.

Overall, traditional meta-analysis software is strong at implementing established statistical models once clean data are available, but weak at supporting the end-to-end workflow. By leaving literature management, screening, and

data extraction largely manual, these tools reinforce fragmented and spreadsheet-centric practices—limitations that motivate the development of integrated, AI-enabled platforms for evidence synthesis (Altena et al., 2019; Scott et al., 2023; Tsafnat et al., 2014).

AI-Assisted Literature Screening and Data Extraction

tasresearch.org incorporates artificial intelligence–assisted mechanisms to support two of the most labor-intensive stages of systematic reviews and meta-analyses: literature screening and quantitative data extraction. These functionalities are implemented as integrated components of a production-ready platform that is publicly accessible and designed for real-world research use rather than experimental prototyping.

For literature screening, **tasresearch.org** employs large language models to analyze full-text PDFs following upload. Rather than relying solely on title and abstract information, the screening process operates on extracted full-text content, enabling the identification of study design characteristics, outcome relevance, and the presence of quantitative data required for meta-analysis. Screening outputs are generated in a structured format that includes an inclusion or exclusion decision, a concise rationale, and an associated confidence score. All AI-generated decisions remain fully transparent to the user and can be reviewed, overridden, or corrected at any stage, ensuring that final eligibility judgments remain under expert control.

Following inclusion, **tasresearch.org** supports AI-assisted extraction of study characteristics and numerical results relevant to effect size computation. The extraction pipeline focuses on identifying sample sizes, group-level descriptive statistics (e.g., means and standard deviations), test statistics (e.g., t, F, and p values), and reported effect sizes when available. Extracted data are returned in a structured schema aligned with downstream meta-analytic modules, allowing direct use in effect size calculation, moderator analysis, and sensitivity analyses without manual reformatting. Automated validation checks are applied to detect missing values, internal inconsistencies, and implausible ranges, flagging records that require manual inspection.

Unlike many existing AI-assisted tools that address screening or extraction as isolated tasks, **tasresearch.org** integrates these capabilities within a single end-to-end workflow. AI-assisted screening, data extraction, validation, and statistical synthesis are performed within a unified platform and database context, reducing data transfer errors and enhancing reproducibility. Long-running AI tasks are executed asynchronously via a worker-based architecture, allowing users to continue interacting with the system while large document collections are processed in the background.

By deploying these AI-assisted components in a publicly available, production-ready environment, **tasresearch.org** demonstrates the practical feasibility of augmenting evidence synthesis workflows with large language models while maintaining methodological rigor, transparency, and user oversight. The implementation emphasizes AI as a supportive tool rather than a replacement for expert judgment, positioning automation as a means to reduce workload and accelerate reviews without compromising scientific standards.

Chapter 6: An AI-Integrated Platform for End-to-End Meta-Analysis: Design, Implementation, and Evaluation of "tasresearch.org"

Objectives and Contributions

This chapter presents tasresearchçorg Meta-Analysis Platform, a web-based system designed to:

1. Automate PDF processing using GPT-4 API for intelligent data extraction
2. Provide comprehensive statistical tools for all stages of meta-analysis
3. Enable scalable processing through worker-based asynchronous architecture
4. Ensure data quality with validation, duplicate detection, and quality assessment
5. Generate publication-ready outputs including tables, graphs, and APA-formatted reports
6. Support multiple disciplines (medicine, psychology, education, engineering, etc.)

Significance

tasresearch.org platform addresses three critical gaps:

- Automation Gap: First platform to integrate AI-powered PDF extraction with meta-analysis
- Accessibility Gap: Web-based interface requires no software installation or programming
- Scalability Gap: Worker-based architecture handles large-scale projects (1000+ studies)

Method

This section describes the methodological framework and system design underlying **tasresearch.org**. The platform was developed to operationalize established principles of systematic review and meta-analysis while incorporating AI-assisted components to reduce manual workload. The following subsections detail the system architecture, data processing pipeline, AI-assisted screening and extraction procedures, and the statistical methods implemented for evidence synthesis, with an emphasis on transparency, reproducibility, and human oversight.

System Architecture

tasesearch platform employs a three-tier architecture optimized for scalability, maintainability, and user experience (Figure 1):

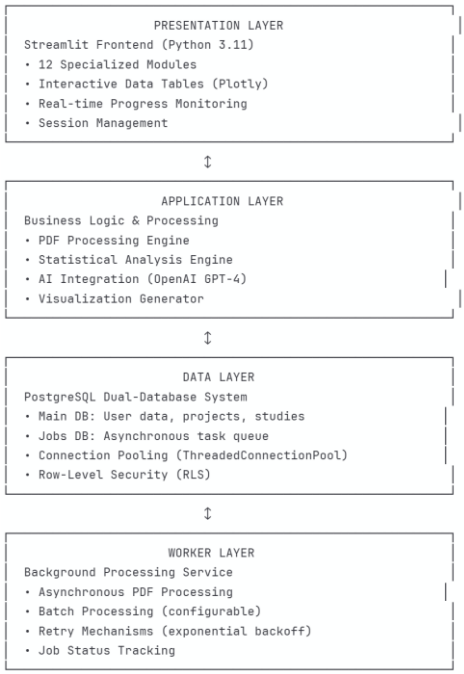


Figure 1. Overall Design Philosophy

Core Technologies

The technology stack of tasresearch.org was designed to support the complete workflow of systematic reviews and meta-analyses in an integrated and accessible manner. The platform is built on Streamlit, which enables interactive, Python-based user interfaces well suited for research-oriented data applications. This choice allows analytical processes and visual outputs to be closely coupled within a single environment.

The backend is implemented in Python 3.11, providing access to a mature scientific ecosystem for statistical analysis and artificial intelligence integration. Data storage relies on PostgreSQL, used both for core application data and background job management, ensuring transactional reliability, reproducibility, and secure multi-user access.

AI-assisted functionality is integrated through the OpenAI API, which supports full-text PDF screening and quantitative data extraction from unstructured research articles. Supporting libraries such as PyPDF2 for PDF processing, pandas for data handling, scipy and statsmodels for meta-analytic calculations, and Plotly for interactive, publication-quality visualizations collectively enable an end-to-end evidence synthesis workflow.
Overall, this technology stack reflects a deliberate emphasis on integration, scalability, and usability, allowing researchers to conduct complex meta-analyses within a single, cohesive platform.

Component	Technology	Version	Justification
Frontend Framework	Streamlit	1.39.0	Rapid development, Python-native, excellent for data apps
Backend Language	Python	3.11	Rich scientific libraries, AI/ML ecosystem
Database (Main)	PostgreSQL	15+	ACID compliance, RLS support, JSON operations
Database (Jobs)	PostgreSQL	15+	Reliable queue management, transaction support
AI/ML	OpenAI API	GPT-4	State-of-art language model for extraction
HTTP Client	httpx	0.27.0	Async support, connection pooling, timeouts
PDF Processing	PyPDF2	3.0.1	Text extraction, metadata reading
Data Analysis	pandas	2.2.0	Dataframe operations, Excel export
Statistics	scipy/statsmodels	1.11.4/0.14.1	Meta-analytic calculations
Visualization	Plotly	5.18.0	Interactive plots, publication quality

Figure 2. Core Technologies

The tasresearch.org Meta-Analysis Platform: Architecture, Functionality, and Design Philosophy

The **tasresearch.org meta-analysis platform** (https://tasresearch.org) is a web-based, AI-assisted system developed to support the complete workflow of systematic reviews and meta-analyses within a single, integrated environment. Unlike traditional approaches that rely on fragmented toolchains and extensive manual intervention, the platform is designed to consolidate literature management, data extraction, statistical analysis, visualization, and reporting into a coherent and reproducible framework.

The platform is implemented primarily in **Python (version 3.11)** using the **Streamlit** framework for the user interface and is deployed in a production environment on **Railway.app**. With an overall codebase of approximately **20,000 lines of Python code** (~975 KB), tasresearch.org reflects a mature, production-ready system rather than a proof-of-concept or experimental prototype (Figure 3).

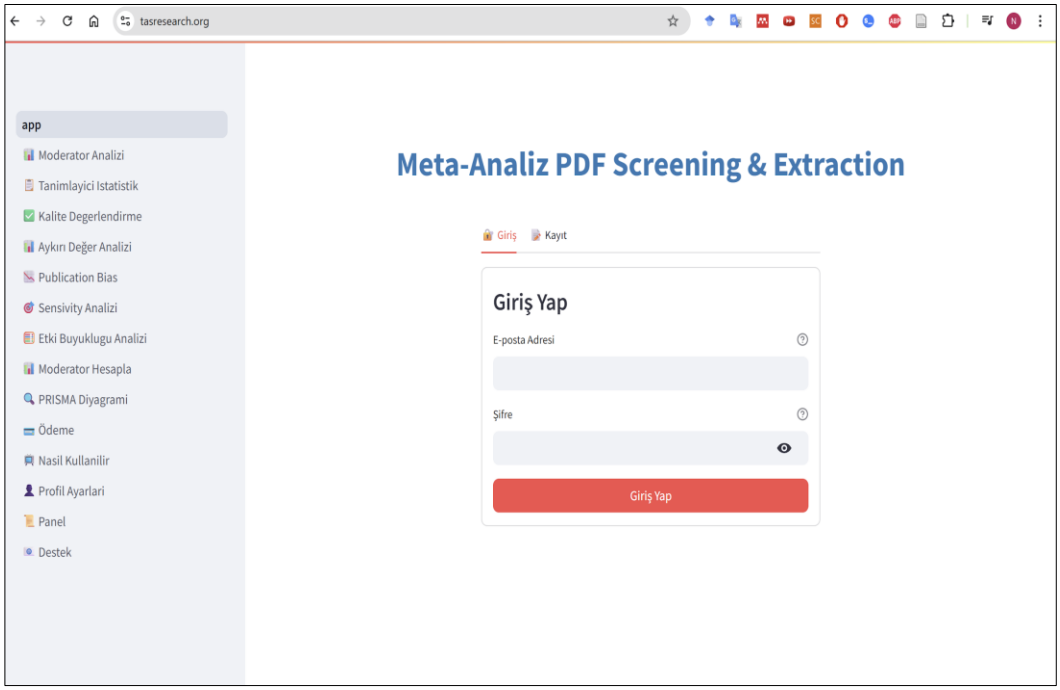


Figure 3. Home Page

Overall System Architecture

tasresearch.org adopts a modular, service-oriented architecture centered on three core components: the interactive web application layer, the database and data management layer, and an asynchronous background processing layer. This separation of concerns allows the system to remain responsive to users while performing computationally intensive tasks such as PDF processing and AI-based data extraction in the background.
At a high level, the architecture is designed to meet three primary goals:

(1) scalability to large collections of full-text studies,

(2) transparency and user oversight in AI-assisted decisions, and

(3) methodological alignment with established standards for systematic reviews and meta-analyses.

Application Layer and User Interaction

The application layer provides a browser-based interface through which users can register, authenticate, and manage meta-analysis projects. Within each project, users can upload large numbers of PDF articles (Figure x), inspect extracted study data, and invoke a wide range of analytical modules. A central dashboard summarizes project-level statistics such as the number of uploaded studies, processed PDFs, and completed analyses, supporting rapid situational awareness during long-running reviews.

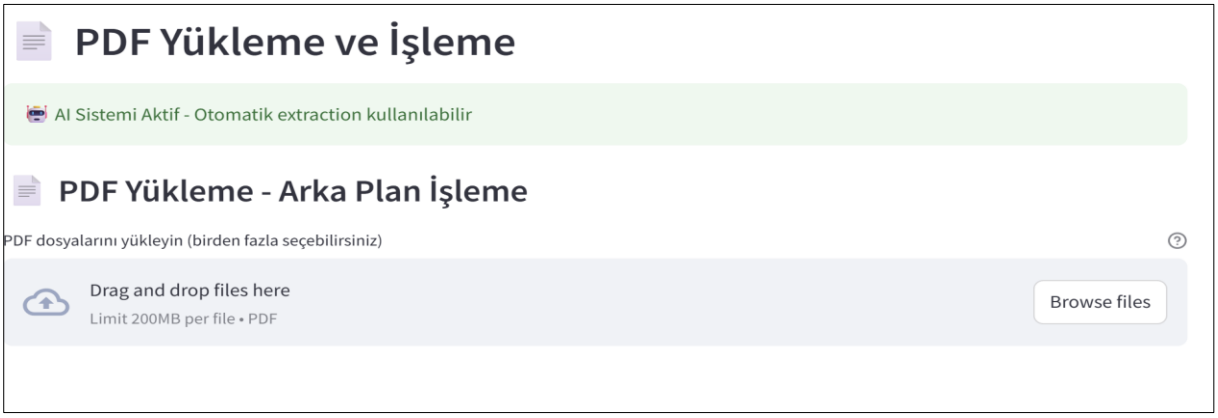


Figure 4. Pdf upload field

The interface is designed to be accessible to researchers without advanced programming expertise. All major actions—PDF upload, study inclusion decisions, moderator coding, quality assessment, and statistical analysis—are performed through graphical controls rather than scripts. At the same time, the system preserves detailed internal state and metadata to ensure reproducibility and auditability.

Data Management and Database Design

Data persistence in tasresearch.org is handled through **PostgreSQL**, which serves as the primary storage backend for users, projects, studies, comparisons, moderator variables, quality assessments, and analysis results. Each uploaded PDF corresponds to a distinct study record, which can contain multiple comparisons, moderators, and outcome measures.

To support multi-user operation and data privacy, the platform employs **row-level security (RLS)**, ensuring that users can access only their own projects and associated data. Connection pooling and retry mechanisms are implemented to improve reliability under high load, particularly when processing large batches of documents.

A key architectural feature is the use of a **separate job-oriented database** for asynchronous task management. Long-running operations, such as PDF processing and AI inference, are recorded as jobs and executed independently of the main application thread. This design prevents blocking behavior in the user interface and allows users to continue working while background tasks are running.

Asynchronous Processing and Background Workers

Computationally intensive tasks are executed by a dedicated background worker component. This worker periodically checks the job queue, retrieves pending tasks, and processes them in controlled batches. Each job typically involves extracting text from a PDF, performing AI-assisted screening, extracting structured data, and writing validated results back to the main database.

The asynchronous architecture is particularly important for real-world meta-analyses, where hundreds or thousands of PDFs may need to be processed. By decoupling document processing from user interaction, tasresearch.org maintains responsiveness and robustness even under heavy workloads.

AI-Assisted Screening and Data Extraction

One of the defining features of tasresearch.org is its integration of **large language models** to support two of the most labor-intensive stages of systematic reviews: study screening and quantitative data extraction. After text extraction from PDFs, the system applies AI-assisted analysis to identify study design, relevance to the research question, and the presence of extractable quantitative data.

For screening, the system generates structured inclusion or exclusion recommendations accompanied by brief rationales. Importantly, these recommendations are never enforced automatically; all decisions remain visible to and editable by the user. This design choice reflects a deliberate emphasis on **human oversight**, positioning AI as an assistive tool rather than an autonomous decision-maker.

For studies deemed eligible, the platform supports AI-assisted extraction of numerical data required for meta-analysis, including sample sizes, group-level descriptive statistics (means and standard deviations), test statistics (e.g., t, F, p values), and reported effect sizes. Automated validation checks are applied to flag missing or implausible values, prompting manual review when necessary.

Analytical Modules and Meta-Analytic Capabilities

tasresearch.org provides a comprehensive set of analytical modules that collectively cover the core requirements of contemporary meta-analysis. These include:

- descriptive summaries of study characteristics,
- effect size computation (e.g., standardized mean differences, odds ratios, correlations),
- fixed-effect and random-effects meta-analytic models,
- heterogeneity assessment using Q, I^2, and τ^2 statistics,
- sensitivity analyses such as leave-one-out and cumulative meta-analysis,
- moderator analyses and meta-regression for both categorical and continuous variables,
- publication bias diagnostics including funnel plots, Egger's and Begg's tests, trim-and-fill procedures, and fail-safe N,
- and automated generation of PRISMA 2020 (Figure 5) flow diagrams based on database records.

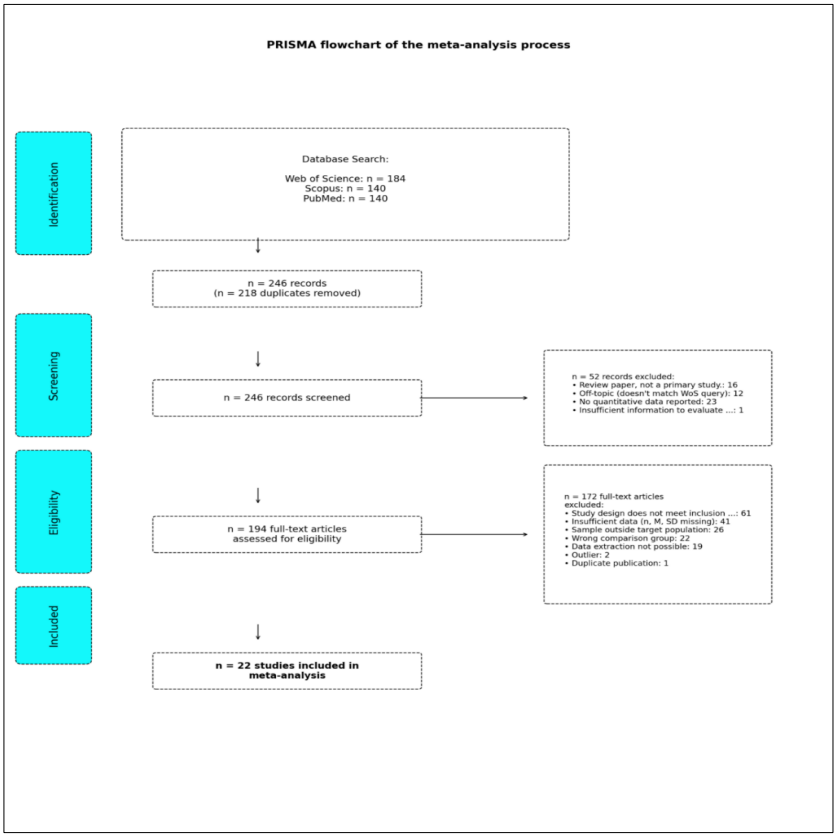


Figure 5. Prisma Example

These modules are tightly integrated, allowing outputs from one stage (e.g., extracted effect sizes) to feed directly into subsequent analyses without manual reformatting or data transfer.

Visualization and Reporting

To facilitate interpretation and dissemination of results, tasresearch.org includes extensive visualization capabilities based primarily on interactive plotting libraries. Users can generate forest plots, funnel plots, influence diagnostics, and meta-regression visualizations directly within the platform. Graphs are designed to be publication-ready and can be exported for inclusion in manuscripts and reports (Figure 6). Tabular outputs and intermediate datasets can be exported in spreadsheet formats, supporting transparency and external verification of results.

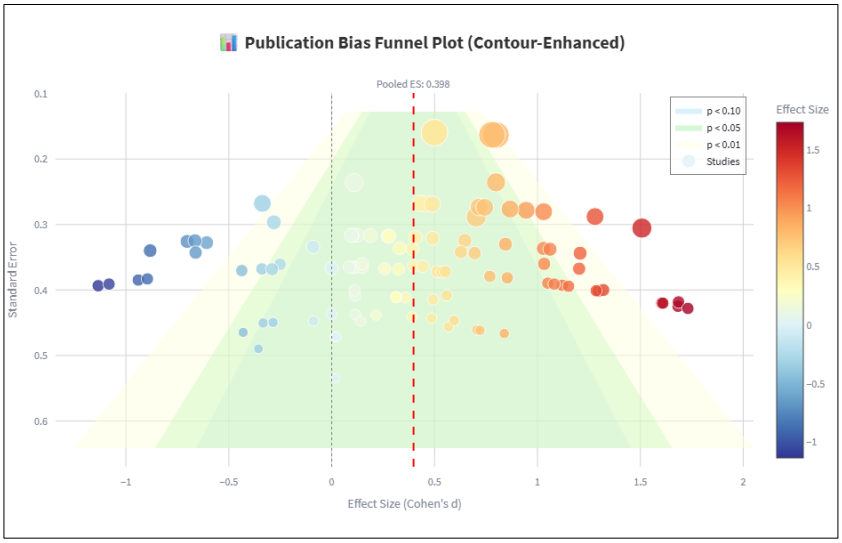


Figure 6. Publication Bias Funnel Plot Example

Deployment, Scalability, and Production Readiness

The platform is deployed in a production environment using container-based infrastructure on Railway.app. Environment-level configuration supports scalability, secure handling of credentials, and separation between development and production settings. The system has been tested on projects involving large numbers of PDFs and concurrent users, demonstrating its suitability for applied research rather than experimental use.

Conceptual Contribution

From a conceptual perspective, tasresearch.org represents a shift from tool-centric to **workflow-centric** support for evidence synthesis. Traditional meta-analysis software excels at statistical modeling once clean data are available but offers little assistance with upstream processes such as screening, extraction, and quality assessment. tasresearch.org addresses this gap by integrating AI-assisted text processing with established statistical methodology in a single, cohesive platform.

By combining automation with explicit user oversight and adherence to methodological standards, the platform illustrates how artificial intelligence can be responsibly embedded into evidence synthesis workflows—reducing workload and accelerating research without compromising scientific rigor.

To better contextualize the contribution of **tasresearch.org**, it is useful to compare the platform with widely used tools for systematic reviews and meta-analyses. Traditional software such as RevMan, Comprehensive Meta-Analysis (CMA), Stata, R (e.g., *metafor*), and JASP has long provided robust and well-validated implementations of statistical synthesis models. These tools excel at effect size computation, model estimation, and hypothesis testing once structured data are available, and they remain indispensable for advanced or highly customized analyses.

However, these platforms largely assume that upstream tasks—literature screening, full-text data extraction, moderator coding, and quality assessment—have already been completed manually. As a result, researchers typically rely on external spreadsheets, scripts, and manual workflows to prepare data prior to analysis. This separation between data preparation and statistical synthesis reinforces fragmented workflows, increases the risk of transcription errors, and limits reproducibility, particularly in large-scale or living meta-analyses.

In contrast, tasresearch.org is explicitly designed to support the **entire evidence synthesis pipeline**. Its primary distinction lies not in replacing established statistical methods, but in integrating AI-assisted PDF screening and data extraction directly into the analytic environment. By embedding these steps within a single web-based platform, tasresearch.org reduces reliance on ad hoc tools and lowers the technical barrier for conducting comprehensive meta-analyses, especially for users without advanced programming expertise.

Compared with desktop-based solutions, tasresearch.org offers several practical advantages: it requires no local installation, supports cross-device access, and enables scalable processing of large PDF collections through asynchronous worker-based architecture. Interactive visualizations (e.g., forest plots, funnel plots, PRISMA diagrams) are generated automatically from the underlying database, reducing manual effort in reporting and figure preparation.

At the same time, established statistical environments retain important strengths. Programming-based tools such as R and Stata provide greater flexibility, a wider range of effect size transformations, and access to advanced methods including network meta-analysis, Bayesian models, and multilevel structures. tasresearch.org currently implements a focused but standard set of meta-analytic techniques and therefore complements rather than replaces these ecosystems. In practice, tasresearch.org can be viewed as an upstream and midstream platform that accelerates data preparation and core analyses, while still allowing export of cleaned datasets for further modeling in specialized statistical software when needed.

Overall, the comparison highlights a shift in emphasis rather than competition: while traditional tools remain dominant at the statistical modeling layer, tasresearch.org addresses long-standing bottlenecks at the workflow and data-extraction layers. By combining AI-assisted processing with conventional meta-analytic methods in a unified environment, the platform fills a critical gap between manual review practices and fully automated, but often experimental, research prototypes.

Table 1. Comparison with other programs

Feature / Dimension	**tasresearch.org**	**RevMan**	**CMA**	**R (metafor)**	**Stata**	**JASP**
Platform Type	Web-based	Desktop	Desktop	Programming-based	Programming-based	Desktop
Installation Required	No	Yes	Yes	Yes	Yes	Yes
Programming Knowledge Required	No	No	No	Yes	Yes	No

Primary Focus	End-to-end workflow	Synthesis & reporting	Statistical synthesis	Advanced modeling	Advanced modeling	Statistical synthesis
AI-Assisted PDF Screening	Yes	No	No	No	No	No
AI-Based Data Extraction	Yes	No	No	No	No	No
Integrated Workflow (Screening → Analysis)	Yes	Partial	Partial	No	No	Partial
Effect Size Types Supported	d, g, OR, RR, r	Limited	Extensive	Extensive	Extensive	Moderate
Meta-Regression	Yes (basic)	Yes	Yes	Yes (advanced)	Yes (advanced)	Yes
Publication Bias Analyses	Yes	Limited	Yes	Yes	Yes	Yes
PRISMA Diagram Generation	Automatic	Manual	Manual	Script-based	Script-based	No
Interactive Visualizations	Yes	No	No	Partial	No	Yes
Scalability (Large PDF Sets)	High (async workers)	Limited	Moderate	Hardware-dependent	Hardware-dependent	Limited
Typical Learning Curve	Low (1–2 hours)	Moderate	Moderate	High	High	Moderate
Primary Strength	Workflow integration & automation	Standardized reporting	Rich ES formats	Methodological flexibility	Statistical power	Ease of use
Primary Limitation	Limited advanced models	Manual data prep	Closed-source, cost	Steep learning curve	Steep learning curve	Limited automation

Interpretive Summary (optional paragraph for immediately after the table)

This comparison illustrates that tasresearch.org does not aim to replace established statistical environments, but rather to complement them by addressing long-standing bottlenecks in evidence synthesis workflows. While traditional tools remain dominant for advanced statistical modeling, they provide limited support for upstream processes such as PDF screening and data extraction. tasresearch.org uniquely integrates these stages into a single, web-based environment, thereby reducing manual workload, improving transparency, and lowering the technical barrier to conducting comprehensive meta-analyses.

Conclusion

This chapter introduced **tasresearch.org**, a web-based, AI-assisted meta-analysis platform designed to support the complete workflow of systematic reviews and meta-analyses within a single, integrated environment. Unlike traditional software that primarily focuses on statistical synthesis after manual data preparation, tasresearch.org

addresses long-standing bottlenecks at earlier stages of evidence synthesis, including full-text PDF screening, quantitative data extraction, moderator coding, and workflow coordination.

By integrating large language models into a human-supervised architecture, the platform demonstrates how artificial intelligence can be responsibly embedded into evidence synthesis without compromising methodological rigor. AI-assisted screening and data extraction substantially reduce manual workload, while transparent interfaces and validation mechanisms ensure that expert judgment remains central to all critical decisions. The worker-based asynchronous architecture further enables scalable processing of large document collections, making the platform suitable for real-world meta-analytic projects rather than experimental use.

Comparative analysis indicates that tasresearch.org complements rather than replaces established statistical environments such as R, Stata, and CMA. While advanced modeling capabilities remain the domain of programming-based tools, tasresearch.org uniquely contributes by unifying upstream and midstream processes—screening, extraction, analysis, visualization, and reporting—within a single web-based system accessible to users without advanced technical expertise.

Overall, tasresearch.org illustrates a shift from tool-centric to workflow-centric support for evidence synthesis. As the volume and complexity of scientific literature continue to grow, platforms that combine AI-assisted automation with transparency, reproducibility, and expert oversight are likely to play an increasingly important role in the future of systematic reviews and meta-analyses.

References

Abbas, A., Hefnawy, M., & Negida, A. (2024). Meta-analysis accelerator: A comprehensive tool for statistical data conversion in systematic reviews with meta-analysis. *BMC Medical Research Methodology*, *24*. https://doi.org/10.1186/s12874-024-02356-6

Ahn, E., & Kang, H. (2018). Introduction to systematic review and meta-analysis. *Korean Journal of Anesthesiology*, *71*, 103–112. https://doi.org/10.4097/kjae.2018.71.2.103

Allen, J., Kim, E. K., & Jimerson, S. (2023). Meta-Analyses and Systematic Reviews Advancing the Practice of School Psychology: The Imperative of Bringing Science to Practice. *School Psychology Review*, *52*, 87–94. https://doi.org/10.1080/2372966x.2023.2178769

Altena, A., Spijker, R., Spijker, R., & Olabarriaga, S. (2019). Usage of automation tools in systematic reviews. *Research Synthesis Methods*, *10*, 72–82. https://doi.org/10.1002/jrsm.1335

Askie, L., & Offringa, M. (2015). Systematic reviews and meta-analysis. *Seminars in Fetal & Neonatal Medicine*, *20 6*, 403–409. https://doi.org/10.1016/j.siny.2015.10.002

Atkinson, C. (2023). Cheap, Quick, and Rigorous: Artificial Intelligence and the Systematic Literature Review. *Social Science Computer Review*, *42*, 376–393. https://doi.org/10.1177/08944393231196281

Blaizot, A., Veettil, S., Saidoung, P., Moreno-García, C. F., Wiratunga, N., Aceves-Martins, M., Lai, N., & Chaiyakunapruk, N. (2022). Using artificial intelligence methods for systematic review in health

sciences: A systematic review. *Research Synthesis Methods*, *13*, 353–362. https://doi.org/10.1002/jrsm.1553

Borah, R., Brown, A., Capers, P., & Kaiser, K. (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open*, *7*. https://doi.org/10.1136/bmjopen-2016-012545

Chen, H., Jiang, Z., Liu, X., Xue, C. C., Yew, S. M. E., Sheng, B., Zheng, Y., Wang, X., Wu, Y., Sivaprasad, S., Wong, T., Chaudhary, V., & Tham, Y. (2025). Can large language models fully automate or partially assist paper selection in systematic reviews? *The British Journal of Ophthalmology*, *109*. https://doi.org/10.1136/bjo-2024-326254

Cipriani, A., Higgins, J., Geddes, J., & Salanti, G. (2013). Conceptual and Technical Challenges in Network Meta-analysis. *Annals of Internal Medicine*, *159*, 130–137. https://doi.org/10.7326/0003-4819-159-2-201307160-00008

Clark, J., Barton, B., Albarqouni, L., Byambasuren, O., Jowsey, T., Keogh, J., Liang, T., Moro, C., O'Neill, H., & Jones, M. (2025). Generative artificial intelligence use in evidence synthesis: A systematic review. *Research Synthesis Methods*, *16*, 601–619. https://doi.org/10.1017/rsm.2025.16

Esterhuizen, T., & Thabane, L. (2016). Con: Meta-analysis: Some key limitations and potential solutions. *Nephrology, Dialysis, Transplantation : Official Publication of the European Dialysis and Transplant Association - European Renal Association*, *31 6*, 882–885. https://doi.org/10.1093/ndt/gfw092

Gates, A., Guitard, S., Pillay, J., Elliott, S., Dyson, M., Newton, A., & Hartling, L. (2019). Performance and usability of machine learning for screening in systematic reviews: A comparative evaluation of three tools. *Systematic Reviews*, *8*. https://doi.org/10.1186/s13643-019-1222-2

Geissbühler, M., Hincapié, C., Aghlmandi, S., Zwahlen, M., Jüni, P., & Da Costa, B. (2021). Most published meta-regression analyses based on aggregate data suffer from methodological pitfalls: A meta-epidemiological study. *BMC Medical Research Methodology*, *21*. https://doi.org/10.1186/s12874-021-01310-0

Gopal, S., Hagan, J., & Pammi, M. (2025). Meta-Analysis in Clinical Research: An Overview of the Statistics of Evidence Synthesis. *Hospital Pediatrics*. https://doi.org/10.1542/hpeds.2024-008223

Harris, J., Brand, J., Cote, M., & Dhawan, A. (2017). Research Pearls: The Significance of Statistics and Perils of Pooling. Part 3: Pearls and Pitfalls of Meta-analyses and Systematic Reviews. *Arthroscopy : The Journal of Arthroscopic & Related Surgery : Official Publication of the Arthroscopy Association of North America and the International Arthroscopy Association*, *33 8*, 1594–1602. https://doi.org/10.1016/j.arthro.2017.01.055

Jeyaraj, A., & Dwivedi, Y. (2020). Meta-analysis in information systems research: Review and recommendations. *Int. J. Inf. Manag.*, *55*, 102226. https://doi.org/10.1016/j.ijinfomgt.2020.102226

Jimenez, R. C., Lee, T., Rosillo, N., Córdova, R., Cree, I., González, Á., & Ruiz, B. I. (2022). Machine learning computational tools to assist the performance of systematic reviews: A mapping review. *BMC Medical Research Methodology*, *22*. https://doi.org/10.1186/s12874-022-01805-4

Johnson, B., & Hennessy, E. (2019). Systematic reviews and meta-analyses in the health sciences: Best practice methods for research syntheses. *Social Science & Medicine*, *233*, 237–251. https://doi.org/10.1016/j.socscimed.2019.05.035

Jung, J., & Aloe, A. (2025). Bayesian workflow for bias-adjustment model in meta-analysis. *Research Synthesis Methods*. https://doi.org/10.1017/rsm.2025.10050

Khalil, H., Ameen, D., & Zarnegar, A. (2021). Tools to support the automation of systematic reviews: A scoping review. *Journal of Clinical Epidemiology*. https://doi.org/10.1016/j.jclinepi.2021.12.005

Legate, A., Nimon, K., & Noblin, A. (2024). (Semi)automated approaches to data extraction for systematic reviews and meta-analyses in social sciences: A living review. *F1000Research*, *13*. https://doi.org/10.12688/f1000research.151493.2

Lorenc, T., Felix, L., Petticrew, M., Melendez-Torres, G., Thomas, J., Thomas, S., O'Mara-Eves, A., & Richardson, M. (2016). Meta-analysis, complexity, and heterogeneity: A qualitative interview study of researchers' methodological values and practices. *Systematic Reviews*, *5*. https://doi.org/10.1186/s13643-016-0366-6

Martínez, E. C., Hasbun, P. G., Vargas, V. S., García-González, O., Madera, M. F., Capistrán, D. R., Carmona, T. C., Cruz, C. S., & Hooper, C. T. (2025). A comprehensive guide to conduct a systematic review and meta-analysis in medical research. *Medicine*, *104*. https://doi.org/10.1097/md.0000000000041868

Metelli, S., & Chaimani, A. (2020). Challenges in meta-analyses with observational studies. *Evidence-Based Mental Health*, *23*, 83–87. https://doi.org/10.1136/ebmental-2019-300129

Mikolajewicz, N., & Komarova, S. (2019). Meta-Analytic Methodology for Basic Research: A Practical Guide. *Frontiers in Physiology*, *10*. https://doi.org/10.3389/fphys.2019.00203

Mikolajewski, D., Rojek, I., Kotlarz, P., Dorozynski, J., & Kopowski, J. (2023). Personalization of the 3D-Printed Upper Limb Exoskeleton Design-Mechanical and IT Aspects. *APPLIED SCIENCES-BASEL*, *13*(12). https://doi.org/10.3390/app13127236

Møller, M., Ioannidis, J., & Darmon, M. (2018). Are systematic reviews and meta-analyses still useful research? We are not sure. *Intensive Care Medicine*, *44*, 518–520. https://doi.org/10.1007/s00134-017-5039-y

Muka, T., Glisic, M., Milic, J., Verhoog, S., Bohlius, J., Bramer, W., Chowdhury, R., & Franco, O. (2019). A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in medical research. *European Journal of Epidemiology*, *35*, 49–60. https://doi.org/10.1007/s10654-019-00576-5

Ofori-Boateng, R., Aceves-Martins, M., Wiratunga, N., & Moreno-García, C. F. (2024). Towards the automation of systematic reviews using natural language processing, machine learning, and deep learning: A comprehensive review. *Artificial Intelligence Review*, *57*. https://doi.org/10.1007/s10462-024-10844-w

Orel, E., Ciglenecki, I., Thiabaud, A., Temerev, A., Calmy, A., Keiser, O., & Merzouki, A. (2023). An Automated Literature Review Tool (LiteRev) for Streamlining and Accelerating Research Using Natural Language Processing and Machine Learning: Descriptive Performance Evaluation Study. *Journal of Medical Internet Research*, *25*. https://doi.org/10.2196/39736

Parums, D. (2021). Editorial: Review Articles, Systematic Reviews, Meta-Analysis, and the Updated Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 Guidelines. *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research*, *27*. https://doi.org/10.12659/msm.934475

Prott, L., Carrasco-Labra, A., Gierthmuehlen, P., & Blatz, M. (2024). How to Conduct and Publish Systematic Reviews and Meta-Analyses in Dentistry. *Journal of Esthetic and Restorative Dentistry*, *37*, 14–27. https://doi.org/10.1111/jerd.13366

Qin, X., Liu, J., Wang, Y., Liu, Y.-M., Deng, K., Yu, Zou, K., Li, L., & Sun, X. (2021). Natural language processing was effective in assisting rapid title and abstract screening when updating systematic reviews. *Journal of Clinical Epidemiology*. https://doi.org/10.1016/j.jclinepi.2021.01.010

Reddy, S., Patel, S., Weyrich, M., Fenton, J., & Viswanathan, M. (2020). Comparison of a traditional systematic review approach with review-of-reviews and semi-automation as strategies to update the evidence. *Systematic Reviews*, *9*. https://doi.org/10.1186/s13643-020-01450-2

Sargeant, J., & O'Connor, A. (2020). Scoping Reviews, Systematic Reviews, and Meta-Analysis: Applications in Veterinary Medicine. *Frontiers in Veterinary Science*, *7*. https://doi.org/10.3389/fvets.2020.00011

Schmidt, L., Olorisade, B., McGuinness, L., Thomas, J., & Higgins, J. (2021). Data extraction methods for systematic review (semi)automation: Update of a living systematic review. *F1000Research*, *10*. https://doi.org/10.12688/f1000research.51117.3

Schmidt, L., Sinyor, M., Webb, R., Marshall, C., Knipe, D., Eyles, E., John, A., Gunnell, D., & Higgins, J. (2023). A narrative review of recent tools and innovations toward automating living systematic reviews and evidence syntheses. *Zeitschrift Fur Evidenz, Fortbildung Und Qualitat Im Gesundheitswesen*. https://doi.org/10.1016/j.zefq.2023.06.007

Scott, A., Glasziou, P., & Clark, J. (2023). We extended the two-week systematic review (2weekSR) methodology to larger, more complex systematic reviews: A case series. *Journal of Clinical Epidemiology*. https://doi.org/10.1016/j.jclinepi.2023.03.007

Siddaway, A., Wood, A., & Hedges, L. (2019). How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. *Annual Review of Psychology*, *70*, 747–770. https://doi.org/10.1146/annurev-psych-010418-102803

Soni, K. (2025). Critical appraisal of systematic reviews and meta-analyses. *Indian Journal of Anaesthesia*, *69*, 161–164. https://doi.org/10.4103/ija.ija_1223_24

Suchmacher, M., & Geller, M. (2018). Systematic reviews and meta-analyses. *Practical Biostatistics*. https://doi.org/10.1891/9781617052392.0028

Tsafnat, G., Glasziou, P., Choong, M., Dunn, A., Galgani, F., & Coiera, E. (2014). Systematic review automation technologies. *Systematic Reviews*, *3*, 74–74. https://doi.org/10.1186/2046-4053-3-74

Vetter, T. (2019). Systematic Review and Meta-analysis: Sometimes Bigger Is Indeed Better. *Anesthesia & Analgesia*, *128*, 575. https://doi.org/10.1213/ane.0000000000004014

Ye, A., Maiti, A., Schmidt, M., & Pedersen, S. (2024). A Hybrid Semi-Automated Workflow for Systematic and Literature Review Processes with Large Language Model Analysis. *Future Internet*, *16*, 167. https://doi.org/10.3390/fi16050167

Zhai, C., & Guyatt, G. (2023). Fixed-effect and random-effects models in meta-analysis. *Chinese Medical Journal*, *137*, 1–4. https://doi.org/10.1097/cm9.0000000000002814

Author Information

Nurullah Taş

https://orcid.org/0000-0002-8312-8733

Atatürk University

Erzurum

Türkiye

Contact e-mail: *nurullah.tas@atauni.edu.tr*

Citation

Taş, N. (2025). An AI-Integrated Platform for End-to-End Meta-Analysis: Design, Implementation, and Evaluation of "tasresearch.org". In A. Kaban & T. Demirel (Eds.), *Interdisciplinary Applications of Artificial Intelligence - 1* (pp. 115-136). ISTES.

Chapter 7- Artificial Intelligence in Instructional Design and Technologies

Ömer Bilen

Chapter Highlights

- Artificial Intelligence (AI) is increasingly integrated into instructional design and technologies, enhancing personalized learning experiences.
- AI-driven tools enable adaptive content delivery, improving learner engagement and knowledge retention.
- The use of AI supports the automation of routine instructional tasks, allowing educators to focus on higher-level pedagogical strategies.
- AI facilitates real-time feedback and assessment, promoting continuous learner improvement and self-regulation.
- Ethical considerations and challenges, such as data privacy and algorithmic bias, are critical in implementing AI in education.
- Collaboration between AI systems and human instructors optimizes instructional effectiveness and learner outcomes.
- Ongoing research is essential to refine AI applications in instructional design, ensuring they meet diverse educational needs.

Introduction

Instructional design is a multidisciplinary field dedicated to systematically enhancing the effectiveness, efficiency, and engagement of learning experiences (ANDREEA, 2022; Trif-Boia, 2022). It involves a structured approach to identifying learning objectives, developing materials, selecting appropriate strategies, and evaluating outcomes, all while considering diverse factors such as learner needs, content characteristics, pedagogical approaches, and available technological resources (Francesa-Alfaro et al., 2016; Ismail & Alkhazali, 2019). In an era of rapid digital transformation, the landscape of education has evolved significantly, moving beyond traditional classrooms to embrace online, blended, and hybrid models. This shift, driven by advancements like the internet, mobile devices, and cloud services, necessitates a redefinition of instructional designers' roles, requiring them to skillfully navigate and design learning experiences within complex digital ecosystems (OECD, 2023; Ибрагимов & Kalimullina, 2020).

Central to this digital evolution is the pervasive influence of Artificial Intelligence, which has emerged as a transformative force in instructional design. AI's capabilities, particularly those of generative AI, offer unprecedented potential for personalizing learning experiences, optimizing design processes, and automating time-consuming tasks such as content creation and assessment (Lan & Zhou, 2025; Luo et al., 2025; Song, 2024). Generative AI models can produce novel text, visuals, interactive simulations, and even entire lesson plans, thereby accelerating material development and enabling richer, more adaptive learning environments (Arslan et al., 2024; Giannakos et al., 2024; Rouabhia, 2024). This transition from rule-based AI systems to generative AI marks a significant shift, empowering instructional designers to craft dynamic and highly individualized educational pathways.

However, the integration of AI into instructional design also presents critical pedagogical, ethical, and design challenges. Issues such as automation bias, the risk of pedagogical incompatibility, algorithmic biases, and concerns around data privacy and security demand careful consideration (Bulut et al., 2024; Li et al., 2025; Zhu et al., 2025). Addressing these limitations necessitates a human-centered design approach, ensuring that AI serves to augment human capabilities rather than replace them, and that technology aligns with core educational principles. Consequently, the role of the instructional designer is being reshaped; they are becoming crucial collaborators with AI, responsible for directing human-AI interactions, fostering AI literacy, and upholding ethical standards to maximize technology's benefits while mitigating its risks (Amado-Salvatierra et al., 2023; Burneo-Arteaga et al., 2025; “September 2024 Full Issue,” 2024).

A Conceptual Introduction to Instructional Design

Instructional design is an interdisciplinary field that employs a systematic approach to enhance the effectiveness, efficiency, and engagement of learning and teaching processes. Its fundamental purpose is to devise learning experiences that facilitate the optimal acquisition of knowledge, skills, and attitudes by learners (ANDREEA, 2022; Trif-Boia, 2022). This process takes into account various factors, including learner needs, content

characteristics, pedagogical approaches, and available technological resources (Francesa-Alfaro et al., 2016). It can be understood as a planned and structured methodology encompassing the identification of learning objectives, the development of materials, the selection of strategies, and the evaluation of outcomes. Fundamentally, instructional design is both a process and a system that involves a series of activities aimed at optimizing learning processes (Francesa-Alfaro et al., 2016), and it can also be seen as the practical application of scientific theories and principles in education and instruction (ANDREEA, 2022). Beyond just developing materials, it involves the holistic planning of the learning environment, interactions, and assessment methods (Ismail & Alkhazali, 2019).

The relationship between instructional design and instructional technologies is dynamic and mutually reinforcing (Mondal, 2025). Instructional technologies provide powerful tools for implementing designed learning experiences, such as learning management systems, virtual reality applications, online interactive platforms, and simulations (Chimalakonda & Nori, 2021). Instructional designers can integrate the possibilities offered by technology in alignment with pedagogical objectives, thereby overcoming the limitations encountered in traditional learning environments and creating richer learning opportunities (Brown et al., 2019). Emerging educational technologies, including learning analytics, artificial intelligence, mixed realities, and adaptive learning systems, promise to deliver more complex and personalized learning experiences (Spector, 2015). Conversely, the systematic approach of instructional design ensures that technology integration is pedagogically meaningful and effective, preventing the arbitrary use of technology (Hanshaw & Hanson, 2019). In this context, technology serves as a tool to achieve instructional design goals, with technological innovations continuously shaping instructional design approaches (West, 2018).

Definition and Purpose of Instructional Design

Instructional design is a planned and structured approach encompassing the processes of determining learning objectives, developing materials, selecting strategies, and evaluating outcomes (ANDREEA, 2022; Trif-Boia, 2022). Its fundamental purpose is to design learning experiences that enable learners to acquire knowledge, skills, and attitudes in the most appropriate manner. This process takes into account numerous factors, including learner needs, content characteristics, pedagogical approaches, and available technological resources (Francesa-Alfaro et al., 2016). Instructional design is also defined as both a process and a system involving a series of activities aimed at optimizing learning processes (Francesa-Alfaro et al., 2016). According to another definition, instructional design can be seen as the practical application of scientific theories and principles in education and instruction (ANDREEA, 2022). The design not only involves material development but also the holistic planning of the learning environment, interactions, and assessment methods (Ismail & Alkhazali, 2019).

Instructional Design – Instructional Technologies Relationship

The relationship between instructional design and instructional technologies is dynamic and mutually reinforcing (Mondal, 2025). Instructional technologies provide powerful tools for implementing designed learning

experiences, such as learning management systems, virtual reality applications, online interactive platforms, and simulations (Chimalakonda & Nori, 2021). Instructional designers have the potential to overcome limitations encountered in traditional learning environments and create richer learning opportunities by integrating the possibilities offered by technology in alignment with pedagogical objectives (Brown et al., 2019). Emerging educational technologies, including learning analytics, artificial intelligence, mixed realities, and adaptive learning systems, promise to deliver more complex and personalized learning experiences (Spector, 2015). Conversely, the systematic approach of instructional design ensures that technology integration is pedagogically meaningful and effective, preventing arbitrary use of technology (Hanshaw & Hanson, 2019). In this context, technology serves as a tool to achieve instructional design goals, with technological innovations continuously shaping instructional design approaches (West, 2018).

Digital Transformation and the Evolution of Learning Environments

In today's world, digital transformation is causing profound changes in every area of education (OECD, 2023a, 2023b). Digitalization refers to the transition from analog data transmission to digital format, representing information more accurately and ensuring its free circulation, processing, and application in networks (Ибрагимов & Kalimullina, 2020). The widespread adoption of the internet, the development of mobile devices, and the rise of cloud-based services have extended learning environments beyond physical classrooms (Hetmańczyk, 2023). Traditional face-to-face education models have been enriched—and even transformed in some cases—by online, blended, and hybrid learning models (Arısoy, 2022). Digital tools and platforms enable learners to access content anytime and anywhere, interact with peers and instructors, and follow personalized learning paths (Rajasekaran et al., 2024; Zou et al., 2025). This evolution is also redefining the roles of instructional designers; they are now required not only to develop content but also to possess the skills to design and manage effective learning experiences in complex digital ecosystems (Godin & Terekhova, 2021). Artificial intelligence technologies, as one of the primary driving forces of this digital transformation, hold the potential to make learning environments smarter and more adaptive (Human, Technologies and Quality of Education, 2023, 2023; Sgay et al., 2025; Song, 2024). AI applications are increasingly integrated into higher education, offering benefits such as personalized learning experiences, real-time feedback, and greater flexibility (Lan & Zhou, 2025; Luo et al., 2025).

Instructional Design Models and Digital Learning

Instructional design is an interdisciplinary field(Luo et al., 2024) that provides a systematic approach(Senadheera et al., 2024; Wintersberg & Pittich, 2025) aimed at making learning and teaching processes more effective, efficient, and engaging(Godsk & Møller, 2024; Luo et al., 2024). This section will examine the fundamental principles of instructional design(Wintersberg & Pittich, 2025), its relationship with instructional technologies(Abuhassna & Alnawajha, 2023; Özkan et al., 2025), and the effects of digital transformation on learning environments(Mulenga & Shilongo, 2024; OECD, 2023) within a conceptual framework before

proceeding to the integration of artificial intelligence into instructional design(Amado-Salvatierra et al., 2023; Luo et al., 2025).

Definition and Purpose of Instructional Design

Instructional design is a planned and structured approach encompassing the processes of determining learning objectives, developing materials, selecting strategies, and evaluating outcomes (ANDREEA, 2022; Trif-Boia, 2022). Its fundamental purpose is to design learning experiences that enable learners to acquire knowledge, skills, and attitudes in the most appropriate manner. This process takes into account numerous factors, including learner needs, content characteristics, pedagogical approaches, and available technological resources (Francesa-Alfaro et al., 2016). Instructional design is also defined as both a process and a system involving a series of activities aimed at optimizing learning processes (Francesa-Alfaro et al., 2016). According to another definition, instructional design can be seen as the practical application of scientific theories and principles in education and instruction (ANDREEA, 2022). The design not only involves material development but also the holistic planning of the learning environment, interactions, and assessment methods (Ismail & Alkhazali, 2019).

Instructional Design – Instructional Technologies Relationship

The relationship between instructional design and instructional technologies is dynamic and mutually reinforcing (Mondal, 2025). Instructional technologies provide powerful tools for implementing designed learning experiences, such as learning management systems, virtual reality applications, online interactive platforms, and simulations (Chimalakonda & Nori, 2021). Instructional designers have the potential to overcome limitations encountered in traditional learning environments and create richer learning opportunities by integrating the possibilities offered by technology in alignment with pedagogical objectives (Brown et al., 2019). Emerging educational technologies, including learning analytics, artificial intelligence, mixed realities, and adaptive learning systems, promise to deliver more complex and personalized learning experiences (Spector, 2015). Conversely, the systematic approach of instructional design ensures that technology integration is pedagogically meaningful and effective, preventing arbitrary use of technology (Hanshaw & Hanson, 2019). In this context, technology serves as a tool to achieve instructional design goals, with technological innovations continuously shaping instructional design approaches (Luo et al., 2024).

Digital Transformation and the Evolution of Learning Environments

In today's world, digital transformation is causing profound changes in every area of education (OECD, 2023). Digitalization refers to the transition from analog data transmission to digital format, representing information more accurately and ensuring its free circulation, processing, and application in networks (Ибрагимов & Kalimullina, 2020). The widespread adoption of the internet, the development of mobile devices, and the rise of cloud-based services have extended learning environments beyond physical classrooms (Hetmańczyk, 2023). Traditional face-to-face education models have been enriched—and even transformed in some cases—by online,

blended, and hybrid learning models (Arısoy, 2022). Digital tools and platforms enable learners to access content anytime and anywhere, interact with peers and instructors, and follow personalized learning paths (Rajasekaran et al., 2024; Zou et al., 2025). This evolution is also redefining the roles of instructional designers; they are now required not only to develop content but also to possess the skills to design and manage effective learning experiences in complex digital ecosystems (Godin & Terekhova, 2021). Artificial intelligence technologies, as one of the primary driving forces of this digital transformation, hold the potential to make learning environments smarter and more adaptive (Sgay et al., 2025; Song, 2024). AI applications are increasingly integrated into higher education, offering benefits such as personalized learning experiences, real-time feedback, and greater flexibility (Lan & Zhou, 2025; Luo et al., 2025).

Integration of Artificial Intelligence into Instructional Design

Instructional design is a systematic approach aimed at improving learning and teaching processes (Amado-Salvatierra et al., 2023). Today, rapid developments in artificial intelligence technologies offer transformative potential in this field (Amado-Salvatierra et al., 2023; Choi et al., 2024). The integration of AI into instructional design is critically important for personalizing learning experiences, increasing efficiency, and optimizing design processes (Luo et al., 2025; McNeill, 2024; Merino-Campos, 2025).

Why is Artificial Intelligence Critical in Instructional Design?

Artificial intelligence plays a critical role in instructional design for various reasons. Firstly, it has the capacity to analyze learners' individual needs, learning styles, and progress to offer personalized learning paths (Sgay et al., 2025; Song, 2024). This personalization provides an adaptation level that is difficult to achieve with traditional teaching methods. Secondly, AI can process large data sets to provide learning analytics, thereby enabling designers to gain a deeper understanding of the effectiveness of instructional materials and strategies (Luo et al., 2025). Thirdly, AI-based tools can automate time-consuming tasks such as content creation, assessment, and feedback provision, reducing the workload on instructional designers and allowing them to focus on more creative and strategic tasks (Lan & Zhou, 2025). One of the primary reasons for using AI in education is the increasing complexity of knowledge and the inadequacy of current education systems in managing this complexity. AI can also enhance teacher-student interactions, assist educators, and improve the quality of learning materials. By exhibiting human-like intelligence in education, artificial intelligence has the potential to enhance the learning and teaching process, making it more efficient and personalized. These capabilities make AI an indispensable tool for the instructional design discipline.

Transition from Rule-Based Systems to Generative AI

The evolution of artificial intelligence in instructional design represents a significant transition from initial rule-based systems to today's generative artificial intelligence models (Guettala et al., 2024; Yang & Zhang, 2024). Traditional AI systems operated within specific rules and algorithms to perform predefined tasks (Hao et al., 2025;

Zhang et al., 2024). For example, simple adaptive learning systems could offer different learning paths within predefined scenarios based on students' responses (Maity & Deroy, 2024). However, these systems were limited in terms of flexibility and creativity (AlAli & Wardat, 2024; Kılınç & Keçecioğlu, 2024).

Generative artificial intelligence, on the other hand, has overcome these limitations with its ability to produce new and original content (Arslan et al., 2024; Giannakos et al., 2024). Large language models and other generative models can generate text, images, audio, and even code (Choi et al., 2024). This offers revolutionary innovations in instructional design (Choi et al., 2024). Generative AI can create personalized learning materials, lesson scenarios, assessment questions, and even interactive simulations tailored to students' needs (Rouabhia, 2024; Y et al., 2024). This transition allows instructional designers to create entirely new and dynamic learning experiences rather than just adapting existing materials. Generative artificial intelligence is a type of AI capable of creating human-like texts, images, or other media formats based on user-provided inputs (Guettala et al., 2024).

Positions of AI in the Design Process

Artificial intelligence can be effectively positioned at every stage of the instructional design process:

- **Analysis**: AI can automate and deepen processes such as learner analysis, content analysis, and context analysis. By analyzing data such as learners' prior knowledge levels, learning preferences, and motivations, it can help create more accurate learner profiles (Song, 2024). Thanks to trends and patterns obtained from large data sets, instructional needs can be determined more precisely (Luo et al., 2025). For example, AI can analyze user interactions on learning platforms to identify topics where students struggle or effective learning strategies (Lan & Zhou, 2025).
- **Design**: AI-supported tools can be used in stages such as determining instructional objectives, selecting learning strategies, and designing assessment methods. AI can suggest the most appropriate pedagogical approaches for specific learning objectives or draft lesson plans according to different scenarios (Song, 2024). Additionally, it can contribute to personalized instructional design by suggesting different learning paths or activities based on learners' characteristics and needs (Sgay et al., 2025).
- **Development**: In this stage, AI plays a significant role in creating learning materials (Luo et al., 2024; mostafa et al., 2024). Generative AI models can produce text-based lesson content, question banks, scenarios, visual materials, and even simple interactive learning objects (Choi et al., 2024; Li et al., 2024; Rouabhia, 2024). This accelerates the material development process and enables designers to produce content in different formats and languages (Choi et al., 2024; Y et al., 2024). AI-supported software can facilitate the production of rich media content such as videos, animations, and simulations (Luo et al., 2024; Netland et al., 2024; Zhang, 2025).
- **Evaluation**: AI also offers powerful tools for evaluating learning outcomes and measuring instructional effectiveness. Automatic exam grading, written assignment analysis, and feedback systems make the evaluation process more efficient (Luo et al., 2025). AI-based learning analytics can continuously monitor student performance data to identify the strengths and weaknesses of instructional programs.

This way, instructional designers can continuously improve their programs and provide students with more timely and personalized feedback (Lan & Zhou, 2025).

AI-Supported Instructional Design Applications

Artificial intelligence technologies offer innovative applications at various stages of instructional design (Choi et al., 2024; Luo et al., 2024), holding the potential to transform learning experiences (Bolick & Silva, 2023). AI-supported applications make learning more personalized(Luo et al., 2025; Merino-Campos, 2025), adaptive, and efficient while providing instructional designers with new tools to optimize their processes (McNeill, 2024).

Personalized Learning Scenarios

AI-supported instructional design enables the creation of personalized learning scenarios tailored to students' individual needs and learning styles (Song, 2024). AI algorithms can analyze the student's previous performance data, learning pace, preferred learning environments, and even cognitive characteristics to recommend the most suitable learning path (Luo et al., 2025; Sgay et al., 2025). This personalization can increase students' motivation and improve learning outcomes (Ifraheem et al., 2024; Kestin et al., 2025; Wang et al., 2024). For example, intelligent tutoring systems can present materials at different difficulty levels according to the student's proficiency in a subject or provide special supplementary materials by identifying deficient areas (Hariyanto et al., 2025; Merino-Campos, 2025). This approach enables each student to progress at their own pace and according to their interests, making the learning process more meaningful (Nguyen et al., 2023; Xiao, 2024). To create individualized instructional experiences, AI continuously evaluates data such as student performance, interactions, and preferences, dynamically adapting learning materials, activities, or feedback (Bayly-Castaneda et al., 2024; Hariyanto et al., 2025; Plooy et al., 2024).

Intelligent Content Generation

Artificial intelligence plays an important role in the processes of creating and updating instructional materials (Choi et al., 2024). Generative artificial intelligence models can produce various learning materials, such as text, visuals, interactive simulations, or video content, on a specific topic (Chiu, 2024; Rouabhia, 2024). This helps instructional designers accelerate their content development processes and present richer materials in diverse formats (Dickey & Bejarano, 2023; mostafa et al., 2024). Intelligent content production systems can automatically adapt content to different learner profiles or needs in various languages (Guettala et al., 2024; Y et al., 2024). Additionally, by tracking current information, they ensure the content remains continuously relevant and accurate (Team et al., 2025). Such systems provide significant efficiency increases, especially for large-scale courses or topics requiring continuous updates (Yao et al., 2025).

Learning Analytics and Adaptive Systems

AI-supported learning analytics provide valuable insights by analyzing large data sets collected from learning processes (Lan & Zhou, 2025). These insights offer information on students' learning behaviors, points of difficulty, success rates, and interaction patterns. Instructional designers can use these analytics to evaluate the effectiveness of instructional strategies and identify areas for improvement (Luo et al., 2025). Adaptive learning systems modify the learning experience in real-time using this analytical data. For example, when a student struggles with a specific topic, the system can automatically offer additional exercises or alternative explanatory materials. These systems enhance the effectiveness of learning processes by providing instant feedback to students and dynamically adjusting learning paths.

AI-Supported Tools for Instructional Designers

Artificial intelligence offers instructional designers a range of tools and platforms to support their workflows (McNeill, 2024). These tools enhance designers' productivity through task automation, content creation assistance, and decision-making support (Bolick & Silva, 2023).

- **Automatic Feedback Systems**: Provide designers with rapid feedback on materials or generate feedback for student work (Luo et al., 2025).
- **Lesson Plan and Learning Objective Generators**: AI algorithms can create draft lesson plans for specific topics and learner profiles or suggest appropriate learning objectives (Luo et al., 2024).
- **Assessment Tool Assistants**: Assist with tasks such as creating question banks, automatic exam analysis, and rubric development (Luo et al., 2025).
- **Data Analysis and Visualization Tools**: Help designers make more informed decisions by interpreting data from learning analytics (Lan & Zhou, 2025).
- **Project Management and Collaboration Platforms**: Optimize project timelines with AI-supported features and enhance collaboration among design teams.

These AI-supported tools enable instructional designers to focus on more complex and creative tasks while reducing routine and time-consuming workloads (McNeill, 2024).

Generative Artificial Intelligence in Instructional Technologies

Generative artificial intelligence technologies offer groundbreaking innovations in the field of education and instructional technologies, reshaping content creation processes, learning experiences, and the roles of instructional designers (Choi et al., 2024; Luo et al., 2024; Qian, 2025; Zhang et al., 2024). This section examines the main applications of generative AI in instructional technologies and the transformations brought by these technologies.

Text, Visual, Video, and Interactive Content Production

Generative artificial intelligence offers significant automation and personalization potential in the development of instructional materials. Text-based generative AI models can rapidly generate content such as lecture notes, explanatory texts, case studies, summaries, and exam questions on specific topics (Song, 2024). This capability greatly accelerates the content development process while enabling the production of variations suitable for different learning levels or languages.

Generative AI has also achieved notable advancements in visual and video content production. AI algorithms can create images, infographics, diagrams, and even scripts or animations for educational videos based on text descriptions. This reduces the cost and time required to produce visually rich and interactive learning materials (Sgay et al., 2025). The ability to generate 3D models or scenarios for virtual and augmented reality environments can provide students with more immersive and experiential learning opportunities. Interactive content production—such as simulations, gamified learning modules, or dialogue-based learning assistants—has become easier and more scalable with generative AI. AI can design adaptive materials that respond to student interactions in real-time and dynamically adjust the learning path (Lan & Zhou, 2025).

Lesson Plan, Learning Outcome, and Assessment Tool Creation

Generative artificial intelligence can assist instructional designers and educators in optimizing their lesson planning processes (Berg, 2024; Chan & Colloton, 2024). Based on a specific curriculum or learning objectives, AI can generate lesson plans, unit contents, and activity suggestions (Huang et al., 2025; Li et al., 2024). Learning outcomes and achievements can be automatically formulated according to the specified learning objectives, helping to establish consistent and measurable targets (Doyle et al., 2025; Sridhar et al., 2023).

The creation of measurement and evaluation tools is also an important application area for generative AI. AI models can produce exam questions in various formats, such as multiple-choice, true-false, fill-in-the-blank, or open-ended questions (Jukiewicz, 2025; Luo et al., 2025; Namoun et al., 2024). Additionally, they can design rubrics and scoring criteria to evaluate student responses (Chan & Colloton, 2024). These capabilities increase the diversity of assessment materials while reducing the time burden on educators in this process (Namoun et al., 2024). AI can also create adaptive tests to detect learners' progress and deficiencies, thereby making the assessment process more personalized (Arslan et al., 2024; Xia et al., 2024).

Transformation of the Teacher-Designer Role

The integration of generative artificial intelligence into instructional technologies brings about a significant transformation in the roles of teachers and instructional designers (Zhaı, 2024). By automating routine and repetitive tasks, AI tools enable these professionals to focus on higher-level thinking, creativity, and strategic planning (McNeill, 2024). For example, AI assistance in content creation and lesson planning allows designers to

devote more time to developing pedagogical strategies, deeply analyzing student needs, and designing innovative learning experiences (Luo et al., 2024; Song, 2024).

This transformation requires teachers and designers to evolve beyond being mere knowledge transmitters or content creators, becoming curators, guides, and facilitators of learning experiences (Giannakos et al., 2024). They are expected to continuously improve learning processes based on AI-provided data, design personalized learning paths, and integrate technology within ethical principles. Coping with AI's complexity makes AI literacy and the critical use of AI-based tools essential for teachers and designers (Brandão et al., 2024). This necessitates acquiring new skills and continuous adaptation in professional development (Luo et al., 2025; Zhaı, 2024).

Pedagogical, Ethical, and Design Limitations

While the integration of artificial intelligence and generative AI into instructional design offers great potential, the pedagogical, ethical, and design limitations and challenges brought by these technologies—such as automation bias, algorithmic biases, lack of pedagogical foundations, data privacy issues, and black-box opacity—must also be carefully addressed (Bulut et al., 2024; Li et al., 2025; Zhu et al., 2025). These limitations can directly affect the effectiveness, fairness, and acceptability of AI-based systems by perpetuating inequities, eroding human oversight, and misaligning with educational principles (Bozkurt et al., 2024; Trevisan et al., 2024).

Automation Bias

Automation bias refers to the tendency of people to overly trust recommendations from automated systems or to prioritize AI decisions over human judgment (Bulut et al., 2024). In instructional design, this can lead to serious pedagogical consequences (Li et al., 2025). Instructional designers and educators may accept lesson plans, content, or evaluation methods suggested by AI algorithms without critical evaluation (Bulut et al., 2024). This can cause any algorithmic biases or incomplete information in AI to reflect in the learning experience (Zhu et al., 2025). AI systems typically reflect patterns and biases from their training datasets, which can result in less effective or unfair learning paths for certain student groups (Bozkurt et al., 2024). Excessive automation bias can lead to the neglect of human expertise and the weakening of critical pedagogical thinking (Li et al., 2025). For example, a personalized learning path suggested by AI may overlook the student's actual needs or emotional state, thereby reducing the student's motivation (Zhai et al., 2024).

Risk of Pedagogical Incompatibility

In integrating AI-based tools into instructional design, there is a risk that the technology may not align with pedagogical goals (Li et al., 2025; Sajja et al., 2025). Although AI offers new possibilities, using it merely because it exists can overlook learning objectives (Koh et al., 2023). Pedagogical incompatibility arises when AI applications conflict with educational philosophy, learning theories, or teaching strategies (“Higher Education for Good,” 2023; Hooshyar et al., 2025). For instance, in a course requiring complex problem-solving skills, using

AI primarily as a tool for information transmission may hinder students' deep learning (Bauer et al., 2025). Moreover, the design of AI systems may not be based on a specific learning theory or may fail to support diverse pedagogical approaches (Hooshyar et al., 2025). This can render the "personalization" or "adaptation" provided by AI superficial, failing to deliver genuine learning benefits (Koh et al., 2023; Verghis et al., 2025). Effective AI integration requires designing the technology to align with pedagogical principles and serve learning objectives (Li et al., 2025; Luo et al., 2025).

Ethical Responsibilities in Instructional Design

The use of artificial intelligence in instructional design entails a series of important ethical responsibilities (Hong, 2024; Zhu et al., 2025). Data privacy and security are foremost among them (Hong, 2024; Huang, 2023). AI systems collect and process students' personal and learning data; misuse, breach, or sharing of this data without consent can lead to serious ethical issues (Ahmad et al., 2023; Holmes et al., 2023). Concerns about student surveillance and monitoring are also central to ethical debates (Huang, 2023; Versteijlen & Groesbeek, 2024). AI systems can analyze students' behaviors, performance, and even emotional states, potentially causing privacy violations and a sense of insecurity in the learning environment (Holmes et al., 2023).

Ethical issues concerning student autonomy and representation are equally significant (Bozkurt et al., 2024; Verghis et al., 2025). AI determining learning paths or influencing student decisions can diminish students' autonomy in their own learning processes (Verghis et al., 2025; Versteijlen & Groesbeek, 2024). The opacity of AI algorithms and the inexplicability of their decisions raise questions about the system's fairness and reliability (Verghis et al., 2025; Zhu et al., 2025). Instructional designers are obligated to adhere to principles of transparency, accountability, fairness, and data protection in the design and implementation of AI systems (Hong, 2024; Ifenthaler et al., 2024).

The Necessity of Human-Centered Design

In integrating artificial intelligence into instructional design, adopting a human-centered approach is critically important (Alfredo et al., 2024; Li et al., 2025). This approach aims to ensure that technology serves human needs, values, and abilities (Alfredo et al., 2024). Human-centered design seeks to make AI systems user-friendly, preserve educators' and students' sense of control, and enhance human interaction (Maity & Deroy, 2024; Roméro et al., 2024). The role of AI in instructional processes should not replace human instructional designers and educators but rather enhance and support their capacities (Li et al., 2025).

Human-centered design incorporates continuous user feedback and collaboration in the development of AI-based systems (Hu & Chan, 2025; Sperling et al., 2023). This allows systems to be tailored to real-world contexts and pedagogical requirements (Li et al., 2025). The goal is to enable students and educators to collaborate with AI to create richer and more meaningful learning experiences, thereby maximizing technology's potential benefits while minimizing risks (Atchley et al., 2024). Human-centered design plays a pivotal role in overcoming AI's ethical and pedagogical limitations in education (Alfredo et al., 2024; Li et al., 2025).

Future Perspective: The New Role of the Instructional Designer

The integration of artificial intelligence and generative artificial intelligence technologies into educational environments is redefining the roles, responsibilities, and competencies of instructional designers (Amado-Salvatierra et al., 2023; Kumar et al., 2024; Luo et al., 2024; “September 2024 Full Issue,” 2024). In the future, instructional designers will be key figures working alongside AI, directing human-AI collaboration in design decisions, and shaping the curricula of education faculties in accordance with this transformation (Amado-Salvatierra et al., 2023; Burneo-Arteaga et al., 2025; Costello et al., 2024; Hilali & Chergui, 2025).

Instructional Designer Working with AI

In the age of artificial intelligence, the role of the instructional designer is undergoing significant change as AI transitions from being merely a supplementary tool to a creative partner (Amado-Salvatierra et al., 2023). Instructional designers are taking on the role of "glue" that brings together technology, content, and pedagogical best practices (“September 2024 Full Issue,” 2024). The emergence of generative AI tools is leading to rapid changes in the job responsibilities of instructional designers (Luo et al., 2024). Instructional designers are now using generative AI tools in various instructional design tasks, such as idea generation, performing low-risk tasks, facilitating design processes, and enhancing collaborations (Luo et al., 2024). This situation demonstrates that they play an important role in transforming AI into beneficial applications for teaching and learning that consider ethical parameters (“September 2024 Full Issue,” 2024). Future instructional designers should possess the competency to design more complex and personalized learning experiences using emerging tools such as learning analytics, artificial intelligence, mixed realities, and adaptive learning systems (Luo et al., 2025).

Human–AI Collaboration in Design Decisions

Collaboration between humans and artificial intelligence is becoming a fundamental element of modern instructional design practice. It is stated that AI enhances human capabilities in complex instructional tasks and sometimes even surpasses them (Luo et al., 2025). However, this collaboration requires human oversight and cooperation to ensure pedagogical appropriateness (Hilali & Chergui, 2025). The task of AI is to support humans in the decision-making process, not to replace human decision-makers (Ren et al., 2024). Particularly in fields like finance, while AI tools analyze large datasets in real time, the responsibility of human experts to interpret these insights and make strategic decisions continues (Paiva et al., 2025).

Human-AI collaboration in education can help teachers determine which pedagogical issues they want to monitor, search for relevant patterns in the data, create interventions to address these issues, and evaluate the effectiveness of the interventions (Paiva et al., 2025). Frameworks like ARCHED provide a human-centered design approach that implements a structured, multi-stage workflow between educators and AI, prioritizing pedagogical rigor and human agency over automation (Li et al., 2025). The aim is to provide education around critical workforce skills such as effective communication and collaboration by using AI as a collaborative partner (Atchley et al., 2024).

Implications for Education Faculties and Instructional Designer Training

The integration of artificial intelligence into instructional design has important implications for education faculties and instructional designer training. In this new era, there is a need for both existing and new competencies (Burneo-Arteaga et al., 2025). There is a lack of guidance, policies, and resources to help teachers and instructional designers integrate AI into teaching and learning processes efficiently and ethically (Ren & Wu, 2025).

For future instructional designers, AI literacy and AI readiness should become fundamental competencies as intelligent technological knowledge (Ren & Wu, 2025). Additionally, AI-supported innovative pedagogy needs to transform the student-teacher relationship and complement the social presence of educators (Ren & Wu, 2025). Education faculties should update instructional design programs to reflect new roles and responsibilities related to AI integration. This should include curriculum changes that provide instructional designers with practical experience in using AI-supported tools, developing human-AI collaboration strategies, and understanding the ethical and pedagogical limitations of AI (Luo et al., 2024). Continuous professional development programs are critically important for instructional designers to adapt to the rapidly evolving nature of AI and acquire new skills (Burneo-Arteaga et al., 2025).

General Evaluation

The integration of artificial intelligence and generative artificial intelligence technologies into educational environments is redefining the roles, responsibilities, and competencies of instructional designers (Amado-Salvatierra et al., 2023; Kumar et al., 2024; Luo et al., 2024; “September 2024 Full Issue,” 2024). In the future, instructional designers will be key figures working alongside AI, directing human-AI collaboration in design decisions, and shaping the curricula of education faculties in accordance with this transformation (Amado-Salvatierra et al., 2023; Burneo-Arteaga et al., 2025; Costello et al., 2024; Hilali & Chergui, 2025).

Instructional Designer Working with AI

In the age of artificial intelligence, the role of the instructional designer is undergoing significant change as AI transitions from being merely a supplementary tool to a creative partner (Amado-Salvatierra et al., 2023). Instructional designers are taking on the role of "glue" that brings together technology, content, and pedagogical best practices (“September 2024 Full Issue,” 2024). The emergence of generative AI tools is leading to rapid changes in the job responsibilities of instructional designers (Luo et al., 2024). Instructional designers are now using generative AI tools in various instructional design tasks, such as idea generation, performing low-risk tasks, facilitating design processes, and enhancing collaborations (Luo et al., 2024). This situation demonstrates that they play an important role in transforming AI into beneficial applications for teaching and learning that consider ethical parameters (“September 2024 Full Issue,” 2024). Future instructional designers should possess the competency to design more complex and personalized learning experiences using emerging tools such as learning analytics, artificial intelligence, mixed realities, and adaptive learning systems (Luo et al., 2025).

Human–AI Collaboration in Design Decisions

Collaboration between humans and artificial intelligence is becoming a fundamental element of modern instructional design practice. It is stated that AI enhances human capabilities in complex instructional tasks and sometimes even surpasses them (Luo et al., 2025). However, this collaboration requires human oversight and cooperation to ensure pedagogical appropriateness (Hilali & Chergui, 2025). The task of AI is to support humans in the decision-making process, not to replace human decision-makers (Ren et al., 2024). Particularly in fields like finance, while AI tools analyze large datasets in real time, the responsibility of human experts to interpret these insights and make strategic decisions continues (Paiva et al., 2025).

Human-AI collaboration in education can help teachers determine which pedagogical issues they want to monitor, search for relevant patterns in the data, create interventions to address these issues, and evaluate the effectiveness of the interventions (Paiva et al., 2025). Frameworks like ARCHED provide a human-centered design approach that implements a structured, multi-stage workflow between educators and AI, prioritizing pedagogical rigor and human agency over automation (Li et al., 2025). The aim is to provide education around critical workforce skills such as effective communication and collaboration by using AI as a collaborative partner (Atchley et al., 2024).

Implications for Education Faculties and Instructional Designer Training

The integration of artificial intelligence into instructional design has important implications for education faculties and instructional designer training. In this new era, there is a need for both existing and new competencies (Burneo-Arteaga et al., 2025). There is a lack of guidance, policies, and resources to help teachers and instructional designers integrate AI into teaching and learning processes efficiently and ethically (Ren & Wu, 2025).

For future instructional designers, AI literacy and AI readiness should become fundamental competencies as intelligent technological knowledge (Ren & Wu, 2025). Additionally, AI-supported innovative pedagogy needs to transform the student-teacher relationship and complement the social presence of educators (Ren & Wu, 2025). Education faculties should update instructional design programs to reflect new roles and responsibilities related to AI integration. This should include curriculum changes that provide instructional designers with practical experience in using AI-supported tools, developing human-AI collaboration strategies, and understanding the ethical and pedagogical limitations of AI (Luo et al., 2024). Continuous professional development programs are critically important for instructional designers to adapt to the rapidly evolving nature of AI and acquire new skills (Burneo-Arteaga et al., 2025).

Acknowledgements and Notes

Disclosure of AI usage

During the preparation of this work, the author(s) used Jenni.ai, ChatGPT in order to generate text, check grammar, and assist with literature search]. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published work.

References

Abuhassna, H., & Alnawajha, S. (2023). Instructional Design Made Easy! Instructional Design Models, Categories, Frameworks, Educational Context, and Recommendations for Future Work [Review of *Instructional Design Made Easy! Instructional Design Models, Categories, Frameworks, Educational Context, and Recommendations for Future Work*]. *European Journal of Investigation in Health Psychology and Education*, *13*(4), 715. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/ejihpe13040054

Ahmad, S. F., Han, H., Alam, M. M., Rehmat, Mohd. K., Irshad, M., Arraño-Muñoz, M., & Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications*, *10*(1). https://doi.org/10.1057/s41599-023-01787-8

AlAli, R., & Wardat, Y. (2024). Opportunities and Challenges of Integrating Generative Artificial Intelligence in Education. *International Journal of Religion*, *5*(7), 784. https://doi.org/10.61707/8y29gv34

Alfredo, R., Echeverría, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martínez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education Artificial Intelligence*, *6*, 100215. https://doi.org/10.1016/j.caeai.2024.100215

Amado-Salvatierra, H. R., Morales, M., & Hernández-Rizzardini, R. (2023). *Combining Human Creativity and AI-Based Tools in the Instructional Design of MOOCs: Benefits and Limitations*. 1. https://doi.org/10.1109/lwmoocs58322.2023.10306023

ANDREEA, E. A. (2022). INSTRUCTIONAL DESIGN IN EDUCATION. *IJAEDU- International E-Journal of Advances in Education*. https://doi.org/10.18768/ijaedu.1204180

Arısoy, B. (2022). Digitalization in education. *Cypriot Journal of Educational Sciences*, *17*(5), 1799. https://doi.org/10.18844/cjes.v17i5.6982

Arslan, B., Lehman, B., Tenison, C., Sparks, J. R., López, A. A., Gu, L., & Zapata-Rivera, D. (2024). Opportunities and challenges of using generative AI to personalize educational assessment [Review of *Opportunities and challenges of using generative AI to personalize educational assessment*]. *Frontiers in Artificial Intelligence*, *7*, 1460651. Frontiers Media. https://doi.org/10.3389/frai.2024.1460651

Atchley, P., Pannell, H., Wofford, K., Hopkins, M. M., & Atchley, R. A. (2024). Human and AI collaboration in the higher education environment: opportunities and concerns. *Cognitive Research Principles and Implications*, *9*(1), 20. https://doi.org/10.1186/s41235-024-00547-9

Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K., & Sailer, M. (2025). Looking Beyond the Hype: Understanding the Effects of AI on Learning. *Educational Psychology Review*, *37*(2). https://doi.org/10.1007/s10648-025-10020-8

Bayly-Castaneda, K., Ramirez-Montoya, M.-S., & Morita-Alexander, A. (2024). Crafting personalized learning paths with AI for lifelong learning: a systematic literature review. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1424386

Berg, G. van den. (2024). Generative AI and Educators: Partnering in Using Open Digital Content for Transforming Education. *Open Praxis*, *16*(2), 130. https://doi.org/10.55982/openpraxis.16.2.640

Bolick, A. D., & Silva, R. L. da. (2023). Exploring Artificial Intelligence Tools and Their Potential Impact to Instructional Design Workflows and Organizational Systems. *TechTrends*, *68*(1), 91. https://doi.org/10.1007/s11528-023-00894-2

Bozkurt, A., Xiao, J., Farrow, R., Bai, J. Y. H., Nerantzi, C., Moore, S., Dron, J., Stracke, C. M., Singh, L., Crompton, H., Koutropoulos, A., Terentev, E., Pazurek, A., Nichols, M., Sidorkin, A. M., Costello, E., Watson, S., Mulligan, D., Honeychurch, S., … Asino, T. I. (2024). The Manifesto for Teaching and Learning in a Time of Generative AI: A Critical Collective Stance to Better Navigate the Future. *Open Praxis*, *16*(4), 487. https://doi.org/10.55982/openpraxis.16.4.777

Brandão, A., Pedro, L., & Zagalo, N. (2024). Teacher professional development for a future with generative artificial intelligence – an integrative literature review. *Digital Education Review*, *45*, 151. https://doi.org/10.1344/der.2024.45.151-157

Brown, K., Larionova, V., Stepanova, N., & Lally, V. (2019). Re-imagining the Pedagogical Paradigm Within a Technology Mediated Learning Environment. *Open Education Studies*, *1*(1), 138. https://doi.org/10.1515/edu-2019-0009

Bulut, O., Beiting-Parrish, M., Casabianca, J. M., Slater, S. C., Jiao, H., Song, D., Ormerod, C. M., Fabiyi, D. G., Ivan, R., Walsh, C., Rios, O., Wilson, J. M., Yildirim-Erbasli, S. N., Wongvorachan, T., Liu, J. X., Tan, B., & Morilova, P. (2024). The Rise of Artificial Intelligence in Educational Measurement: Opportunities and Ethical Challenges. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2406.18900

Burneo-Arteaga, P., Lira, Y., Murzi, H., Balula, A., & Costa, A. P. (2025). Capability-based training framework for generative AI in higher education. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1594199

Chan, C. K. Y., & Colloton, T. (2024). *Generative AI in Higher Education*. https://doi.org/10.4324/9781003459026

Chimalakonda, S., & Nori, K. V. (2021). A patterns based approach for the design of educational technologies. *Interactive Learning Environments*, *31*(4), 2114. https://doi.org/10.1080/10494820.2021.1875000

Chiu, T. K. F. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: a case of ChatGPT. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-024-10366-w

Choi, G. W., Kim, S. H., Lee, D., & Moon, J. (2024). Utilizing Generative AI for Instructional Design: Exploring Strengths, Weaknesses, Opportunities, and Threats. *TechTrends*, *68*(4), 832. https://doi.org/10.1007/s11528-024-00967-w

Costello, E., Brunton, J., Otrel-Cass, K., Lyngdorf, N. E. R., & Brown, M. (2024). Hacking Happier Futures : An AI-augmented student hackathon to address affective and ethical digital learning challenges. *Research Portal Denmark*, 9. https://local.forskningsportal.dk/local/dki-cgi/ws/cris-link?src=aau&id=aau-40fe5df2-680d-4e0a-8285-69066ff15b33&ti=Hacking%20Happier%20Futures%20%3A%20An%20AI-augmented%20student%20hackathon%20to%20address%20affective%20and%20ethical%20digital%20learning%20challenges

Dickey, E., & Bejarano, A. (2023). GAIDE: A Framework for Using Generative AI to Assist in Course Content Development. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2308.12276

Doyle, A., Sridhar, P., Agarwal, A., Šavelka, J., & Sakr, M. (2025). A comparative study of AI-generated and human-crafted learning objectives in computing education. *Journal of Computer Assisted Learning*, *41*(1). https://doi.org/10.1111/jcal.13092

Francesa-Alfaro, A., Leiva-Chinchilla, P., Chacón, M. P., & Garita, C. (2016). *User-friendly instructional design tool to facilitate course planning*. 1. https://doi.org/10.1109/laclo.2016.7751763

Giannakos, M. N., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernández-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2024). The promise and challenges of generative AI in education. *Behaviour and Information Technology*, *44*(11), 2518. https://doi.org/10.1080/0144929x.2024.2394886

Godin, V. V., & Terekhova, A. (2021). Digitalization of Education: Models and Methods. *International Journal of Technology*, *12*(7), 1518. https://doi.org/10.14716/ijtech.v12i7.5343

Godsk, M., & Møller, K. L. (2024). Engaging students in higher education with educational technology. *Education and Information Technologies*, *30*(3), 2941. https://doi.org/10.1007/s10639-024-12901-x

Guettala, M., Bourekkache, S., Kazar, O., & Harous, S. (2024). Generative Artificial Intelligence in Education: Advancing Adaptive and Personalized Learning. *Acta Informatica Pragensia*, *13*(3), 460. https://doi.org/10.18267/j.aip.235

Hanshaw, G., & Hanson, J. (2019). Using Microlearning and Social Learning to Improve Teachers' Instructional Design Skills: A Mixed Methods Study of Technology Integration in Teacher Professional Development. *International Journal of Learning and Development*, *9*(1), 145. https://doi.org/10.5296/ijld.v9i1.13713

Hao, Z., Luo, H., Chen, Y., & Zhang, Y. (2025). Unpacking Graduate Students' Learning Experience with Generative AI Teaching Assistant in A Quantitative Methodology Course. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2506.02966

Hariyanto, H., Kristianingsih, F. X. D., & Maharani, R. (2025). Artificial intelligence in adaptive education: a systematic review of techniques for personalized learning. *Discover Education*, *4*(1). https://doi.org/10.1007/s44217-025-00908-6

Hetmańczyk, P. (2023). Digitalization and its impact on labour market and education. Selected aspects. *Education and Information Technologies*, *29*(9), 11119. https://doi.org/10.1007/s10639-023-12203-8

Higher Education for Good. (2023). In *Open Book Publishers*. https://doi.org/10.11647/obp.0363

Hilali, K., & Chergui, M. (2025). Toward a New Instructional Design Methodology in the Era of Generative AI. In *Lecture notes on data engineering and communications technologies* (p. 3). Springer International Publishing. https://doi.org/10.1007/978-3-031-98476-1_1

Holmes, W., Iniesto, F., Anastopoulou, S., & Boticario, J. G. (2023). Stakeholder Perspectives on the Ethics of AI in Distance-Based Higher Education. *The International Review of Research in Open and Distributed Learning*, *24*(2), 96. https://doi.org/10.19173/irrodl.v24i2.6089

Hong, C. (2024). The Ethical Challenges of Educational Artificial Intelligence and Coping Measures: A Discussion in the Context of the 2024 World Digital Education Conference. *Science Insights Education Frontiers*, *20*(2), 3263. https://doi.org/10.15354/sief.24.re339

Hooshyar, D., Šír, G., Yang, Y., Kikas, E., Hämäläinen, R., Kärkkäinen, T., Gašević, D., & Azevedo, R. (2025). Towards responsible AI for education: Hybrid human-AI to confront the Elephant in the room. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2504.16148

Hu, W. W., & Chan, C. K. Y. (2025). From user needs to AI solutions: a human-centered design approach for AI-powered virtual teamwork competency training. *International Journal of Educational Technology in Higher Education*, *22*(1). https://doi.org/10.1186/s41239-025-00551-z

Huang, L. (2023). Ethics of Artificial Intelligence in Education: Student Privacy and Data Protection. *Science Insights Education Frontiers*, *16*(2), 2577. https://doi.org/10.15354/sief.23.re202

Huang, S., Zeng, X., & Luo, W. (2025). Knowledge-enhanced large language models for automatic lesson plan generation. *Humanities and Social Sciences Communications*, *12*(1). https://doi.org/10.1057/s41599-025-06004-2

Human, Technologies and Quality of Education, 2023. (2023). https://doi.org/10.22364/htqe.2023

Ifenthaler, D., Majumdar, R., Gorissen, P., Judge, M., Mishra, S., Raffaghelli, J. E., & Shimada, A. (2024). Artificial Intelligence in Education: Implications for Policymakers, Researchers, and Practitioners. *Technology Knowledge and Learning*. https://doi.org/10.1007/s10758-024-09747-0

Ifraheem, S., Rasheed, M., & Siddiqui, A. (2024). Transforming Education Through Artificial Intelligence: Personalization, Engagement and Predictive Analytics. *Deleted Journal*, *13*(2), 250. https://doi.org/10.62345/jads.2024.13.2.22

Ismail, A. M. A., & Alkhazali, T. M. N. (2019). Assessing the Impact of an Instructional Design Course on Arabian Gulf University Distance Learning Students' Instructional Design Competencies. *Open Journal of Social Sciences*, *7*(9), 106. https://doi.org/10.4236/jss.2019.79009

Jukiewicz, M. (2025). How generative artificial intelligence transforms teaching and influences student wellbeing in future education. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1594572

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, *15*(1), 17458. https://doi.org/10.1038/s41598-025-97652-6

Kılınç, H. K., & Keçecioğlu, Ö. F. (2024). Generative Artificial Intelligence: A Historical and Future Perspective. *Academic Platform Journal of Engineering and Smart Systems*, *12*(2), 47. https://doi.org/10.21541/apjess.1398155

Koh, J., Cowling, M., Jha, M., & Sim, K. N. (2023). The Human Teacher, the AI Teacher and the AIed-Teacher Relationship. *Journal of Higher Education Theory and Practice*, *23*(17). https://doi.org/10.33423/jhetp.v23i17.6543

Kumar, S., Gunn, A., Rose, R. J., Pollard, R., Johnson, M., & Ritzhaupt, A. D. (2024). The Role of Instructional Designers in the Integration of Generative Artificial Intelligence in Online and Blended Learning in Higher Education. *Online Learning*, *28*(3). https://doi.org/10.24059/olj.v28i3.4501

Lan, M., & Zhou, X. (2025). A qualitative systematic review on AI empowered self-regulated learning in higher education. *Npj Science of Learning*, *10*(1), 21. https://doi.org/10.1038/s41539-025-00319-0

Li, H., Fang, Y., Zhang, S., Lee, S. M., Wang, Y., Trexler, M., & Botelho, A. F. (2025). ARCHED: A Human-Centered Framework for Transparent, Responsible, and Collaborative AI-Assisted Instructional Design. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2503.08931

Li, H., Xu, T., Zhang, C., Chen, E., Liang, J., Fan, X., Li, H., Tang, J., & Wen, Q. (2024). Bringing Generative AI to Adaptive Learning in Education. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2402.14601

Luo, J., Chang, Z., Yin, J., & Teo, H. (2025). Design and assessment of AI-based learning tools in higher education: a systematic review [Review of *Design and assessment of AI-based learning tools in higher education: a systematic review*]. *International Journal of Educational Technology in Higher Education*, *22*(1). Springer Nature. https://doi.org/10.1186/s41239-025-00540-2

Luo, T., Muljana, P. S., Ren, X., & Young, D. (2024). Exploring instructional designers' utilization and perspectives on generative AI tools: A mixed methods study. *Educational Technology Research and Development*, *73*(2), 741. https://doi.org/10.1007/s11423-024-10437-y

Maity, S., & Deroy, A. (2024a). Human-Centric eXplainable AI in Education. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2410.19822

Maity, S., & Deroy, A. (2024b). *Generative AI and Its Impact on Personalized Intelligent Tutoring Systems*. https://doi.org/10.36227/techrxiv.172954174.48041095/v1

McNeill, L. (2024). Automation or Innovation? A Generative AI and Instructional Design Snapshot. *IAFOR International Conference on Education, Official Conference Proceedings*, 187. https://doi.org/10.22492/issn.2189-1036.2024.17

Merino-Campos, C. (2025). The Impact of Artificial Intelligence on Personalized Learning in Higher Education: A Systematic Review [Review of *The Impact of Artificial Intelligence on Personalized Learning in Higher Education: A Systematic Review*]. *Trends in Higher Education*, *4*(2), 17. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/higheredu4020017

Mondal, G. (2025). Integrating Learning Theories in Web-Based Instruction: Towards a Unified Conceptual Framework for Pedagogical Design. *Universal Journal of Educational Research*, *4*(3), 305. https://doi.org/10.64637/629057

mostafa, haitham, abdelazia, abdelazia, & Saleh, N. (2024). Employing Generative Artificial Intelligence in Instructional Design Based on The ADDIE Model., *14*(4), 61. https://doi.org/10.21608/idj.2024.282953.1136

Mulenga, R., & Shilongo, H. (2024). Hybrid and Blended Learning Models: Innovations, Challenges, and Future Directions in Education. *Acta Pedagogia Asiana*, *4*(1), 1. https://doi.org/10.53623/apga.v4i1.495

Namoun, A., Ibrahim, I. A., Mustafa, E., Alrehaili, A., Tufail, A., Shuja, J., & Bilal, K. (2024). Generative artificial intelligence in education: an umbrella review of applications and challenges. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-4892155/v1

Namoun, A., Ibrahim, I. A., Mustafa, E., Alrehaili, A., Tufail, A., Shuja, J., Bilal, K., & Alanazi, M. H. (2024). Generative artificial intelligence in education: an umbrella review of applications and challenges. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-4892155/v2

Netland, T. H., Dzengelevski, O. von, Tesch, K., & Kwasnitschka, D. (2024). Comparing human-made and AI-generated teaching videos: An experimental study on learning effects. *Computers & Education*, *224*, 105164. https://doi.org/10.1016/j.compedu.2024.105164

Nguyen, T., Nguyễn, M. T., & Tran, H. T. (2023). Artificial intelligent based teaching and learning approaches: A comprehensive review [Review of *Artificial intelligent based teaching and learning approaches: A*

comprehensive review]. *International Journal of Evaluation and Research in Education (IJERE)*, *12*(4), 2387. Institute of Advanced Engineering and Science (IAES). https://doi.org/10.11591/ijere.v12i4.26623

OECD. (2023a). Shaping Digital Education. In *OECD Publishing eBooks*. https://doi.org/10.1787/bac4dc9f-en

OECD. (2023b). OECD Digital Education Outlook 2023. In *Digital education outlook*. Organization for Economic Cooperation and Development. https://doi.org/10.1787/c74f03de-en

Özkan, Â., Çevik, İ., Saylan, E., & Çakıroğlu, Ü. (2025). The Past and Present of Instructional Design in Online Learning: Trends and Emerging Directions. *The International Review of Research in Open and Distributed Learning*, *26*(3), 126. https://doi.org/10.19173/irrodl.v26i3.8507

Paiva, R., Xisto, J., Sobrinho, Á., Silva, A., Sarmento, F., Recch, F., Tenório, S., Carvalho, A. M. X., Bittencourt, I. I., & Isotani, S. (2025). Expanding the resilience of the Brazilian education system by supporting the evaluation of digital textbooks. *Humanities and Social Sciences Communications*, *12*(1). https://doi.org/10.1057/s41599-025-05489-1

Plooy, E. du, Casteleijn, D., & Franzsen, D. (2024). Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement [Review of *Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement*]. *Heliyon*, *10*(21). Elsevier BV. https://doi.org/10.1016/j.heliyon.2024.e39630

Qian, Y. (2025). Pedagogical Applications of Generative AI in Higher Education: A Systematic Review of the Field [Review of *Pedagogical Applications of Generative AI in Higher Education: A Systematic Review of the Field*]. *TechTrends*, *69*(5), 1105. Springer Science+Business Media. https://doi.org/10.1007/s11528-025-01100-1

Rajasekaran, S., Adam, T., & Tilmes, K. (2024). Digital Pathways for Education: Enabling Greater Impact for All. In *Washington, DC: World Bank eBooks*. https://doi.org/10.1596/42386

Ren, C., Pardos, Z. A., & Li, Z. (2024). Human-AI Collaboration Increases Skill Tagging Speed but Degrades Accuracy. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2403.02259

Ren, X., & Wu, M. L. (2025). Examining Teaching Competencies and Challenges While Integrating Artificial Intelligence in Higher Education. *TechTrends*, *69*(3), 519. https://doi.org/10.1007/s11528-025-01055-3

Roméro, M., Frøsig, T. B., Taylor-Beswick, A. M. L., Laru, J., Bernasco, B., Urmeneta, A., Strutynska, O., & Girard, M.-A. (2024). *Manifesto in Defence of Human-Centred Education in the Age of Artificial Intelligence* (p. 157). https://doi.org/10.1007/978-3-031-55272-4_12

Rouabhia, D. (2024). Artificial Intelligence Driven Course Generation: A Case Study Using ChatGPT., *5*(3), 287. https://doi.org/10.70091/atras/ai.18

Sajja, R., Sermet, Y., Cwiertny, D. M., & Demir, İ. (2025). Integrating AI and Learning Analytics for Data-Driven Pedagogical Decisions and Personalized Interventions in Education. *Technology Knowledge and Learning*. https://doi.org/10.1007/s10758-025-09897-9

Senadheera, V. V., Ediriweera, D. S., & Rupasinghe, T. P. (2024). Instructional Design Models for Digital Learning in Higher Education — A Scoping Review [Review of *Instructional Design Models for Digital Learning in Higher Education — A Scoping Review*]. *Journal of Learning for Development*, *11*(1), 15. https://doi.org/10.56059/jl4d.v11i1.973

September 2024 Full Issue. (2024). *Online Learning*, *28*(3). https://doi.org/10.24059/olj.v28i3.4714

Sgay, M. A., Borji, A., Bahramzadeh, N., Abedi, M., & Golizadeh, S. (2025). Identifying Indicators of Optimizing Learning Environments Using Artificial Intelligence in Curriculum Design. *Deleted Journal*, *3*, 1. https://doi.org/10.61838/kman.ec.psynexus.3.3

Song, D. (2024). Artificial intelligence for human learning: A review of machine learning techniques used in education research and a suggestion of a learning design model [Review of *Artificial intelligence for human learning: A review of machine learning techniques used in education research and a suggestion of a learning design model*]. *American Journal of Education and Learning*, *9*(1), 1. https://doi.org/10.55284/ajel.v9i1.1024

Spector, J. M. (2015). *Foundations of Educational Technology*. https://doi.org/10.4324/9781315764269

Sperling, K., Stenliden, L., Nissen, J., & Heintz, F. (2023). Behind the Scenes of Co-designing AI and LA in K-12 Education. *Postdigital Science and Education*, *6*(1), 321. https://doi.org/10.1007/s42438-023-00417-5

Sridhar, P., Doyle, A., Agarwal, A., Bogart, C., Šavelka, J., & Sakr, M. (2023). Harnessing LLMs in Curricular Design: Using GPT-4 to Support Authoring of Learning Objectives. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2306.17459

Team, L., Google, AUTHOR_ID, N., Martín, A., Globerson, A., Wang, A., Shekhawat, A., Iurchenko, A., Choudhury, A., Hassidim, A., Çakmakli, A., Evron, A. S., Yang, C., Heldreth, C., Akrong, D., Elidan, G., Hairong, M., Li, I., Cohen, I., … Matias, Y. (2025). Towards an AI-Augmented Textbook. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2509.13348

Trevisan, O., Christensen, R., Drossel, K., Friesen, S., Forkosh-Baruch, A., & Phillips, M. (2024). Drivers of Digital Realities for Ongoing Teacher Professional Learning. *Technology Knowledge and Learning*, *29*(4), 1851. https://doi.org/10.1007/s10758-024-09771-0

Trif-Boia, A. E. (2022). *INSTRUCTIONAL DESIGN IN EDUCATION*. https://doi.org/10.47696/adved.202201

Verghis, A. M., Jose, B., Sheeja, P. R., Varghese, S., Mumthas, S., & Shaiju, K. S. (2025). Beyond personalization: autonomy and agency in intelligent systems education. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1610239

Versteijlen, M., & Groesbeek, M. J. (2024). Perspective Chapter: Educational Technology under Scrutiny in Higher Education – A Framework for Balancing Environmental, Economic and Social Aspects in a Blended Design. In *IntechOpen eBooks*. IntechOpen. https://doi.org/10.5772/intechopen.1005117

Wang, X., Huang, R., Sommer, M., Pei, B., Shidfar, P., Rehman, M. S., Ritzhaupt, A. D., & Martin, F. (2024). The Efficacy of Artificial Intelligence-Enabled Adaptive Learning Systems From 2010 to 2022 on Learner Outcomes: A Meta-Analysis. *Journal of Educational Computing Research*, *62*(6), 1348. https://doi.org/10.1177/07356331241240459

West, R. E. (2018). *Foundations of Learning and Instructional Design Technology*. https://doi.org/10.59668/3

Wintersberg, L., & Pittich, D. (2025). Toward a universal definition of instructional design: a systematic review [Review of *Toward a universal definition of instructional design: a systematic review*]. *Discover Education*, *4*(1). Springer Science+Business Media. https://doi.org/10.1007/s44217-025-00491-w

Xia, Q., Weng, X., Ouyang, F., Lin, T., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education [Review of *A scoping review on how generative*

artificial intelligence transforms assessment in higher education]. *International Journal of Educational Technology in Higher Education*, *21*(1). Springer Nature. https://doi.org/10.1186/s41239-024-00468-z

Xiao, J. (2024). Will Artificial Intelligence Enable Open Universities to Regain their Past Glory in the 21st Century? *Open Praxis*, *16*(1), 11. https://doi.org/10.55982/openpraxis.16.1.618

Y, W., Jiang, Y.-H., Liu, J., Qi, C., & Rui, J. (2024). The Advancement of Personalized Learning Potentially Accelerated by Generative AI. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2412.00691

Yang, J., & Zhang, H. (2024). Development And Challenges of Generative Artificial Intelligence in Education and Art. *Highlights in Science Engineering and Technology*, *85*, 1334. https://doi.org/10.54097/vaeav407

Yao, H., Xu, W., Turnau, J., Kellam, N., & Wei, H. (2025). *Instructional Agents: LLM Agents on Automated Course Material Generation for Teaching Faculties*. https://doi.org/10.48550/ARXIV.2508.19611

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review [Review of *The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review*]. *Smart Learning Environments*, *11*(1). Springer Nature. https://doi.org/10.1186/s40561-024-00316-7

Zhaı, X. (2024). Transforming Teachers' Roles and Agencies in the Era of Generative AI: Perceptions, Acceptance, Knowledge, and Practices. *Journal of Science Education and Technology*. https://doi.org/10.1007/s10956-024-10174-0

Zhang, D. (2025). Transforming Higher Education with AI-Powered Video Lectures. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2511.20660

Zhang, L., Jackson, H., Yang, S., Qian, X., Carter, R., Diliberto, J. J., & Zhang, J. (2024). Prototyping A Multi-Agent System to Enhance AI-Human Collaboration in Individualized Education Program Development. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-5339247/v1

Zhang, X., Zhang, P., Shen, Y., Liu, M., Wang, Q., Gašević, D., & Fan, Y. (2024). A Systematic Literature Review of Empirical Research on Applying Generative Artificial Intelligence in Education. *Frontiers of Digital Education.*, *1*(3), 223. https://doi.org/10.1007/s44366-024-0028-5

Zhu, H., Sun, Y., & Yang, J. (2025). Towards responsible artificial intelligence in education: a systematic review on identifying and mitigating ethical risks. *Humanities and Social Sciences Communications*, *12*(1). https://doi.org/10.1057/s41599-025-05252-6

Zou, Y. L., Kuek, F., Feng, W., & Cheng, X. (2025). Digital learning in the 21st century: trends, challenges, and innovations in technology integration. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1562391

Ибрагимов, Г. И., & Kalimullina, A. A. (2020). Theoretical and Practical Issues of Education as Part of a Digitalization Process. *ARPHA Proceedings*, *1*, 807. https://doi.org/10.3897/ap.2.e0807

Author Information

Ömer Bilen

https://orcid.org/0000-0001-7288-7606

Atatürk University

Erzurum

Türkiye

Contact e-mail: omerbilen76@gmail.com

Citation

Bilen, Ö. (2025). Artificial Intelligence in Instructional Design and Technologies. In A. Kaban & T. Demirel (Eds.), *Interdisciplinary Applications of Artificial Intelligence - 1* (pp. 137-160). ISTES.

Chapter 8- Use of Generative Artificial Intelligence in Developing Writing Skills in Language Education

Hamza Polat

Chapter Highlights

- GenAI integration is framed through sociocultural theory, Vygotsky, Bloom, cognitive apprenticeship, posthumanist pedagogy, and AI–human co-authorship models.
- Core writing pedagogies—process, genre, and integrated process–genre approaches—are outlined to show how writing is recursive, strategic, and socially situated.
- Technical foundations of generative AI and LLMs are explained, including transformer architectures, autoregressive token prediction, and decoding strategies.
- GenAI is shown to support all stages of writing—pre-writing ideation, drafting and co-writing, revision and editing, and formatting/polishing—improving cohesion, accuracy, and confidence.
- Instructional design guidelines emphasize explicitly defined learning goals, scaffolded activities, reflection, agency, ethical use, and co-regulated learning.
- AI-based assessment is reviewed, showing high reliability in constrained tasks but mixed validity, with strongest value as a formative aid alongside teacher judgment.

Introduction

Generative artificial intelligence (GenAI) is rapidly reshaping what it means to teach, learn, and assess writing. Tools powered by large language models have moved far beyond simple spellcheckers to become ever-present composing partners that can generate ideas, extend text, reorganize arguments, and evaluate drafts at scale. Their growing presence in students' everyday writing practices forces educators to revisit foundational questions: What counts as writing and authorship when texts are co-produced with machines? Under which conditions can AI be said to *support* learning rather than merely automate text production?

This chapter situates GenAI within robust theoretical and conceptual frameworks before examining its practical roles across the writing process. It first revisits core writing pedagogies – process, genre, and integrated process–genre approaches – and introduces key developments in generative language technologies and learning theories that justify or problematize AI use in writing instruction. It then synthesizes emerging research on how learners and teachers deploy AI for pre-writing, drafting and co-writing, revision and editing, formatting and publishing, instructional design, and assessment. Finally, it critically interrogates the opportunities, challenges, and ethical concerns surrounding AI-mediated writing, arguing that the educational value of GenAI hinges on thoughtful task design, transparent policies, and an explicit commitment to preserving learner agency, critical thinking, and intellectual ownership.

Theoretical and Conceptual Foundations

Generative artificial intelligence (GenAI) has rapidly become an integral component of students' daily composing practices, prompting writing educators to reassess not only instructional approaches but also foundational definitions of writing and authorship (Barrett & Pack, 2023; Wang, 2024). Current scholarship argues that GenAI should not be positioned as a supplemental technological tool; rather, its incorporation into writing pedagogy must be situated within rigorous theoretical and technological frameworks that protect learner agency, cultivate critical and metacognitive thinking, and transform – without displacing – core writing processes such as idea generation, drafting, feedback, and revision (Han, 2025; Tate et al., 2025; Colby, 2025; Hong et al., 2025).

From a sociocultural standpoint, GenAI applications may operate as "more knowledgeable others," offering scaffolded, dialogic guidance within students' zones of proximal development. This pedagogical value becomes particularly salient when instructional models intentionally draw on Vygotskian theory, constructivism, Bloom's taxonomy, and cognitive apprenticeship to structure iterative human–AI interaction around increasingly complex reasoning, strategic decision-making, and self-regulation (Han, 2025; Hong et al., 2025; Xia, 2025). Concurrently, frameworks rooted in human–AI collaboration, posthumanist pedagogy, and socio-culturally grounded human-centered designs emphasize shared authorship, transparency, and reflective AI literacy to ensure that efficiency and productivity gains do not eclipse creativity, ethical accountability, or students' intellectual ownership of their work (Han, 2025; Maphoto et al., 2024; Seran et al., 2025; Yusuf et al., 2024).

The technological rationale for GenAI integration is founded on the development of AI-mediated learning environments that actively operationalize these theories into concrete features and user practices. Empirical research on AI-supported writing systems—such as PapyrusAI, ChatGPT-based platforms, and Grammarly—illustrates that when these tools are embedded within structured pedagogical sequences (e.g., "think first, then prompt, then reflect"), employ scaffolded prompting interfaces, and support staged writing processes (pre-writing, drafting, revision) alongside blended AI–teacher feedback, GenAI can effectively enhance linguistic accuracy, organizational coherence, metacognitive reflection, and students' feedback literacy, without diminishing cognitive engagement or transferring responsibility for learning to the tool itself (Tate et al., 2025; Escalante et al., 2023; Kong & Yang, 2024; Chen, 2025; Chanpradit, 2025).

Systematic reviews and empirical studies across EFL, ENL, distance education, and doctoral writing contexts converge on a depiction of GenAI as a configurable pedagogical scaffold and feedback partner—capable of improving cohesion, clarity, and learner confidence, yet simultaneously capable of creating over-dependence, cognitive dissonance, and ethical dilemma when divorced from explicit institutional policies and usage guidelines (Escalante et al., 2023; Maphoto et al., 2024; Zamorano, 2025; Chanpradit, 2025; Al-Saliti et al., 2025). Against this backdrop, scholarly focus increasingly shifts from the question of whether GenAI should be integrated into writing pedagogy to how—and under which theoretical, pedagogical, and technological commitments – its integration can credibly claim to support meaningful learning rather than merely automate text production.

Writing Pedagogies

Writing pedagogy has undergone substantial conceptual evolution over recent decades, shifting away from conceptions of writing as a static, one-shot product toward recognition of it as a recursive, strategic, and socially situated act. Contemporary practice is most prominently influenced by two complementary pedagogical orientations—process-based instruction and genre-based approaches—which, together and in tension, shape how learners are taught not only to produce texts but also to understand the social purposes and discourse communities that anchor written communication. Against this backdrop, the following section delineates the foundational assumptions, instructional features, and empirical outcomes of process pedagogy, genre pedagogy, and their subsequent integration.

Core pedagogical orientations within contemporary writing instruction are commonly grounded in process writing and genre-based teaching, both of which reposition writing away from a static, one-shot product and toward a complex, iterative, and socially embedded activity. The process approach underscores the cognitive and recursive dimensions of composing, whereby learners engage in planning, drafting, revising, and editing across multiple drafts, typically accompanied by systematic feedback, peer and teacher conferencing, and collaborative task structures (Jarunthawatchai et al., 2025; Hanusova et al., 2020; Bin-Hady et al., 2020; Badger & White, 2000). This perspective foregrounds learning *how* to write—emphasizing the gradual development of strategies for idea generation, text organization, linguistic refinement, and writerly self-regulation over time.

Conversely, genre-based pedagogy conceptualizes writing as a socially situated form of action. In this model, texts are understood as purposeful, recurrent responses to specific contexts, audiences, and communicative aims, characterized by recognizable rhetorical stages and linguistic configurations (Hyland, 2003; Hyland, 2007; Phạm & Bui, 2021; Orfanò & Wingate, 2024). Genre pedagogy often materializes through a structured teaching–learning cycle in which learners build contextual knowledge, analyze model texts, engage in jointly constructed writing with teacher scaffolding, and subsequently produce independent texts within the target genre (Jarunthawatchai et al., 2025; Bin-Hady et al., 2020; Phạm & Bui, 2021; Ganapathy et al., 2022). Empirical evidence demonstrates that explicit genre instruction enhances students' command of discourse moves, organizational patterns, and genre-specific linguistic features, particularly within high-stakes academic and professional writing tasks (Hyland, 2007; Zhang & Zhang, 2021; Ganapathy et al., 2022; Orfanò & Wingate, 2024).

Yet, neither model—on its own—offers a sufficiently comprehensive pedagogical foundation. Process-oriented instruction may inadvertently minimize explicit attention to language forms and genre conventions, while a genre-exclusive approach risks obscuring the nonlinear, recursive realities of composing. These limitations have contributed to the development of integrated process–genre pedagogies, which synthesize cognitive and sociocultural insights by drawing simultaneously on process, product, and genre traditions (Jarunthawatchai et al., 2025; Hanusova et al., 2020; Bin-Hady et al., 2020; Badger & White, 2000). Within such frameworks, learners are supported in mastering both the staged procedures of composing and the ways texts are shaped by discourse communities, communicative purposes, and audience expectations—often through iterative cycles of reading model texts, collaborative drafting, and autonomous revision (Jarunthawatchai et al., 2025; Hanusova et al., 2020; Bin-Hady et al., 2020). Research conducted in EFL and university contexts indicates that these blended approaches yield improvements in content quality, textual organization, cohesion, and learner confidence at levels surpassing single-method instruction (Jarunthawatchai et al., 2025; Hanusova et al., 2020; Insuwan & Thongrin, 2025; Bin-Hady et al., 2020; Zhang & Zhang, 2021).

Taken together, this body of work demonstrates that no singular pedagogical model is sufficient in isolation. Instead, integrated process–genre frameworks offer the strongest basis for supporting students' development as writers—cultivating both strategic, recursive composing and explicit genre awareness. As writing needs continue to diversify across academic and professional landscapes, future instructional design must strategically blend cognitive and sociocultural perspectives, ensuring that pedagogy remains adaptable, contextually responsive, and capable of sustaining learners' growth as autonomous and confident writers.

Generative AI and Language Technologies

Generative artificial intelligence refers to a class of computational models designed to produce new content – such as text, images, or code – by learning underlying statistical patterns from large-scale datasets and subsequently sampling from these learned distributions to create original yet data-consistent outputs (Banh & Strobel, 2023; Ooi et al., 2023). Within this broader family, large language models (LLMs) constitute the subset specialized for natural language processing. These models are transformer-based neural architectures, frequently comprising

billions of parameters, trained on extensive text corpora to predict successive tokens in a sequence, thereby enabling advanced language understanding and generation across a wide array of NLP tasks (Hagos et al., 2024; Li et al., 2022; Kumar, 2024; Xu et al., 2023). In contemporary NLP practice, text generation is conceptualized as an end-to-end mapping from an input (e.g., a natural language prompt, source text, or structured data) to an output sequence, optimized such that the model's conditional probability distributions yield fluent, coherent, and contextually appropriate language (Li et al., 2022; Kumar, 2024).

The foundation of LLM-based text production lies in autoregressive language modeling. During pre-training, the model learns to estimate the probability of each upcoming token given its preceding context, employing self-attention mechanisms to capture contextual dependencies and long-range linguistic relationships (Li et al., 2022; Kumar, 2024; Venkatasubramanian & Chakraborty, 2024). Once trained, generation unfolds through an iterative sampling process in which the model selects tokens from its learned probability distribution; decoding strategies such as greedy search or beam search are used to balance determinism, coherence, and linguistic diversity (Kumar, 2024). These technical capabilities allow LLMs to perform a wide spectrum of text-based operations—including translation, summarization, question answering, and conversational interaction—often in zero-shot or few-shot configurations where task instructions are conveyed solely through natural-language prompts (Hagos et al., 2024; Li et al., 2022; Xu et al., 2023).

More broadly, generative AI systems extend NLP beyond traditional discriminative paradigms (e.g., classification) toward creative, open-ended language generation. While this shift enables transformative applications in education, communication, and knowledge work, it simultaneously introduces concerns regarding reliability, bias, ethical use, and vulnerability to misuse (Hagos et al., 2024; Banh & Strobel, 2023; Sengar et al., 2024; Ooi et al., 2023).

Learning Theories Supporting AI Integration

Core learning theories provide an essential foundation for understanding why and how artificial intelligence can be pedagogically legitimized within writing instruction. Rather than conceptualizing AI as a neutral, context-free tool, contemporary scholarship situates its use within established perspectives on how learners think, interact, and develop over time. These theoretical lenses help determine not only whether AI should be used, but under which instructional conditions it can support – not substitute – human cognition and authorship.

From a sociocultural and dialogic standpoint, learning is a socially mediated, language-dependent process. Within this framing, AI chatbots and generative writing tools can be understood as mediating artifacts or "digital peers" capable of participating in dialogue, providing feedback within learners' zones of proximal development, and scaffolding drafts, outlines, and argumentation (Yao et al., 2025; Kim et al., 2024; Li & Wilson, 2025; Parker et al., 2024; Wegerif & Casebourne, 2025). Work drawing on activity theory conceptualizes AI-supported feedback as one component within a broader activity system – comprising subject, tool, community, and rules – that complements rather than replaces teacher response in L2 writing (Yao et al., 2025). Dialogic theories further

justify the integration of generative AI when it is positioned as expanding cultural conversation and co-construction of meaning, rather than simply delivering finished textual products (Wegerif & Casebourne, 2025).

Cognitive and self-regulated learning theories offer an additional rationale by framing AI as a mechanism for reducing cognitive load, externalizing linguistic subskills, and strengthening metacognition throughout the writing process. AI-enhanced NLP tools, for instance, support grammar, vocabulary, and textual organization, thereby freeing cognitive resources for higher-order planning, idea development, and argumentation (Zhao, 2024; Kong & Yang, 2024; Song & Song, 2023; Lai, 2025). Evidence from experimental and quasi-experimental studies indicates that AI-generated feedback can enhance writing quality while promoting goal-setting, monitoring, and iterative revision in alignment with self-regulated learning models (Song & Song, 2023; Zhang et al., 2025; Fan et al., 2024; Lai, 2025). Nonetheless, findings related to "metacognitive laziness" underscore the importance of deliberate instructional design that prompts students to evaluate, question, and refine AI-generated suggestions rather than passively adopting them (Fan et al., 2024).

Constructivist and social cognitive perspectives further support AI-mediated writing environments that emphasize active meaning-making, modeling, and dialogic feedback. Tools that enable personalized, adaptive interaction and immediate response align with constructivist notions of learners building knowledge through rich engagement with tasks and resources (Gibson et al., 2023; Bahari et al., 2025). Social cognitive theory has also been applied to frame AI-supported language learning as a context in which learners can observe, imitate, and then appropriate modeled language and rhetorical strategies, with AI functioning as a conversational partner capable of enhancing self-efficacy – particularly when accompanied by human scaffolding (Guan et al., 2024; Chiu et al., 2023). Across these theoretical traditions, the pedagogical legitimacy of AI in writing ultimately rests on conceptualizing AI as a scaffold that strengthens agency, dialogue, and self-regulation – rather than a shortcut that displaces human thinking or diminishes intellectual ownership.

Applications of Generative AI Across the Writing Process

Generative AI – particularly applications powered by large language models such as ChatGPT, Gemini, and Copilot – is rapidly reshaping how students engage with the writing process, extending its influence from the earliest emergence of an idea through the refinement of a final manuscript. Rather than functioning merely as automated correction tools, these systems increasingly operate as multipurpose writing assistants, virtual tutors, and digital peers that learners strategically mobilize at different stages of composing in response to varied needs (Kim et al., 2024; Kim et al., 2025; Alpar, 2025). Findings from emerging classroom-based and interview-driven research indicate that students already incorporate GenAI into cycles of planning, drafting, revising, and reflecting, often engaging in iterative, back-and-forth patterns of human–AI co-composition (Kim et al., 2024; Wang, 2024; Kim et al., 2025).

At the pre-writing stage, GenAI tools support learners in brainstorming topics, generating research questions, locating background information, and sketching preliminary outlines, ultimately lowering the cognitive and affective barriers to initiating complex academic tasks (Kim et al., 2024; Kim et al., 2025; Alpar, 2025). During

drafting, students frequently rely on AI to expand initial ideas into full paragraphs, reorganize arguments for improved coherence, or experiment with alternative syntactic and rhetorical structures – often progressing "from broad concepts to intricacies" through AI-generated scaffolds that they subsequently revise and appropriate into their own voice (Kim et al., 2024; Wang, 2024; Kim et al., 2025). In the revision and editing phases, GenAI systems offer immediate feedback on both global dimensions – such as argument quality, organizational structure, and coherence – and local linguistic features including grammar, diction, and stylistic expression, with some studies noting the added value of blended AI–teacher feedback for promoting richer revision practices (Wang, 2024; Tran, 2025; Escalante et al., 2023).

Research also underscores students' affective and metacognitive gains when AI is incorporated intentionally. Continuous, on-demand support has been shown to reduce writing anxiety, maintain motivation, and prompt reflective engagement, as learners evaluate, adapt, and sometimes contest AI-generated suggestions rather than adopting them uncritically (Kim et al., 2024; Kim et al., 2025). At the same time, concerns regarding overreliance, erosion of authorial voice, and risks to academic integrity emphasize the need for explicit pedagogical guidance and critical AI literacy, ensuring that generative systems augment rather than supplant core composing skills (Kim et al., 2024; Wang, 2024; Raza et al., 2025; Escalante et al., 2023). Viewed through this framing, generative AI is not a shortcut to producing finished texts but a flexible scaffold that – when aligned purposefully with different writing stages – can support students in becoming more strategic, reflective, and agentive writers.

Pre-Writing Support

AI tools are increasingly embedded in the earliest phases of the writing process, when learners may still be struggling to articulate a topic, generate ideas, or outline a preliminary plan. Research across school and higher education contexts demonstrates that students turn to AI chatbots and other AI-enabled systems to overcome "blank page" paralysis, elaborate on nascent ideas, and explore multiple argumentative directions before committing to a draft (Pratiwi et al., 2025; Wang, 2024; Utami et al., 2023; Alpar, 2025; Woo et al., 2022). In this pre-writing space, AI operates less as an automatic text generator and more as a cognitive partner – one that suggests possibilities, poses provocative questions, and provides structural scaffolds to support the organization of emerging thoughts.

Studies with undergraduate and doctoral writers show that tools such as ChatGPT and similar generative platforms are commonly used to brainstorm subthemes, identify angles on a research topic, and connect early ideas with relevant theories or scholarly sources – resources that students then independently revise and develop (Pratiwi et al., 2025; Kim et al., 2024; Utami et al., 2023). Survey and mixed-method studies further indicate that AI-supported "idea exploration" and outlining functionalities facilitate topic selection and help learners move from broad thematic concepts to more focused research questions or thesis claims (Utami et al., 2023; Alpar, 2025). Within language learning contexts, self-built chatbots and GenAI systems have been shown to generate lists of arguments, counterarguments, narrative elements, or provisional outlines, enabling students to visualize a text's overall structure before composing (Guo & Li, 2024; Alpar, 2025; Woo et al., 2022).

At the same time, empirical findings reveal meaningful tensions. Many students report that AI-assisted ideation and planning not only enhance efficiency and reduce writing-related anxiety but also deepen topic knowledge, confidence, and motivation (Pratiwi et al., 2025; Kim et al., 2024; Goldman et al., 2025; Guo & Li, 2024). Yet parallel concerns emerge regarding the potential dilution of originality or the risk of "laziness" in critical and creative thinking when AI is heavily relied upon at this early stage (Pratiwi et al., 2025; Wang, 2024; Utami et al., 2023; Escalante et al., 2023). Such concerns underscore the necessity of guided and ethical AI use – positioning generative tools as scaffolds that support exploration, organization, and questioning, while ensuring that students remain accountable for evaluating suggestions, making deliberate choices, and ultimately owning the plan they carry into drafting.

Drafting and Co-Writing

AI-powered writing tools have evolved far beyond the role of simple spellcheckers; they increasingly function as co-writers that generate, reshape, and extend text alongside human authors. Across academic and creative domains, students and researchers employ large language models (LLMs) to draft sections, rephrase passages, and suggest continuations, engaging in iterative "machine-in-the-loop" workflows in which human and AI contributions become tightly interwoven (Nguyen et al., 2024; Wang, 2024; Singh et al., 2022; Fang et al., 2024). Process-tracing studies of doctoral and undergraduate writers reveal that proficient users do not simply accept AI output at face value. Instead, they engage in cycles of prompting, critiquing, copying, pasting, and revising AI-generated text – gradually transforming it into integrated components of their own manuscripts (Nguyen et al., 2024; Wang, 2024; Kim et al., 2025).

As generative text assistants, AI systems can operate as entry-level writers, producing initial drafts or sample paragraphs that learners subsequently expand, reorganize, and personalize – reducing the cognitive and emotional burden of "getting something on the page," while still maintaining human control over argumentation, authorial voice, and audience awareness (Wang, 2024; Kim et al., 2024; Altmäe et al., 2023). Tools such as Wordtune and Sudowrite exemplify this support function by offering alternative phrasings, tonal variations, expanded or condensed versions of user-generated sentences, and stylistic continuations – resources that are especially beneficial for language learners who may struggle to sustain narrative flow or explore linguistic options spontaneously (Zhao, 2022; Fang et al., 2024; Rahmi et al., 2024). Empirical work conducted in L2 and EFL contexts further reveals that AI-assisted writing can enhance grammatical accuracy, cohesion, and fluency, even though content density and nuanced conceptual development often still require deliberate human supplementation and refinement (Zhao, 2022; Yan, 2023; Rahmi et al., 2024).

Beyond surface-level text improvement, AI can also function as a collaborative partner in knowledge construction, offering counterarguments, examples, or alternative framings that writers evaluate, adapt, and weave into more complex argumentative structures (Cress & Kimmerle, 2023; Kim et al., 2025; Rashid et al., 2025). However, this co-writing potential introduces significant tensions. Research documents perceived gains in efficiency, confidence, and idea development, but also substantial concerns regarding overreliance, erosion of authorial identity, and risks

to academic integrity when AI-generated text is presented as original work (Dergaa et al., 2023; Wang, 2024; Alberts et al., 2023; Barrett & Pack, 2023; Bedington et al., 2024). In response, emerging design-based studies highlight the need for interfaces and pedagogies that preserve human agency – supporting learners in planning, monitoring, and selectively using AI suggestions – so that co-writing with AI strengthens rather than replaces students' strategic, critical, and creative writing capacities (Nguyen et al., 2024; Kim et al., 2025; Krajka & Olszak, 2024).

Revision and Editing

AI-powered writing tools are increasingly reshaping how students receive feedback on their drafts, shifting from basic spellchecking to the provision of multi-layered feedback on grammar, coherence, clarity, and organization. Contemporary automated writing evaluation (AWE) systems and large language model (LLM) – based assistants can identify sentence-level errors, flag awkward or ambiguous phrasing, and evaluate global features such as structural coherence and logical progression – often in real time and at scale (Xuan, 2025; Zheng & Zhang, 2025; Lee et al., 2025; Shi & Aryadoust, 2024; Zhang, 2025). Recent developments in hybrid deep-learning architectures – combining transformer models with sequential networks or incorporating knowledge-graph components – enable assessments that span grammatical accuracy, discourse coherence, and organizational integrity, while generating individualized suggestions that increasingly approximate human evaluators' ratings of writing quality (Xuan, 2025; Zhang, 2025).

Across educational contexts, AI-generated feedback has been shown to support measurable improvements in second-language writers' performance. Tools such as Grammarly, BERT-based diagnostic systems, and ChatGPT-assisted instructional models have demonstrated positive effects on grammatical accuracy, textual coherence, and organizational structure, with randomized trials and classroom-based studies reporting gains in holistic writing quality, cohesion, and coherence compared to traditional instruction alone (Lee et al., 2025; Jaramillo et al., 2025; Wei et al., 2023; Zhai & Xiaomei, 2022; Song & Song, 2023). Among EFL and ESL learners, AI-supported feedback serves as a scaffold for iterative revision cycles: students receive color-coded or itemized comments on errors, clarity, and transitions, revise their drafts accordingly, and gradually produce more logically ordered, cohesive, and fluent texts (Hwang et al., 2023; Jaramillo et al., 2025; Wang et al., 2022; Zhou, 2025).

Despite these benefits, research also identifies notable limitations and design challenges. Systematic reviews emphasize that most AI feedback still disproportionately focuses on surface-level linguistic features – grammar, spelling, punctuation – with comparatively fewer systems offering reliable guidance on higher-order aspects such as argumentation, discourse-level coherence, and rhetorical clarity (Wang et al., 2022; Shi & Aryadoust, 2024). Some studies further indicate that automated diagnostic feedback may outperform student self-assessment in areas such as organization and task fulfillment, yet can paradoxically reduce learners' confidence when comments are perceived as overly critical, opaque, or insufficiently contextualized (Lee et al., 2025). Moreover, empirical evidence suggests that the most substantial improvements in cohesion and organization occur when AI-generated

feedback is deliberately combined with teacher and peer input – positioning AI as one component within a broader feedback ecology rather than a stand-alone evaluator (Escalante et al., 2023; Zhang et al., 2025; Yang et al., 2025; Zhai & Xiaomei, 2022).

Taken together, this emerging body of work suggests that AI-mediated feedback can function as a scalable, "always-on" writing coach capable of helping students refine linguistic form, sharpen clarity, and strengthen organizational logic – provided its use is carefully integrated with human guidance to sustain comprehension, engagement, and long-term writer autonomy.

Formatting and Publishing

AI tools are increasingly capable of transforming rough student drafts into polished, shareable texts by supporting multiple stages of refinement – from sentence-level correction to global restructuring. Large language model (LLM) – based assistants and specialized writing tools now offer grammatical correction, vocabulary enhancement, and smoothing of awkward phrasing, thus producing more fluent and academically conventional language without necessarily altering the writer's underlying ideas (Kim et al., 2024; Song & Song, 2023; Black & Tomlinson, 2025). In practice, learners often paste self-written paragraphs into chatbots and request improvements such as "strengthen this writing," "use more precise vocabulary," or "improve flow." The model then generates a cleaner, more cohesive, and more professional-sounding version of the original text (Song & Song, 2023; Black & Tomlinson, 2025).

Beyond surface-level editing, AI can assist students in reorganizing content so that the final product follows a clearer argumentative trajectory. EFL instructors report that tools such as QuillBot, Wordtune, and ChatGPT support content development and enhance overall organization, leading to essays that more closely align with expected academic patterns (Widiati et al., 2023; Zhao, 2022). Chatbots can generate or revise outlines, recommend more logical sequencing of ideas, tighten introductions and conclusions, and identify missing conceptual links between paragraphs—thereby helping writers to transform loosely structured drafts into more coherent and reader-friendly texts (Kim et al., 2024; Usher & Amzalag, 2025; Song & Song, 2023).

AI-generated feedback also facilitates iterative polishing through repeated cycles of revision. Automated systems and chatbots can provide immediate, targeted feedback on coherence, clarity, and organization, which students then incorporate into successive drafts until the text satisfies assignment or publication expectations (Escalante et al., 2023; Huang & Wilson, 2021; Cotos et al., 2020; Deep & Chen, 2025). Studies indicate that when learners actively engage with this feedback – critiquing AI suggestions, selectively incorporating them, and revising accordingly – the final product tends to show significant gains in cohesion, accuracy, and sophistication of expression (Nguyen et al., 2024; Khuder, 2025; Jin et al., 2024; Song & Song, 2023).

However, research consistently emphasizes that AI should serve as a writing assistant, not a ghostwriter. Both students and tutors highlight the importance of revising AI-generated text to restore personal authorial voice, ensure factual accuracy, and prevent plagiarism – thus ensuring that the final polished version authentically

reflects the student's understanding and effort (Wang, 2024; Dergaa et al., 2023; Eleftheriou et al., 2025; Deep & Chen, 2025; Perkins et al., 2023). When used in such a collaborative manner, AI can streamline the transition from initial draft to final, shareable document while simultaneously supporting – rather than replacing – students' development as independent, reflective writers.

Instructional Design and Classroom Scenarios

Teachers are increasingly expected to navigate the pedagogical implications of AI while continuing to teach writing as a cognitively demanding, meaning-making activity. In real classroom environments – often shaped by time limitations, large class sizes, and uneven access to technological resources – AI cannot simply be appended onto instruction as an optional add-on; rather, it must be deliberately embedded within tasks that promote language development, critical thinking, and ethical use instead of offering shortcuts that bypass learning (Mills et al., 2025; Liu et al., 2021; Tran, 2025). Recent research in AI-supported EFL/ESL writing demonstrates that learning gains, self-efficacy, and engagement are highest when AI tools are positioned as cognitive partners, feedback agents, and reflective prompts within carefully structured instructional sequences (Lai, 2025; Mills et al., 2025; Liu et al., 2021; Rashid et al., 2025; Tran, 2025).

Designing effective AI-mediated writing tasks therefore begins with clarifying the learning goals – whether focused on genre features, argumentation, cohesion, register, or audience – and then determining what roles AI should and should not play. Experimental and classroom-based studies highlight successful approaches in which students draft or outline independently before turning to tools such as ChatGPT or Grammarly to refine ideas, obtain feedback on language and organization, or compare their drafts with AI-generated exemplars (Lai, 2025; Crompton et al., 2024; Mills et al., 2025; Liu et al., 2021; Ghafouri et al., 2024). Sequenced, genre-specific pedagogies that weave AI into brainstorming, outlining, revision, and self-editing – rather than one-click text production – have been shown to improve writing quality, resilience, and self-regulation in both blended and self-directed learning settings (Lai, 2025; Liu et al., 2021; Aladini et al., 2025; Saleem et al., 2025).

A further design principle concerns fostering reflection, agency, and ethical awareness. AI-supported learning environments that explicitly prompt students to evaluate, explain, and selectively incorporate AI feedback tend to produce deeper learning and stronger outcomes than "conventional" AI use alone (Liu et al., 2021). Task structures grounded in co-regulation or active-learning frameworks – where AI is conceptualized as a "more capable other" or "cognitive partner," and positioned alongside peer dialogue and teacher guidance – support the development of metacognitive strategies and sustained creativity, reducing the risk of overreliance (Rashid et al., 2025; Aladini et al., 2025; Li & Wilson, 2025).

At the same time, real-world implementations must attend to practical constraints and risks. Research conducted in under-resourced contexts and career-oriented ESL courses underscores the need for design decisions that consider technological infrastructure, cultural and genre-specific expectations, and explicit guidelines on appropriate AI use (Nguyen et al., 2025; Saleem et al., 2025; Pakdel et al., 2025; Almusharraf et al., 2025). When

thoughtfully designed, AI-supported writing tasks can expand opportunities for feedback, personalization, and engagement – provided teachers anchor them in transparent objectives, scaffolded learning processes, and explicit commitments to integrity and learner autonomy.

Assessment of Writing and Generative AI

AI is now capable of evaluating student writing at scale, ranging from holistic scoring to fine-grained formative feedback. Automated writing evaluation (AWE) systems and more recent large language model (LLM) – based tools analyze dimensions such as grammar, vocabulary, organization, coherence, and rubric alignment, generating scores and comments that increasingly resemble human rater output (Shi & Aryadoust, 2024; Zhang, 2025; Nie, 2025; Atkinson & Palma, 2025; Ramesh & Sanampudi, 2021). Methodological approaches vary along a spectrum – from feature-based scoring models that draw on linguistic indicators of quality, to deep-learning systems (e.g., SBERT, transformer architectures, LSTMs) that capture discourse-level structure and semantic relationships in order to more closely approximate human evaluation (Nie, 2025; Atkinson & Palma, 2025; Ramesh & Sanampudi, 2021; Tan et al., 2025).

Evidence regarding reliability – the extent to which AI ratings are consistent with human scores – is increasingly strong in several learning contexts. Automated essay scoring systems and rubric-based LLM graders commonly achieve high correlations or intraclass correlations with trained human raters, in some cases exceeding 0.9, suggesting near–human-level scoring consistency in constrained tasks such as short responses or rubric-guided essay assessments (Seneviratne & Manathunga, 2025; Atkinson & Palma, 2025; Banihashem et al., 2024; Yavuz et al., 2024). Hybrid architectures and knowledge-graph-enhanced scoring further improve stability and mitigate annotation bias by integrating linguistic layers and applying cross-validation strategies (Zhang, 2025; Nie, 2025; Tan et al., 2025).

Findings on validity – whether AI-generated scores accurately represent the construct of writing ability – are more mixed and context-specific. Systematic reviews of AI-based written feedback report stronger outcomes related to usability and student learning impact than to empirically rigorous construct validation, with concerns persisting around limited attention to content relevance, higher-order reasoning, and genre-specific rhetorical demands (Shi & Aryadoust, 2024; Khojasteh et al., 2025; Ramesh & Sanampudi, 2021). Comparative studies consistently demonstrate that, although AI systems can match humans on surface-level accuracy and criteria-based scoring metrics, expert teachers continue to outperform automated models in prioritizing essential writing features, interpreting nuance, and aligning evaluative decisions with complex educational objectives (Steiss et al., 2024; Al-Obaydi & Pikhart, 2025; Gao, 2021; Ramesh & Sanampudi, 2021).

From an instructional standpoint, AI-mediated writing evaluation appears most effective when used as a formative aid rather than a stand-alone assessor. In large or online courses, AI-generated scoring and feedback can dramatically expand access to timely responses on drafts, support peer and self-assessment, and facilitate iterative revision. Multiple studies document improvements in feedback quality, writing performance, and feedback literacy when AI tools are thoughtfully integrated with teacher guidance and classroom assessment practices (Shi

& Aryadoust, 2024; Khojasteh et al., 2025; Xuan, 2025; Lee, 2023; Guo et al., 2024; Gozali et al., 2024). Simultaneously, research cautions against risks of overreliance, opacity, and algorithmic bias—underscoring the importance of blended human-AI scoring models where automated systems augment, rather than replace, teacher judgment and where learners receive explicit instruction in interpreting and critically evaluating automated feedback (Shi & Aryadoust, 2024; Khojasteh et al., 2025; Al-Obaydi & Pikhart, 2025; Tan et al., 2025; Gozali et al., 2024).

Opportunities, Challenges, and Ethical Concerns

AI is rapidly becoming embedded in both writing and assessment systems, compelling educators to balance its pedagogical benefits against serious concerns related to fairness, plagiarism, and broader long-term educational consequences. Studies involving EFL, ENL, and L2 writers demonstrate that when its use is structured and guided, AI-assisted instruction can enhance organization, coherence, grammatical accuracy, vocabulary development, motivation, and learner engagement (Song & Song, 2023; Escalante et al., 2023; Chen & Gong, 2025; Li & Long, 2025; Zhang, 2025). Students frequently experience AI tools as multitasking assistants, virtual tutors, and digital peers that make drafting and revision more efficient and less anxiety-provoking (Kim et al., 2024; Song & Song, 2023; Deep & Chen, 2025; Malik et al., 2023; Chen & Gong, 2025). At the programmatic level, AI can expand access to feedback, support personalized and differentiated learning pathways, and help cultivate emerging 21st-century literacies – including multimodal composition and sustainability-oriented awareness (Crompton et al., 2024; Yıldız, 2025).

Yet these advantages are tightly intertwined with significant pedagogical risks. Learners and teachers consistently report tendencies toward overreliance, diminished critical thinking, and shallow cognitive engagement when AI is used to generate – rather than support – writing, particularly in unguided settings or high-stakes academic contexts (Song & Song, 2023; Chan & Hu, 2023; Pakdel et al., 2025; Chen & Gong, 2025; Özçelik & Ekşi, 2024; Zhang, 2025). Risks related to accuracy, hallucinated references, and limited contextual comprehension are especially pronounced in disciplines that rely on evidence-based reasoning and domain-specific knowledge (Song & Song, 2023; Chan & Hu, 2023; Pakdel et al., 2025). Educators also express concerns that habitual dependence on AI could erode human writing skills, reduce creativity, and weaken the perceived academic value of higher education credentials (Gustilo et al., 2024; Chan & Hu, 2023; Malik et al., 2023; Pakdel et al., 2025).

Issues of fairness and academic integrity underscore these debates. Access to advanced AI tools remains uneven across students, institutions, and socioeconomic contexts, raising concerns that AI-enhanced writing may widen existing inequities rather than close them (Gustilo et al., 2024; Pakdel et al., 2025; Alotaibi, 2024). Complicating matters further, AI-generated text can evade plagiarism detection and blur traditional standards of authorship, making it increasingly difficult to distinguish legitimate support from forms of contract cheating or undisclosed co-authorship (Perkins, 2023; Gustilo et al., 2024; Chan & Hu, 2023; Yan, 2023; Yeo, 2023). Scholars increasingly argue for transparency, updated academic policies, and explicit AI literacy education – reframing

fairness as the expectation of disclosed, critically mediated AI collaboration rather than hidden automation (Perkins, 2023; Yeo, 2023; Zhang, 2025; Eager & Brunton, 2023).

The long-term implications of AI for writing pedagogy and higher education remain uncertain yet profound. Optimistic perspectives highlight opportunities for scalable and inclusive instructional support, particularly for students who have been historically underserved, as well as possibilities for new forms of human-AI collaborative authorship (Deep & Chen, 2025; Crompton et al., 2024; Zhang, 2025; Alotaibi, 2024). More cautionary viewpoints warn that if AI replaces rather than augments student work, core educational goals – writing proficiency, critical thinking, epistemic agency, and the meaning of academic credentials – may be fundamentally undermined (Song & Song, 2023; Chan & Hu, 2023; Pakdel et al., 2025; Nguyen et al., 2024). Sustainable integration of AI into writing instruction therefore requires addressing not only immediate classroom affordances and risks but also how contemporary design choices will shape future norms of authorship, learning, and expertise.

Conclusion

Across theoretical, technological, and pedagogical perspectives, this chapter has portrayed generative AI not as a neutral tool, but as a powerful and ambivalent mediator of writing and assessment. When grounded in process–genre pedagogies and learning theories, and when embedded in carefully sequenced tasks, AI can function as a configurable scaffold: expanding access to feedback, supporting pre-writing and revision, enhancing organization and cohesion, and fostering greater confidence and engagement among diverse writers. It can also help teachers manage large classes, extend formative assessment, and cultivate new literacies that prepare students for increasingly AI-saturated communicative landscapes.

At the same time, the chapter has underscored persistent risks related to overreliance, reduced critical thinking, fairness, academic integrity, and the potential erosion of human authorship and expertise. Evidence suggests that AI works best as a formative, “always-on” assistant whose suggestions are evaluated, questioned, and selectively adopted within a broader ecology of teacher and peer feedback – not as an opaque judge or invisible ghostwriter. Any sustainable integration of GenAI into writing pedagogy must therefore remain theory-informed, context-sensitive, and policy-aware, explicitly framing AI use as disclosed, critically mediated collaboration. Ultimately, the future of writing in education will be shaped less by what AI can technically do, and more by how educators, institutions, and students choose to align these capabilities with enduring goals of writing proficiency, epistemic agency, and ethical participation in knowledge-making communities.

Acknowledgements or Notes

Generative artificial intelligence tools were employed to support the conceptualization of this book chapter, assist in the literature review, contribute to the drafting process, and facilitate language editing.

References

Al-Obaydi, L., & Pikhart, M. (2025). AI partner versus human partner: comparing AI-based peer assessment with human-generated peer assessment in examining writing skills. *Language Testing in Asia*, 15. https://doi.org/10.1186/s40468-025-00375-8

Al-Saliti, R., Ismail, A., Elmorsy, G., & Shahpo, S. (2025). Teaching Writing Skills Using Generative AI: The Paradox of Adoption and Resistance Among Language Educators. *World Journal of English Language*. https://doi.org/10.5430/wjel.v15n8p28

Aladini, A., Ismail, S., Khasawneh, M., & Shakibaei, G. (2025). Self-directed writing development across computer/AI-based tasks: Unraveling the traces on L2 writing outcomes, growth mindfulness, and grammatical knowledge. *Computers in Human Behavior Reports*. https://doi.org/10.1016/j.chbr.2024.100566

Alberts, I., Mercolli, L., Pyka, T., Prenosil, G., Shi, K., Rominger, A., & Afshar-Oromieh, A. (2023). Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be?. *European Journal of Nuclear Medicine and Molecular Imaging*, 50, 1549 - 1552. https://doi.org/10.1007/s00259-023-06172-w

Almusharraf, A., Bailey, D., Almusharraf, N., & Alotaibi, T. (2025). Students' perceptions of generative AI in EFL writing: Strategies, self-efficacy, satisfaction and behavioural intention. *Australasian Journal of Educational Technology*. https://doi.org/10.14742/ajet.10045

Alotaibi, N. (2024). The Impact of AI and LMS Integration on the Future of Higher Education: Opportunities, Challenges, and Strategies for Transformation. *Sustainability*. https://doi.org/10.3390/su162310357

Alpar, Ö. (2025). Evaluating Generative AI Tools for Improving English Writing Skills: A Preliminary Comparison of ChatGPT-4, Google Gemini, and Microsoft Copilot. *European Journal of Educational Research*. https://doi.org/10.12973/eu-jer.14.4.1291

Altmäe, S., Sola-Leyva, A., & Salumets, A. (2023). Artificial intelligence in scientific writing: a friend or a foe?. *Reproductive biomedicine online*. https://doi.org/10.1016/j.rbmo.2023.04.009

Atkinson, J., & Palma, D. (2025). An LLM-based hybrid approach for enhanced automated essay scoring. *Scientific Reports*, 15. https://doi.org/10.1038/s41598-025-87862-3

Badger, R., & White, G. (2000). A process genre approach to teaching writing. *Elt Journal*, 54, 153-160. https://doi.org/10.1093/elt/54.2.153

Bahari, A., Han, F., & Strzelecki, A. (2025). Integrating CALL and AIALL for an interactive pedagogical model of language learning. *Education and Information Technologies*, 30, 14305 - 14333. https://doi.org/10.1007/s10639-025-13388-w

Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, 33, 1-17. https://doi.org/10.1007/s12525-023-00680-1

Banihashem, S., Kerman, N., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: peer-generated or AI-generated feedback?. *International Journal of Educational Technology in Higher Education*, 21. https://doi.org/10.1186/s41239-024-00455-4

Barrett, A., & Pack, A. (2023). Not quite eye to A.I.: student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education*, 20, 1-24. https://doi.org/10.1186/s41239-023-00427-0

Bedington, A., Halcomb, E., McKee, H., Sargent, T., & Smith, A. (2024). Writing with generative AI and human-

machine teaming: Insights and recommendations from faculty and students. *Computers and Composition*. https://doi.org/10.1016/j.compcom.2024.102833

Bin-Hady, W., Nasser, A., & Al-Kadi, A. (2020). A Pre-Experimental Study on a Process-Genre Approach for Teaching Essay Writing. *NRU HSE: Journal of Language & Education (Topic)*. https://doi.org/10.2139/ssrn.3862325

Black, R., & Tomlinson, B. (2025). University students describe how they adopt AI for writing and research in a general education course. *Scientific Reports*, 15. https://doi.org/10.1038/s41598-025-92937-2

Chan, C., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20. https://doi.org/10.1186/s41239-023-00411-8

Chanpradit, T. (2025). Generative artificial intelligence in academic writing in higher education: A systematic review. *Edelweiss Applied Science and Technology*. https://doi.org/10.55214/25768484.v9i4.6128

Chen, C., & Gong, Y. (2025). The Role of AI-Assisted Learning in Academic Writing: A Mixed-Methods Study on Chinese as a Second Language Students. *Education Sciences*. https://doi.org/10.3390/educsci15020141

Chen, Q. (2025). Opportunities and Limitations of Using Grammarly to Support EFL Writing with Generative AI. *Business and Social Sciences Proceedings*. https://doi.org/10.71222/gswrgp34

Chiu, T., Moorhouse, B., Chai, C., & Ismailov, M. (2023). Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot. *Interactive Learning Environments*, 32, 3240 - 3256. https://doi.org/10.1080/10494820.2023.2172044

Colby, R. (2025). Playing the digital dialectic game: Writing pedagogy with generative AI. *Computers and Composition*. https://doi.org/10.1016/j.compcom.2025.102915

Cotos, E., Huffman, S., & Link, S. (2020). Understanding Graduate Writers' Interaction with and Impact of the Research Writing Tutor during Revision. *Journal of Writing Research*. https://doi.org/10.17239/jowr-2020.12.01.07

Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSCL perspective. *International Journal of Computer-Supported Collaborative Learning*, 18, 607-614. https://doi.org/10.1007/s11412-023-09409-w

Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *Br. J. Educ. Technol.*, 55, 2503-2529. https://doi.org/10.1111/bjet.13460

Deep, P., & Chen, Y. (2025). The Role of AI in Academic Writing: Impacts on Writing Skills, Critical Thinking, and Integrity in Higher Education. *Societies*. https://doi.org/10.3390/soc15090247

Dergaa, I., Chamari, K., Żmijewski, P., & Saad, B. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing.. *Biology of sport*, 40 2, 615-622. https://doi.org/10.5114/biolsport.2023.125623

Eager, B., & Brunton, R. (2023). Prompting Higher Education Towards AI-Augmented Teaching and Learning Practice. *Journal of University Teaching and Learning Practice*. https://doi.org/10.53761/1.20.5.02

Eleftheriou, M., Ahmer, M., & Fredrick, D. (2025). Balancing ethics and support: Peer tutors' experiences with AI tools in student writing. *Contemporary Educational Technology*. https://doi.org/10.30935/cedtech/16554

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, 1-20. https://doi.org/10.1186/s41239-023-00425-2

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, 1-20. https://doi.org/10.1186/s41239-023-00425-2

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2024). Beware of Metacognitive Laziness: Effects of Generative Artificial Intelligence on Learning Motivation, Processes, and Performance. *ArXiv*, abs/2412.09315. https://doi.org/10.1111/bjet.13544

Fang, X., Guo, K., & Ng, D. (2024). Sudowrite: Co-Writing Creative Stories with Artificial Intelligence. *RELC Journal*. https://doi.org/10.1177/00336882241250109

Ganapathy, M., Kaur, M., Jamal, M., & Phan, J. (2022). THE EFFECT OF A GENRE-BASED PEDAGOGICAL APPROACH ON ORANG ASLI STUDENTS' EFL WRITING PERFORMANCE. *Malaysian Journal of Learning and Instruction*. https://doi.org/10.32890/mjli2022.19.1.4

Gao, J. (2021). Exploring the Feedback Quality of an Automated Writing Evaluation System Pigai. *Int. J. Emerg. Technol. Learn.*, 16. https://doi.org/10.3991/ijet.v16i11.19657

Ghafouri, M., Hassaskhah, J., & Mahdavi-Zafarghandi, A. (2024). From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill. *Language Teaching Research*. https://doi.org/10.1177/13621688241239764

Gibson, D., Kovanović, V., Ifenthaler, D., Dexter, S., & Feng, S. (2023). Learning theories for artificial intelligence promoting learning processes. *Br. J. Educ. Technol.*, 54, 1125-1146. https://doi.org/10.1111/bjet.13341

Goldman, S., Smith, S., & Carreon, A. (2025). Artificial Intelligence to Support Writing Outcomes for Students With Disabilities. *Journal of Special Education Technology*, 40, 418 - 426. https://doi.org/10.1177/01626434251326328

Gozali, I., Wijaya, A., Lie, A., Cahyono, B., & Suryati, N. (2024). Leveraging the potential of ChatGPT as an automated writing evaluation (AWE) tool: Students' feedback literacy development and AWE tools integration framework. *The JALT CALL Journal*. https://doi.org/10.29140/jaltcall.v20n1.1200

Guan, L., Zhang, E., & Gu, M. (2024). Examining generative AI–mediated informal digital learning of English practices with social cognitive theory: a mixed-methods study. *ReCALL*. https://doi.org/10.1017/s0958344024000259

Guo, K., & Li, D. (2024). Understanding EFL students' use of self-made AI chatbots as personalized writing assistance tools: A mixed methods study. *System*. https://doi.org/10.1016/j.system.2024.103362

Guo, K., Pan, M., Li, Y., & Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. *Internet High. Educ.*, 63, 100962. https://doi.org/10.1016/j.iheduc.2024.100962

Gustilo, L., Ong, E., & Lapinid, M. (2024). Algorithmically-driven writing and academic integrity: exploring educators' practices, perceptions, and policies in AI era. *International Journal for Educational Integrity*, 20, 1-43. https://doi.org/10.1007/s40979-024-00153-8

Hagos, D., Battle, R., & Rawat, D. (2024). Recent Advances in Generative AI and Large Language Models: Current Status, Challenges, and Perspectives. *IEEE Transactions on Artificial Intelligence*, 5, 5873-5893. https://doi.org/10.1109/tai.2024.3444742

Han, Y. (2025). Beyond the Algorithm: Reconciling Generative AI and Human Agency in Academic Writing

Education. *International Journal of Learning and Teaching*. https://doi.org/10.18178/ijlt.11.1.39-42

Hanusova, S., Dontcheva-Navratilova, O., Vališová, M., & Matulová, M. (2020). Process genre approach to L2 academic writing: An intervention study. *XLinguae*. https://doi.org/10.18355/xl.2020.13.04.03

Hong, H., Vate-U-Lan, P., & Viriyavejakul, C. (2025). Generative AI–mediated scaffolds for enhanced critical thinking in EFL writing. *Edelweiss Applied Science and Technology*. https://doi.org/10.55214/25768484.v9i6.7751

Huang, Y., & Wilson, J. (2021). Using automated feedback to develop writing proficiency. *Computers and Composition*. https://doi.org/10.1016/j.compcom.2021.102675

Hwang, W., Nurtantyana, R., Purba, S., Hariyanti, U., Indrihapsari, Y., & Surjono, H. (2023). AI and Recognition Technologies to Facilitate English as Foreign Language Writing for Supporting Personalization and Contextualization in Authentic Contexts. *Journal of Educational Computing Research*, 61, 1008 - 1035. https://doi.org/10.1177/07356331221137253

Hyland, K. (2003). Genre-Based Pedagogies: A Social Response to Process.. *Journal of Second Language Writing*, 12, 17-29. https://doi.org/10.1016/s1060-3743(02)00124-8

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, 16, 148-164. https://doi.org/10.1016/j.jslw.2007.07.005

Insuwan, C., & Thongrin, S. (2025). Empowering Thai EFL Learners as Critical Thinkers and Skilled Writers: A Genre-Based Approach with Critical Pedagogy. *rEFLections*. https://doi.org/10.61508/refl.v32i1.280405

Jaramillo, J., Chiappe, A., & Delgado, F. (2025). From Struggle to Mastery: AI-Powered Writing Skills in ESL Education. *Applied Sciences*. https://doi.org/10.3390/app15148079

Jarunthawatchai, O., Jarunthawatchai, W., & Gilbert, L. (2025). Connecting Reading and Writing in Foreign Language Instruction: A Process-genre Approach. *LEARN Journal: Language Education and Acquisition Research Network*. https://doi.org/10.70730/sogp3675

Jin, F., Lin, C., & Lai, C. (2024). Modeling AI-assisted writing: How self-regulated learning influences writing outcomes. *Comput. Hum. Behav.*, 165, 108538. https://doi.org/10.1016/j.chb.2024.108538

Khojasteh, L., Kafipour, R., Pakdel, F., & Mukundan, J. (2025). Empowering medical students with AI writing co-pilots: design and validation of AI self-assessment toolkit. *BMC Medical Education*, 25. https://doi.org/10.1186/s12909-025-06753-3

Khuder, B. (2025). Enhancing disciplinary voice through feedback-seeking in AI-assisted doctoral writing for publication. *Applied Linguistics*. https://doi.org/10.1093/applin/amaf022

Kim, J., Lee, S., Detrick, R., Wang, J., & Li, N. (2025). Students-Generative AI interaction patterns and its impact on academic writing. *Journal of Computing in Higher Education*. https://doi.org/10.1007/s12528-025-09444-6

Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on Generative AI-assisted academic writing. *Education and Information Technologies*, 30, 1265 - 1300. https://doi.org/10.1007/s10639-024-12878-7

Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on Generative AI-assisted academic writing. *Education and Information Technologies*, 30, 1265 - 1300. https://doi.org/10.1007/s10639-024-12878-7

Kim, S., So, H., & Park, K. (2025). Supporting learner agency in collaborative writing with generative AI. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.70015

Kong, S., & Yang, Y. (2024). A Human-Centered Learning and Teaching Framework Using Generative Artificial Intelligence for Self-Regulated Learning Development Through Domain Knowledge Learning in K–12 Settings. *IEEE Transactions on Learning Technologies*, 17, 1588-1599. https://doi.org/10.1109/tlt.2024.3392830

Krajka, J., & Olszak, I. (2024). Artificial Intelligence Tools in Academic Writing Instruction: Exploring the Potential of On-Demand AI Assistance in the Writing Process. *Roczniki Humanistyczne*. https://doi.org/10.18290/rh247206.8

Kumar, P. (2024). Large language models (LLMs): survey, technical frameworks, and future challenges. *Artificial Intelligence Review*, 57. https://doi.org/10.1007/s10462-024-10888-y

Lai, Z. (2025). The Impact of AI-Assisted Blended Learning on Writing Efficacy and Resilience. *International Journal of Computer-Assisted Language Learning and Teaching*. https://doi.org/10.4018/ijcallt.377174

Lai, Z. (2025). The Impact of AI-Assisted Blended Learning on Writing Efficacy and Resilience. *International Journal of Computer-Assisted Language Learning and Teaching*. https://doi.org/10.4018/ijcallt.377174

Lee, A. (2023). Supporting students' generation of feedback in large-scale online course with artificial intelligence-enabled evaluation. *Studies in Educational Evaluation*. https://doi.org/10.1016/j.stueduc.2023.101250

Lee, M., Jang, E., & Hannah, L. (2025). Automated Diagnostic Feedback vs. Self-Assessment: Rethinking Feedback Mechanisms on Academic Writing Development. *TESOL Quarterly*. https://doi.org/10.1002/tesq.70032

Li, C., & Long, J. (2025). Comparing AI-Assisted and Traditional Teaching in College English: Pedagogical Benefits and Learning Behaviors. *Information*. https://doi.org/10.3390/info16100895

Li, J., Tang, T., Zhao, W., Nie, J., & Wen, J. (2022). Pre-Trained Language Models for Text Generation: A Survey. *ACM Computing Surveys*, 56, 1 - 39. https://doi.org/10.1145/3649449

Li, M., & Wilson, J. (2025). AI-Integrated Scaffolding to Enhance Agency and Creativity in K-12 English Language Learners: A Systematic Review. *Information*. https://doi.org/10.3390/info16070519

Liu, C., Hou, J., Tu, Y., Wang, Y., & Hwang, G. (2021). Incorporating a reflective thinking promoting mechanism into artificial intelligence-supported English writing environments. *Interactive Learning Environments*, 31, 5614 - 5632. https://doi.org/10.1080/10494820.2021.2012812

Malik, A., Pratiwi, Y., Andajani, K., Numertayasa, I., Suharti, S., Darwis, A., & , M. (2023). Exploring Artificial Intelligence in Academic Essay: Higher Education Student's Perspective. *International Journal of Educational Research Open*. https://doi.org/10.1016/j.ijedro.2023.100296

Maphoto, K., Sevnarayan, K., Mohale, N., Suliman, Z., Ntsopi, T., & Mokoena, D. (2024). Advancing Students' Academic Excellence in Distance Education: Exploring the Potential of Generative AI Integration to Improve Academic Writing Skills. *Open Praxis*. https://doi.org/10.55982/openpraxis.16.2.649

Mills, N., Hok, H., Dressen, A., & Veillas, Q. (2025). The design and evaluation of an interactive AI companion for foreign language writing. *Foreign Language Annals*. https://doi.org/10.1111/flan.12790

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, 49, 847 - 864. https://doi.org/10.1080/03075079.2024.2323593

Nguyen, L., Dinh, H., Dao, T., & Tran, N. (2025). Teachers' Perceptions and Students' Strategies in Using AI-Mediated Informal Digital Learning for Career ESL Writing. *Education*

Sciences. https://doi.org/10.3390/educsci15101414

Nie, Y. (2025). Automated essay scoring with SBERT embeddings and LSTM-Attention networks. *PeerJ Computer Science*, 11. https://doi.org/10.7717/peerj-cs.2634

Ooi, K., Tan, G., Al-Emran, M., Al-Sharafi, M., Căpățînă, A., Chakraborty, A., Dwivedi, Y., Huang, T., Kar, A., Lee, V., Loh, X., Micu, A., Mikalef, P., Mogaji, E., Pandey, N., Raman, R., Rana, N., Sarker, P., Sharma, A., Teng, C., Wamba, S., & Wong, L. (2023). The Potential of Generative Artificial Intelligence Across Disciplines: Perspectives and Future Directions. *Journal of Computer Information Systems*, 65, 76 - 107. https://doi.org/10.1080/08874417.2023.2261010

Orfanò, B., & Wingate, U. (2024). Teaching Academic Writing to Nursing Students: Developing Criticality Through a Genre-Based Approach. *ESP Today*. https://doi.org/10.18485/esptoday.2024.12.1.6

Özçelik, N., & Ekşi, G. (2024). Cultivating writing skills: the role of ChatGPT as a learning assistant—a case study. *Smart Learning Environments*, 11. https://doi.org/10.1186/s40561-024-00296-8

Pakdel, F., Khojasteh, L., Kafipour, R., & Shahsavar, Z. (2025). Navigating AI writing tools in medical education: A SWOT analysis of L2 academic writing perspectives. *Language Teaching Research*. https://doi.org/10.1177/13621688251322953

Parker, J., Richard, V., Acabá, A., Escoffier, S., Flaherty, S., Jablonka, S., & Becker, K. (2024). Negotiating Meaning with Machines: AI's Role in Doctoral Writing Pedagogy. *International Journal of Artificial Intelligence in Education*, 35, 1218 - 1238. https://doi.org/10.1007/s40593-024-00425-x

Perkins, M. (2023). Academic integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*. https://doi.org/10.53761/1.20.02.07

Perkins, M., Roe, J., Postma, D., McGaughran, J., Vietnam, D., , V., Singapore, J., & , S. (2023). Detection of GPT-4 Generated Text in Higher Education: Combining Academic Judgement and Software to Identify Generative AI Tool Misuse. *Journal of Academic Ethics*, 22, 89-113. https://doi.org/10.1007/s10805-023-09492-6

Phạm, V., & Bui, T. (2021). Genre-based Approach to Writing in EFL Contexts. *World Journal of English Language*. https://doi.org/10.5430/wjel.v11n2p95

Pratiwi, H., , S., , H., & Ridha, M. (2025). Between Shortcut and Ethics: Navigating the Use of Artificial Intelligence in Academic Writing Among Indonesian Doctoral Students. *European Journal of Education*. https://doi.org/10.1111/ejed.70083

Rahmi, R., Amalina, Z., Andriansyah, A., & Rodgers, A. (2024). Does it really help? Exploring the impact of AI-Generated writing assistant on the students' English writing. *Studies in English Language and Education*. https://doi.org/10.24815/siele.v11i2.35875

Ramesh, D., & Sanampudi, S. (2021). An automated essay scoring systems: a systematic literature review. *Artificial Intelligence Review*, 55, 2495 - 2527. https://doi.org/10.1007/s10462-021-10068-2

Rashid, S., Malik, S., & Ghauri, F. (2025). The Active Learning–GenAI Synergy Framework: Ethical Integration of Generative AI in EFL/ESL Writing in Resource-Constrained Contexts. *F1000Research*. https://doi.org/10.12688/f1000research.171435.1

Raza, M., Jahangir, Z., Riaz, M., Saeed, M., & Sattar, M. (2025). Industrial applications of large language models. *Scientific Reports*, 15. https://doi.org/10.1038/s41598-025-98483-1

Saleem, T., Saleem, A., & Aslam, M. (2025). Integrating AI in Pakistani ESL classrooms: Teachers' practices,

perspectives, and impact on student performance. *PLOS One*, 20. https://doi.org/10.1371/journal.pone.0333352

Seneviratne, H., & Manathunga, S. (2025). Artificial intelligence assisted automated short answer question scoring tool shows high correlation with human examiner markings. *BMC Medical Education*, 25. https://doi.org/10.1186/s12909-025-07718-2

Sengar, S., Hasan, A., Kumar, S., & Carroll, F. (2024). Generative artificial intelligence: a systematic review and applications. *Multimedia Tools and Applications*, 84, 23661 - 23700. https://doi.org/10.1007/s11042-024-20016-1

Seran, C., Tan, M., Karim, H., & Aldahoul, N. (2025). A conceptual exploration of generative AI-induced cognitive dissonance and its emergence in university-level academic writing. *Frontiers in Artificial Intelligence*, 8. https://doi.org/10.3389/frai.2025.1573368

Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*. https://doi.org/10.1017/s0958344023000265

Singh, N., Bernal, G., Savchenko, D., & Glassman, E. (2022). Where to Hide a Stolen Elephant: Leaps in Creative Writing with Multimodal Machine Intelligence. *ACM Transactions on Computer-Human Interaction*, 30, 1 - 57. https://doi.org/10.1145/3511599

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14. https://doi.org/10.3389/fpsyg.2023.1260843

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2024.101894

Tan, L., Hu, S., Yeo, D., & Cheong, K. (2025). A Comprehensive Review on Automated Grading Systems in STEM Using AI Techniques. *Mathematics*. https://doi.org/10.3390/math13172828

Tate, T., Harnick-Shapiro, B., Ritchie, D., Tseng, W., Dennin, M., & Warschauer, M. (2025). Incorporating generative AI into a writing-intensive undergraduate course without off-loading learning. *Discover Computing*, 28. https://doi.org/10.1007/s10791-025-09563-9

Tran, T. (2025). Enhancing EFL Writing Revision Practices: The Impact of AI- and Teacher-Generated Feedback and Their Sequences. *Education Sciences*. https://doi.org/10.3390/educsci15020232

Usher, M., & Amzalag, M. (2025). From Prompt to Polished: Exploring Student–Chatbot Interactions for Academic Writing Assistance. *Education Sciences*. https://doi.org/10.3390/educsci15030329

Utami, S., Andayani, A., Winarni, R., & Sumarwati, S. (2023). Utilization of artificial intelligence technology in an academic writing class: How do Indonesian students perceive?. *Contemporary Educational Technology*. https://doi.org/10.30935/cedtech/13419

Venkatasubramanian, V., & Chakraborty, A. (2024). Quo Vadis ChatGPT? From Large Language Models to Large Knowledge Models. *Comput. Chem. Eng.*, 192, 108895. https://doi.org/10.1016/j.compchemeng.2024.108895

Wang, C. (2024). Exploring Students' Generative AI-Assisted Writing Processes: Perceptions and Experiences from Native and Nonnative English Speakers. *Technology, Knowledge and Learning*, 30, 1825 - 1846. https://doi.org/10.1007/s10758-024-09744-3

Wang, Y., Luo, X., Liu, C., Tu, Y., & Wang, N. (2022). An Integrated Automatic Writing Evaluation and SVVR Approach to Improve Students' EFL Writing Performance. *Sustainability*. https://doi.org/10.3390/su141811586

Wegerif, R., & Casebourne, I. (2025). A dialogic theoretical foundation for integrating generative AI into pedagogical design. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.70026

Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: a randomized controlled trial. *Frontiers in Psychology*, 14. https://doi.org/10.3389/fpsyg.2023.1249991

Widiati, U., Rusdin, D., Indrawati, I., , M., & Govender, N. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, 10. https://doi.org/10.1080/2331186x.2023.2236469

Woo, D., Wang, Y., Susanto, H., & Guo, K. (2022). Understanding English as a Foreign Language Students' Idea Generation Strategies for Creative Writing With Natural Language Generation Tools. *Journal of Educational Computing Research*, 61, 1464 - 1482. https://doi.org/10.1177/07356331231175999

Xia, C. (2025). Research on Strategies for Generative Artificial Intelligence Empowering High School English Writing Teaching. *International Journal of Education and Social Development*. https://doi.org/10.54097/5d8cx254

Xu, D., Chen, W., Peng, W., Zhang, C., Xu, T., Zhao, X., Wu, X., Zheng, Y., & Chen, E. (2023). Large language models for generative information extraction: a survey. *Frontiers of Computer Science*, 18. https://doi.org/10.1007/s11704-024-40555-y

Xuan, Y. (2025). Transformer-LSTM Models for Automatic Scoring and Feedback in English Writing Assessment. *IEEE Access*, 13, 82084-82096. https://doi.org/10.1109/access.2025.3562493

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28, 13943-13967. https://doi.org/10.1007/s10639-023-11742-4

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28, 13943-13967. https://doi.org/10.1007/s10639-023-11742-4

Yang, H., Zhang, Y., & Guo, J. (2025). Exploring the Effectiveness of Cooperative Pre-Service Teacher and Generative AI Writing Feedback on Chinese Writing. *Behavioral Sciences*, 15. https://doi.org/10.3390/bs15040518

Yao, Y., Zhu, X., Xiao, L., & Lu, Q. (2025). Secondary school English teachers' application of artificial intelligence-guided chatbot in the provision of feedback on student writing: An activity theory perspective. *Journal of Second Language Writing*. https://doi.org/10.1016/j.jslw.2025.101179

Yavuz, F., Çelik, Ö., & Çelik, G. (2024). Utilizing large language models for EFL essay grading: An examination of reliability and validity in rubric-based assessments. *Br. J. Educ. Technol.*, 56, 150-166. https://doi.org/10.1111/bjet.13494

Yeo, M. (2023). Academic integrity in the age of Artificial Intelligence (AI) authoring apps. *TESOL Journal*. https://doi.org/10.1002/tesj.716

Yıldız, T. (2025). Sustainability Education in L2 Writing: AI-Based Multimodal Awareness and Engagement. *Sustainability*. https://doi.org/10.3390/su17219376

Yusuf, A., Pervin, N., Román-González, M., & Noor, N. (2024). Generative AI in education and research: A systematic mapping review. *Review of Education*. https://doi.org/10.1002/rev3.3489

Zamorano, C. (2025). Enhancing academic writing in English language education through generative AI integration. *Research Studies in English Language Teaching and Learning*. https://doi.org/10.62583/rseltl.v3i3.87

Zhai, N., & , X. (2022). The Effectiveness of Automated Writing Evaluation on Writing Quality: A Meta-Analysis. *Journal of Educational Computing Research*, 61, 875 - 900. https://doi.org/10.1177/07356331221127300

Zhang, C. (2025). Optimising AI writing assessment using feedback and knowledge graph integration. *PeerJ Computer Science*, 11. https://doi.org/10.7717/peerj-cs.2893

Zhang, H., Chen, C., Liu, Y., & Peng, L. (2025). From Dependence to Development: The Role of AI in Enhancing Writing Proficiency Among Chinese English Majors. *Journal of Posthumanism*. https://doi.org/10.63332/joph.v5i8.3159

Zhang, M. (2025). Optimizing academic engagement and mental health through AI: an experimental study on LLM integration in higher education. *Frontiers in Psychology*, 16. https://doi.org/10.3389/fpsyg.2025.1641212

Zhang, T., & Zhang, L. (2021). Taking Stock of a Genre-Based Pedagogy: Sustaining the Development of EFL Students' Knowledge of the Elements in Argumentation and Writing Improvement. *Sustainability*. https://doi.org/10.3390/su132111616

Zhang, Y., Jalaluddin, I., Mamat, R., & Zhao, T. (2025). The Effects of Automated Writing Evaluation (AWE)+Human Feedback on the Quality of Argumentative Writing by Chinese EFL Learners. *Journal of Language Teaching and Research*. https://doi.org/10.17507/jltr.1603.20

Zhao, D. (2024). The impact of AI-enhanced natural language processing tools on writing proficiency: an analysis of language precision, content summarization, and creative writing facilitation. *Education and Information Technologies*, 30, 8055 - 8086. https://doi.org/10.1007/s10639-024-13145-5

Zhao, X. (2022). Leveraging Artificial Intelligence (AI) Technology for English Writing: Introducing Wordtune as a Digital Writing Assistant for EFL Writers. *RELC Journal*, 54, 890 - 894. https://doi.org/10.1177/00336882221094089

Zheng, X., & Zhang, J. (2025). The usage of a transformer based and artificial intelligence driven multidimensional feedback system in english writing instruction. *Scientific Reports*, 15. https://doi.org/10.1038/s41598-025-05026-9

Zhou, L. (2025). A Case Study on the Impact of ChatGPT Feedback on Underachieving Students' Writing Ability in Senior High School English Instruction. *International Journal of Educational Development*. https://doi.org/10.63313/ijed.9008

Author Information

Hamza Polat

https://orcid.org/0000-0002-9646-7507

Atatürk University, Erzurum, Türkiye

Contact e-mail: *hamzapolat@atauni.edu.tr*

Citation

Polat, H., (2025). Use of Generative Artificial Intelligence in Developing Writing Skills in Language Education. In A. Kaban, & T. Demirel (Eds.), *Interdisciplinary Applications of Artificial Intelligence - 1* (pp. 161-184). ISTES.

Chapter 9- Artificial Intelligence in Education: Transformation, Applications, and Future Perspective

Bünyami Kayalı

Chapter Highlights

- It approaches artificial intelligence not merely as a technological tool, but as a pedagogical actor that transforms learning processes.
- It has examined artificial intelligence applications within the framework of Education 4.0 and digital transformation from a holistic perspective.
- It critically addresses the ethical, legal, and social issues associated with the use of artificial intelligence in the educational context.
- It presents future-oriented policy recommendations for the sustainable, human-centred, and responsible adoption of artificial intelligence in the context of education.

Introduction

Artificial intelligence (AI) has emerged as a significant force in today's world, focusing on problem-solving and symbolic reasoning since its inception. In the 21st century, advancements in machine learning and big data have led to significant paradigm shifts in various areas, including education (Funa & Gabay, 2024). In this regard, AI, in the context of education, is positioned as a strategic element that fundamentally transforms learning, teaching, and institutional structures, beyond being merely a technological tool (Singh et al., 2025). This section examines the rise of AI in education within the framework of historical context, digital transformation, and the Education 4.0 paradigm. Furthermore, the innovative applications it brings to educational activities, its existing potential, and its transformative effects are discussed.

The Rising Role of AI in Education

The use of AI in education has gained significant momentum in recent years. As AI's role in education becomes increasingly central, applications such as learning analytics, adaptive learning platforms, and automated assessment systems are emerging. This is because AI supports learner-centred approaches and strengthens data-driven decision-making processes. Furthermore, its ability to monitor student performance in real-time enables the personalization of learning processes, transforming traditional education models by increasing the scalability of educational services (Shaikh & Kıranlı Güngör, 2025). Furthermore, AI is emerging as an assistant that supports and strengthens teaching processes not only for learners but also for teachers. AI tools reduce teachers' workload both inside and outside the classroom, enabling them to focus more on their core pedagogical roles (Fabiana & Fabian, 2025). In these respects, AI is considered a comprehensive transformation tool that reshapes both learning processes and teaching practices in education.

Digital Transformation and Education 4.0 Context

Digital transformation is creating structural impacts on education systems, directing educational institutions towards increasing their innovation capacity and utilising the full potential of digital technologies. Along with this transformation, the Education 4.0 approach, which expresses the fundamental change that digitalisation has brought about in education systems, has come to the fore. Emerging as an extension of Industry 4.0, Education 4.0 is characterised by elements such as flexible learning environments, lifelong learning, and a competency-based approach to education (Kamarudin, 2024). The key to digital transformation in education lies in innovative pedagogical approaches that incorporate components such as intelligent teaching systems and data-driven decision-making processes (Garzón et al., 2025). In this context, digital technologies, particularly AI, play a central role in their potential to transform learning and teaching practices. AI is considered a strategic tool that enables more inclusive, effective, and personalised learning experiences by overcoming the limitations of traditional education models (McCarthy et al., 2023).

Research Problem and Purpose of the Section

The increasing impact of AI applications in education necessitates that teachers and educators develop new pedagogical, digital, and ethical competencies. However, conceptual ambiguities, pedagogical alignment issues, and ethical debates regarding the use of AI in education persist in the literature (Zhu et al., 2025). In particular, technical issues such as algorithmic bias, data privacy, and 'black box' algorithms, as well as the tension between the multidimensional nature of educational contexts and the requirements for technical standardisation, are among the fundamental problems that must be considered in the integration of AI into education (Chen et al., 2020).

In this context, the aim of this section is to conduct an in-depth review of the existing literature on the integration of AI applications into education and their pedagogical effectiveness; to analyse the transformation taking place in education from a multidimensional, critical, and holistic perspective; and to identify research directions for the potential future of this transformation. Thus, the aim is to both contribute to the theoretical gaps in the literature for researchers and provide practical guidance for practitioners on the problems encountered in integrating AI into educational processes (Córdova-Esparza, 2025).

Scope and Structure of the Section

This chapter comprehensively addresses the theoretical foundations of AI in education, its role in pedagogical transformation within the context of Education 4.0, and its impact on learning and teaching processes. The chapter examines the rising role of AI in education and the dynamics of digital transformation, followed by an evaluation of its primary application areas and pedagogical effects. Critical issues, including ethics, data privacy, and transparency, are also addressed. Finally, future perspectives offered by AI for education systems, along with research and policy recommendations, are presented. This structure aims to provide the reader with the opportunity to understand the multidimensional effects of AI in education within a clear and systematic framework.

AI and Education: Conceptual and Theoretical Foundations

This section emphasises that the transfer of AI to the field of education is not merely a technical process; rather, it is a multidimensional phenomenon grounded in cognitive, pedagogical, and epistemological foundations (Ouyang & Jiao, 2021). AI, in the context of education, goes beyond being a mere technological tool; it reconstructs learning and teaching processes as an analogy of human intelligence. It is considered a phenomenon with profound theoretical implications (Rismanchian & Doroudi, 2025). In this regard, the theoretical framework of the relationship between AI and education should be positioned within the axes of cognitive processes, learning theories, and the philosophy of knowledge. This section aims to establish this theoretical framework and create a solid conceptual foundation for subsequent sections.

AI Concepts and Their Intersection with Education

AI is defined as the ability of computer systems to perform complex cognitive tasks, such as reasoning, learning, problem-solving, and decision-making, which are typically associated with human intelligence (Langley, 2019). In the context of education, these fundamental AI capabilities are directly linked to learning and teaching processes. For example, AI's reasoning ability is reflected in intelligent teaching systems that support students in making logical inferences and help them solve complex problems step-by-step (Bellas et al., 2022). Similarly, the concept of learning from an AI perspective refers to algorithms improving their performance by extracting patterns from data, while in education, it manifests itself through the delivery of personalised and adaptive learning activities (Ciaschi & Barone, 2025). Furthermore, decision-making and autonomy capabilities gain an educational function through AI's ability to generate predictions about student performance, provide feedback, and adapt pedagogical strategies (Gennari et al., 2023).

Within this framework, AI plays three fundamental roles in the context of education: as a tool, a teaching environment, and a pedagogical actor. As a tool, AI supports teachers in administrative processes and offers functions such as automated assessment and personalised feedback. In its role as a learning environment, it provides interactive and dynamic learning experiences, particularly through adaptive platforms (Létourneau et al., 2025). As a pedagogical actor, AI applications directly influence learning processes by offering individual guidance to students through intelligent teaching systems and chatbots. In this context, it should be emphasised that the transfer of AI to the field of education is not merely a technical issue; it is a multi-layered phenomenon based on cognitive, pedagogical, and epistemological foundations (Rismanchian & Doroudi, 2025).

The Relationship Between Cognitive Science, Learning Science, and AI

Cognitive science and learning sciences provide conceptual and theoretical frameworks that explain the mental processes of human learning. These two concepts provide a robust theoretical foundation for designing AI systems and their practical application in educational environments. Information processing theory approaches the human mind as a system that acquires, processes, stores, and retrieves information. This approach aligns with AI's ability to process perceptual inputs, creating internal cognitive representations and making decisions based on these representations (Hameed, 2020). Similarly, constructivist learning theory argues that learners actively construct knowledge rather than passively receiving it. AI-supported learning environments similarly support these knowledge construction processes through interactive scenarios (Voskoglou, 2023). The sociocultural approach emphasises that learning occurs within social interactions and cultural contexts. AI-based collaborative learning environments and learning analytics applications enhance the educational applications of this theoretical framework by facilitating increased peer interaction (Balzan et al., 2025). In this regard, cognitive science theories inform the pedagogical design of AI-supported teaching systems, contributing to the alignment of teaching technologies with the principles of learning psychology (Gibson, 2023).

However, these theoretical frameworks also require a critical assessment of the relationship between AI systems and human cognition. Although AI systems can model certain aspects of human learning, they lack the complex

characteristics of human cognition, such as flexibility, creativity, emotional depth, and contextual sensitivity (Deng et al., 2025). Indeed, current AI models cannot fully replicate some of the dynamic and multi-layered mechanisms of the human brain (Lake & Baroni, 2023). In this context, a theoretical approach that considers the limits and complementarity of the relationship between AI and human cognition is critical for positioning AI's role in education in a more realistic and pedagogically meaningful way.

AI Models Used in Education

AI models used in education are generally categorised into three main types: symbolic AI, machine learning (ML) AI, and generative AI (GenAI). This classification is important for understanding how AI is positioned in education and which pedagogical requirements it addresses. Each approach is based on different theoretical foundations and has unique contributions and limitations in terms of learning and teaching processes.

Symbolic AI encompasses rule-based systems where knowledge is represented through explicit rules, symbols, and logical inference mechanisms. Early intelligent tutoring systems (ITS) used this approach to guide students' problem-solving steps according to predefined rules (Broisin, 2020). The strongest aspects of symbolic AI are transparency, explainability, and traceability of decision processes. These features support conceptual understanding and the "why" question from a pedagogical perspective. However, its limitations in dealing with complex and ambiguous data, the need for a manual knowledge base, and low adaptability are fundamental limitations of this approach in educational settings.

Machine learning is a data-driven approach that improves the performance of algorithms by training them on data (Mallik & Gangopadhyay, 2023). ML is widely used in education in areas such as predicting student performance, identifying learning difficulties, creating adaptive learning environments, and learning analytics applications (Song, 2024). The black box problem in decision-making processes, the risk of algorithmic bias, and the need for high-quality and large-volume data are significant pedagogical limitations (Wazan et al., 2023).

GenAI encompasses models that can generate new content such as text, images, sound, or code. It has rapidly spread in the field of education, particularly with the emergence of large language models such as ChatGPT (Kohnke & Foung, 2024). GenAI applications have significant potential in areas such as the production of teaching materials, automated feedback, scenario-based learning design, and support for the assessment process. Content accuracy, ethical risks, disinformation, and potential adverse effects on students' critical thinking skills are among the fundamental pedagogical limitations (Yan et al., 2024).

The Historical Transformation of Educational Technologies and the Position of AI

The historical transformation of educational technologies is a process that began with computer-assisted instruction. Over time, this process has evolved towards e-learning, mobile learning, intelligent teaching systems, and adaptive learning environments (Willet & Na, 2024). In the early days, tools such as slide projectors, radio,

and television were used in education, promising innovation. In later periods, computers, internet-based platforms, and mobile devices were integrated into education (Gràcia & Gil, 2021). These technologies contributed to the diversification of teaching methods by facilitating access to information and knowledge. However, in the educational context, they have generally been positioned as tools that support content delivery and interaction (Almulla & Ali, 2024). In a later stage, intelligent tutoring systems (ITS) and adaptive platforms changed the direction of educational technologies by offering structures capable of limited adaptations based on student performance.

Within this historical trajectory, AI represents a qualitative turning point in educational technologies. While previous technologies primarily focused on facilitating knowledge transfer, AI has created a fundamental shift through its dimensions of pedagogical interaction, adaptability, and data-driven decision-making (Bauer et al., 2025). AI-supported adaptive learning systems can dynamically adapt content and feedback based on students' behavioural patterns. This forms the basis of personalised learning experiences (Castro et al., 2025). Today, learning analytics, adaptive platforms, and customised AI tools have become central to educational technologies. Data-driven teaching processes have become widespread. Various systematic reviews have demonstrated that these systems optimise learning pathways, increase student engagement, and enhance the quality of teaching outcomes (Wang et al., 2024).

The position of AI in educational technologies is regarded not merely as a technological advancement, but as a broader paradigm shift that requires a rethinking of pedagogical and organizational structures (Wang et al., 2024). In this context, AI is positioned not merely as a tool that automates existing processes but as an element of transformation that redefines the nature of learning and teaching processes. This section aims to provide a theoretical framework for why and how AI plays a transformative role in education, thereby establishing a conceptual bridge for subsequent sections.

The Transformative Effects of AI in Education

The transformation brought about by AI in education represents a paradigm shift that goes beyond a simple digitalization process, profoundly affecting the fundamental components and structural dimensions of education (Kabanda, 2025). This process not only challenges traditional educational models but also redefines approaches to learning and assessment (Zhang & Fenton, 2024). In this context, AI has brought about a multi-layered transformation in educational processes and structures. In this respect, AI stands out as a fundamental paradigm that reshapes not only how education is delivered but also how it is conceptualized.

The Transformation of Learning Processes

Artificial intelligence-supported learning environments go far beyond traditional "one-size-fits-all" learning models. They offer flexible and dynamic environments that can be adapted to students' individual learning styles and paces (Kestin et al., 2025). This transformation essentially eliminates limitations such as temporal, spatial,

and content sequencing constraints on learning. It reconfigures learning processes across cognitive, affective, and metacognitive dimensions (Ren & Wu, 2025). AI enhances the learning experience by making it more interactive and goal-oriented through personalized learning, adaptive learning, and instant feedback mechanisms (Wang et al., 2024). These systems have the potential to enhance learners' capacity to manage their own learning processes and their motivation to learn. Therefore, they are considered not merely as assistive technologies in learning environments but as an integral and effective part of the learning experience (Bauer et al., 2025).

Personalised Learning

Personalised learning aims to provide learning experiences tailored to students' knowledge levels, learning speeds, and individual characteristics, based on a learner-centred and constructivist approach (Gupta, 2024). Artificial intelligence-based systems support this process by creating learning activities and content using student data and learning analytics (Merino Campos, 2025). This approach has the potential to increase learning efficiency and individual success, particularly in classrooms where students with varying academic achievement levels are together (Dumbuya, 2024). Despite this potential, ethical and structural risks, such as the digital divide, algorithmic bias, and data privacy concerns, are frequently raised in relation to AI-supported personalization (Rasheed et al., 2025). Therefore, the personalised learning process should be transparent and ethically designed fairly and inclusively.

Adaptive Learning

Adaptive learning systems are artificial intelligence-based platforms that dynamically organise learning activities and content based on students' prior knowledge, learning speeds, and immediate performance (Gupta, 2024). These systems aim to create the most suitable learning plan for each individual by continuously collecting data from student interactions. Thus, the goal is to go beyond a one-size-fits-all teaching approach and increase learning efficiency (Ginting et al., 2024; Hariyanto, 2025). AI-supported adaptive learning can strengthen students' perception of success and learning motivation by presenting content at the right time and at an appropriate level of difficulty (Lan & Zhou, 2025). Furthermore, by providing greater control over learning processes, it supports learner autonomy. Nevertheless, overly personalised learning activities carry the risk of limiting students' discovery, critical thinking, and self-regulation skills (Verghis et al., 2025). Therefore, in the design of adaptive learning systems, a balanced pedagogical approach should be adopted between dimensions such as cognitive challenge and learner autonomy.

Feedback

Immediate feedback is considered a fundamental element that enhances the effectiveness of learning, particularly in behavioral and cognitive learning theories. Constructivist approaches, on the other hand, emphasize that feedback should be supportive in nature, helping students to construct their own understanding. Artificial intelligence-supported systems can provide instant, formative, and personalized feedback by analyzing

student interactions and performance data in real time (Durak & Onan, 2025; Salameh, 2025). The instant feedback loop enables students to quickly identify their mistakes, monitor their learning processes, and make the necessary corrections without delay. This positively supports motivation and academic achievement (Wang et al., 2024). However, the quality and pedagogical meaning of automated feedback must be carefully considered. Feedback that is context-free, superficial, or contains algorithmic biases can weaken rather than support learning. It may also lead to a reduction in student–teacher interaction (Zhu et al., 2025). Therefore, an effective automated feedback system should not be limited to providing right or wrong information. It should take on a function that encourages learning by providing insights into the student's thought process.

The Transformation of Teaching Processes

AI is transforming teaching processes, restructuring the teacher's role from the traditional position of "knowledge transmitter" to that of learning designer, guide, and facilitator (Ley et al., 2025). This transformation is strengthened by AI reducing teachers' routine workload, supporting personalised learning, and making instructional decision-making processes data-driven (Ren & Wu, 2025). AI-supported applications contribute to the implementation of pedagogical approaches in a more strategic, flexible, and student-centred manner through the innovative solutions they offer in course design, content production, and assessment (Amado-Salvatierra et al., 2023). In this respect, AI is positioned as a fundamental, transformative element that reshapes teaching practices, rather than merely serving as a supportive tool in the teaching process.

AI-Supported Course Design

AI significantly impacts instructional design by determining learning objectives, content sequencing, and selecting teaching strategies, thereby making instructional design more data-driven and adaptable. AI-supported tools can analyse student performance and interaction data to suggest learning paths, content adjustments, and teaching strategies tailored to different student profiles, enabling courses to be structured in a more targeted and dynamic manner (Choi et al., 2024). This approach enables teachers to base their teaching decisions on analytical insights rather than intuitive preferences, thereby strengthening the process of achieving teaching objectives. However, there is a risk that AI-generated recommendations may be disconnected from the pedagogical context or remain superficial, which could weaken learner-centredness and design intent (Amado-Salvatierra et al., 2023). Therefore, in AI-supported course design, AI should be positioned not as an element that replaces the instructional designer's pedagogical expertise, but as a tool that supports and enriches it.

Content Production and Material Development

GenAI can rapidly and scalably generate learning materials, including text, visuals, audio, and interactive simulations, with the potential to enhance learning outcomes. In this respect, AI is transforming content production and material development processes (Bozkurt et al., 2024). This allows teachers to reduce their workload in this area and focus more on their fundamental pedagogical roles, such as guidance, feedback, and enriching the learning experience. However, the accuracy, timeliness, cultural sensitivity, and pedagogical appropriateness of

materials produced by AI require critical evaluation (Nguyen, 2025). The teacher's pedagogical control and ethical responsibility are decisive at this point. Reviewing these materials in line with expertise and adapting them to the context will eliminate the risk of inaccurate or biased content negatively affecting learning processes (Nadim & Fuccio, 2025). Therefore, GenAI should be positioned as an assistant that supports teachers in creating content and developing materials.

AI in Measurement and Evaluation Systems

AI-supported measurement and assessment approaches go beyond traditional assessment models, offering more efficient, personalised, and process-oriented assessment opportunities. Formative, competency-based, and adaptive assessment models are becoming more feasible thanks to AI algorithms (Waladi & Sefian, 2023). Student performance can be continuously monitored, and instant feedback can be provided (Durak & Onan, 2025). Automatic scoring and AI-based assessment tools accelerate the process, particularly in large-scale and open-ended assessments, thereby reducing teachers' administrative burden (Johnson & Zhang, 2024). They also enable students to track their own progress more effectively (Wang et al., 2024).

However, AI-supported assessment systems require critical evaluation in terms of fairness and trust. Algorithmic biases and a lack of transparency regarding how the systems operate create significant limitations in the assessment processes. Furthermore, the limited ability to assess higher-level cognitive skills such as creativity and contextual thinking carries the risk of undermining the reliability of the results obtained (Simbeck, 2023). Therefore, explainability, human oversight, and adherence to ethical principles are fundamental requirements in AI-supported measurement and assessment applications. The final assessment must be carried out in conjunction with human expertise (Li, 2025).

Corporate and Systemic Transformation

AI is transforming the management, planning, and quality assurance processes of educational institutions, enabling them to achieve a more efficient and data-driven structure (Siminto et al., 2023). This transformation also encompasses institutional-level decision-making mechanisms and the shaping of educational policies (Garzón et al., 2025). The increasing role of data in institutional decision-making processes stands out as one of the key factors determining the macro-level effects of AI on education systems (Ifenthaler & Yau, 2020).

Learning Analytics

Learning analytics is a field that analyzes large datasets obtained from learners and learning environments to provide insights into learning processes and support informed decision-making (Ifenthaler & Yau, 2020). AI-supported analytical systems create early warning mechanisms by extracting meaningful patterns from student performance data, enabling the timely identification of at-risk students (Herodotou et al., 2020; Fahd & Miah, 2023). This approach enables the development of data-driven strategies aimed at improving student success (Zhu

et al., 2025). However, it is essential to note that ethical issues, such as data privacy and algorithmic bias, must be carefully managed (Garzón et al., 2025).

AI-Based Learning Management Systems

AI-based learning management systems (LMS) optimise educational institutions' administrative and academic processes through automation and data analytics (Sain et al., 2024). Automating functions such as student enrolment, course planning, and performance tracking increases institutional efficiency. Furthermore, the learning experience is supported through pedagogically personalised content, reminders, and progress reports (Wang et al., 2024). The data-driven recommendations provided by these systems contribute to enhancing teaching processes and improving student success. Furthermore, maintaining a balance between automation and human oversight is crucial for ensuring transparent and reliable institutional decision-making (Sain et al., 2024).

Digitalisation and Its Impact on Education Policies

The digitisation of education policies highlights the impact of AI on fundamental policy areas such as ethical and legal oversight, accessibility, and equality (Garzón et al., 2025). AI is transforming education policies in areas such as the restructuring of curricula at national and institutional levels, the development of teachers' digital competencies, and the design of new learning environments (Schiff, 2021). However, the cost and infrastructure requirements of AI-based solutions carry the risk of deepening the digital divide (Dlamini & Ndzinisa, 2025). Algorithmic biases, data privacy, and transparency issues necessitate an ethical management approach (Balan, 2024). Therefore, developing comprehensive policy frameworks that ensure the fair, ethical, and sustainable use of AI in education is emerging as a fundamental priority at both global and local levels.

This section has critically examined the transformative effects of AI in education at the policy, governance, and system levels. The following section will then examine the concrete applications of AI in education in detail, illustrating how it is used in learning, teaching, and assessment processes through specific examples.

AI Applications in Education

AI has now moved beyond being an abstract technology in the field of education and has become an integral component of learning environments, assessment processes, and discipline-specific teaching practices. In line with the theoretical foundations discussed in previous sections and the transformative effects of AI in the educational context, this section addresses the practical applications of AI in education. These applications encompass a wide range, from designing learning environments to measuring and assessing processes, and from discipline-specific teaching approaches to interactive learning experiences (Wang et al., 2024). Within this framework, the pedagogical contributions, structural limitations, and ethical risks associated with AI-based educational applications will be examined through an analytical and critical approach.

AI-Supported Learning Environments

AI-based learning environments provide students with dynamic and adaptable learning experiences that surpass traditional approaches in terms of learner interaction, personalization, and learning continuity. Learning environments have become more student-centered with the use of AI. Within the scope of personalised learning, they enable active participation and the delivery of tailored course content (Garzón et al., 2025). Furthermore, instant feedback mechanisms and learning analytics-based monitoring tools contribute to the continuous support and enrichment of learning processes.

Intelligent Teaching Systems

Intelligent Teaching Systems (ITS) are among the most established and widely used applications of AI-based learning environments. They differ from traditional teaching in terms of learner interaction, personalisation, and learning continuity (Mousavinasab et al., 2018). They aim to provide student-centred and dynamic learning experiences through learning activities and content delivery tailored to individual learning needs, along with instant feedback mechanisms (Zerkouk et al., 2025).

Findings in the literature indicate that AÖSs support deep learning and academic achievement by optimising cognitive load through these capabilities (Steenbergen-Hu & Cooper, 2013). However, maintaining a balance between the teacher's role and system autonomy is critical for the effective integration of AÖS. Excessive system autonomy can lead to a weakening of pedagogical control. Furthermore, data privacy and algorithmic biases underscore the need for ethical and responsible approaches in AI design (Zerkouk et al., 2025).

Chatbots and Conversation-Based Learning

AI-powered chatbots are tools that simulate human-like conversations thanks to their natural language processing capabilities. These systems, which can provide learners with question-and-answer, guidance, and interactive support, are playing an increasingly widespread role in education (Neji et al., 2023). These tools answer student questions, facilitate access to course materials, and provide homework reminders. They support the learning process, particularly in language education, by providing safe and non-judgmental conversation practice (Labadze & Grigolia, 2023). They can also tailor learning experiences to individual needs (Pratiwi et al., 2024). In this respect, chatbots have the potential to enhance interactive learning while reducing the routine support burden on teachers (Garzón et al., 2025).

However, there are some limitations and risks associated with the use of chatbots in education. The presentation of misleading or outdated information, interaction remaining at a superficial level, and students' tendency to accept system outputs without questioning them can negatively affect the quality of learning (Polverini & Gregorcic, 2024; Yuen & Schlote, 2024). Furthermore, data privacy, ethical responsibility, and algorithmic biases are among the factors that directly affect the reliability of chatbots (Yuan, 2025). Therefore, it is important to observe

pedagogical principles that support accuracy, transparency, and critical thinking in the design and use of chatbots in education.

GenAI for Course Material Production

GenAI tools enable the rapid production of teaching materials such as text, images, audio, and video, particularly through large language models. In this respect, they significantly accelerate educators' content development processes. GenAI enables the efficient creation of personalised content, summaries, quizzes, and scenario-based learning materials in line with learning objectives and student profiles (Yaacoub et al., 2025). In this way, GenAI can reduce the time and workload teachers devote to content production, allowing them to focus more on their core pedagogical roles.

However, the pedagogical quality and ethical dimensions of materials generated by GenAI require critical evaluation. The accuracy, cultural appropriateness, and pedagogical depth of content may be insufficient without teacher guidance. Copyright infringements, the reflection of algorithmic biases in content, and the spread of misinformation are other significant risks (Nguyen, 2025). Furthermore, the sustainable and scalable implementation of AI-supported material production depends on the continuous monitoring of ethical principles and data privacy standards. In this context, GenAI should be considered not as a substitute for teachers but as an assistant that enhances their pedagogical expertise and supports creativity.

AI-Supported Assessment

AI both automates measurement and assessment processes and enables more detailed and process-oriented analyses (Zhao, 2024). AI-supported assessment approaches facilitate formative and competency-based assessment. Tracking students' progress in a more systematic manner enables instant feedback and enhances the quality of assessment processes.

Automatic Marking

AI-based automatic marking systems have the ability to evaluate student responses quickly and consistently. Thanks to this ability, they are widely used, particularly for multiple-choice, short-answer, and structured response types. These systems automate and standardise assessment processes in large-scale examinations, while providing instant feedback and detailed analyses, thereby increasing assessment efficiency and reducing the workload on teachers (Rotou & Rupp, 2020; Zhao, 2024).

However, automated grading applications have significant limitations in terms of validity, reliability, and pedagogical fairness. In particular, when evaluating open-ended responses that require creativity, contextual thinking, and higher-order cognitive skills, it is difficult for algorithms to match the level of interpretation achieved by human evaluators. Furthermore, algorithmic biases can produce unfair results for students from different linguistic and cultural backgrounds (Xu et al., 2021). Therefore, hybrid assessment approaches that

incorporate transparency, continuous validation, and human oversight are essential in the design and use of automated scoring systems.

Rubric-Based and Process-Focused Assessment

AI-supported rubric-based assessment approaches enable the analysis of developmental processes in student performance. Through structured rubrics aligned with specific learning outcomes and success criteria, AI can systematically evaluate student work and provide consistent and detailed feedback. This increases assessment transparency while enabling students to see their strengths and areas for improvement more clearly (Xie et al., 2024a). Particularly in skill-based areas such as language learning, AI-supported rubrics make the learning process more measurable and traceable, encouraging self-regulated learning (Köylü, 2025).

In the context of process-oriented assessment, AI can track not only students' final products but also their problem-solving steps, collaboration processes, and the thinking strategies they use. This tracking contributes to more comprehensive analyses of the learning process and strengthens the feedback loop. It also enables the identification of errors made by students throughout the process and the provision of immediate corrective feedback (Guo et al., 2024). In this respect, AI takes on a function that supports the continuity of learning and transforms assessment into a pedagogical learning tool.

Competency-Based Assessment

AI-supported competency-based assessment approaches aim to evaluate students' proficiency in specific knowledge, skills, and attitudes more accurately and holistically through data-driven analyses. AI algorithms that analyse performance data obtained from students' different learning activities and assessment processes provide scoring and feedback based on competency levels by tracking cognitive processes. This significantly contributes to the identification of individual learning activities and the creation of personalised development plans (Zhao, 2024).

This approach enables the assessment of not only students' knowledge levels but also 21st-century skills such as problem-solving, critical thinking, collaboration, and creativity. Furthermore, AI-supported assessment systems can offer scalable and sustainable solutions by accelerating and standardising assessment processes in large student groups. Monitoring algorithmic biases, regularly updating systems, and observing ethical principles are critical to the reliability and long-term applicability of these applications.

Discipline-Specific Applications

The applications of AI in education exhibit significant diversity in line with the unique pedagogical needs and learning objectives of different disciplines. AI technologies adapted to discipline-specific learning requirements

support discipline-specific teaching practices, unlike general teaching approaches. They also make their impact on learning processes more tangible and visible (Memari & Ruggles, 2025).

AI in STEM Education

AI applications in STEM (Science, Technology, Engineering, and Mathematics) education make learning processes more effective and student-centred through intelligent teaching systems, adaptive learning environments, and automated assessment tools. Systematic reviews reveal that AI increases learning motivation in STEM fields and has meaningful and positive effects on academic performance (Memari & Ruggles, 2025).

The structure of STEM education, which focuses on fundamental skills such as problem solving, analytical thinking, experimentation, and data analysis, is supported by AI-assisted simulations, virtual laboratories, and data analytics tools, enabling more in-depth learning experiences. It offers safe and interactive experimental environments, particularly in fields such as physics and chemistry. Data analytics tools help students develop their ability to interpret complex data sets and discover patterns, contributing to the concretisation of abstract STEM concepts (Deckker & Sumanasekara, 2025; León et al., 2025).

Language Education with AI

AI has become an important tool in language learning, supporting the development of students' listening, speaking, reading, and writing skills through personalised feedback and automated assessment capabilities. AI-based applications accelerate the learning process by analysing students' language production in real time and offer teaching experiences that are sensitive to individual learning needs (Zhao, 2024; Xu & Ouyang, 2022).

The multidimensional nature of language acquisition, encompassing pronunciation, grammar, vocabulary, and cultural understanding, highlights the importance of AI-supported adaptive learning platforms, virtual conversation partners, and automated feedback systems. These systems enhance language fluency by offering opportunities for real-time speech practice, pronunciation analysis-based feedback, and the detection of grammatical errors (Xodabande et al., 2025). They also boost students' confidence, thereby reducing their anxiety about practising (Ding & Yusof, 2025).

AI in Healthcare and Engineering Simulations

AI has become a key component in healthcare and engineering education, enriching practical learning through simulations and virtual learning environments. Particularly in these high-risk fields requiring advanced practical skills, AI-supported simulations enable students to experience applications such as surgical procedures, complex engineering problems, and emergency scenarios in safe and controlled environments by providing realistic simulations (Garzón et al., 2025).

Such simulations offer pedagogical contributions aimed at developing students' decision-making, problem-solving, and application skills. AI's performance analysis and learning analytics capabilities enable the provision of personalised feedback and the systematic monitoring of skill development. In this respect, AI-supported simulations stand out as powerful teaching tools that support applied learning, offer error-free learning opportunities, and increase the effectiveness of learning processes.

Art and Design Education with AI

Art and design education is a multidimensional structure encompassing creativity, aesthetic decision-making, concept development, and production processes. In this context, GenAI stands out as a tool that supports creative production by offering students new opportunities in the creation of artworks and design prototypes, idea development, and collaborative production processes (Wang et al., 2024). AI-supported production tools contribute to students experimenting with different styles and aesthetic approaches, accelerating creative processes, and discovering new visual possibilities. However, ethical and philosophical debates concerning originality, copyright, and the artist's creative role are criticisms levelled at these technologies. Therefore, AI should be positioned as a tool that supports creativity, with the final aesthetic decision-making process remaining human.

The sustainability and scalability of discipline-specific productive AI applications depend on their optimisation in line with the pedagogical and technical requirements of the field. This process necessitates continuous collaboration between discipline experts and AI developers, the integration of up-to-date content, and the development of flexible systems capable of adapting to the changing needs of the arts and design fields. The development of cost-effective solutions, particularly in resource-intensive simulation and virtual production environments, is critical for the widespread adoption and long-term viability of these applications.

AI in Interactive Learning Environments

AI plays a central role in the development of interactive learning environments, not limited to content production or assessment processes in education. In the context of interaction, experience, and multi-sensory learning, AI-supported environments offer students more engaging and personalised learning experiences. They increase learning motivation and participation levels in class, thereby enhancing the quality and effectiveness of the learning process.

Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR)-Based Learning

AR, VR, and MR-based learning environments offer students realistic, immersive, and interactive experiences thanks to AI integration. These technologies enable the concretisation of complex and abstract concepts, the safe experience of hazardous or inaccessible environments, and practical training in real-world scenarios (Garzón et al., 2025). They increase the effectiveness of the learning process (Dhaas, 2024).

AI can adapt the learning experience by analysing students' interactions, movements, and responses in AR/VR/MR environments in real time. This supports deeper cognitive understanding and increased learning motivation and curiosity from an affective perspective (Asoodar et al., 2024). Despite all these contributions, potential negative effects such as high costs, advanced technical infrastructure requirements, and technical problems are significant limitations to the widespread use of these technologies.

Educational Games and Gamification

AI-supported educational games and gamification applications enhance student motivation and learning continuity by making learning processes more engaging, interactive, and trackable. AI can dynamically adapt the difficulty level of games based on student performance, provide personalised feedback, and generate new game scenarios or tasks aligned with learning objectives (Gao, 2024). In this respect, gamification stands out as a pedagogical tool that makes learning more lasting and meaningful.

However, the excessive use of gamification components such as badges, leaderboards, and competitive elements carries the risk of developing dependence on external motivation and weakening the intrinsic desire to learn in students (Limonova et al., 2024). Furthermore, despite AI's potential to optimise gamification elements according to individual student profiles, game designs must be aligned with pedagogical goals and avoid purely entertainment-focused approaches (Wadhwa et al., 2024).

AI in Multimedia (Video, Audio, Graphics) Content

Multimedia learning is an approach based on presenting information through different sensory channels such as video, audio, graphics, and text (Twabu, 2025). AI enriches teaching materials with its capacity to produce, personalise, and optimise multimedia content in areas such as video analysis, speech recognition, and graphic design (Garzón et al., 2025). It makes the learning experience more accessible and effective with functions such as automatic summarisation, key concept extraction, multilingual subtitling, and translation (Navarrete et al., 2025). This integration contributes to presenting complex information in a more understandable and memorable way while catering to different learning styles.

However, the pedagogical effectiveness of AI-supported multimedia content depends on continuously improving it by analysing students' interactions with the content and designing it in line with fundamental teaching principles such as cognitive load theory (Saxena et al., 2023). While sustainability can be supported by reusing and enriching existing content, the long-term viability of these systems depends on the continuity of the technological infrastructure, educators' AI literacy, and careful management of ethical and pedagogical boundaries.

The AI applications examined in this section demonstrate meaningful transformations in learning and teaching processes across different layers of education. While these applications have the potential to enrich the learner experience, they require the holistic management of pedagogical consistency, ethical responsibilities, and practical limitations. Sustainability depends on the technological infrastructure and educators' AI literacy and institutional

adaptation capacity, while scalability is directly related to the development of cost-effective solutions and systematic dissemination strategies.

Ethical, Legal and Social Dimensions

The increasing use of AI in education raises fundamental issues such as ethical responsibility, legal accountability, and social justice. While the AI applications discussed in previous sections demonstrate meaningful transformations in learning and teaching processes, these transformations necessitate comprehensive management within the framework of pedagogical consistency, ethical responsibilities, and practical limitations. In this context, the integration of AI into education systems requires the adoption of a human, rights and responsibility-centred approach through issues such as data privacy, algorithmic biases, trust, copyright and liability. Sustainability and scalability are directly linked to technological infrastructure, educators' AI literacy and institutional adaptation capacity.

Data Privacy, Student Data and Confidentiality

AI-supported education systems track student performance and learning preferences to enable learning analytics and personalised teaching applications. In some cases, they even collect and process large-scale and sensitive personal data, including biometric data (Huang, 2023). This necessitates strict adherence to fundamental privacy principles such as data minimisation, explicit consent, purpose limitation, and data security. It requires transparent definition of who processes the data, for what purpose, and for how long (Hariharan, 2025). The literature emphasises that privacy violations arising in data collection, storage, and sharing processes, along with the "black box" nature of AI systems, pose significant risks in terms of ethical issues (Guan et al., 2023).

In this context, the responsibilities of students, teachers, educational institutions, and technology providers must be clearly and fairly defined. Educational institutions must implement robust data protection measures such as encryption, secure storage, and controlled data transmission. It is fundamental that students and parents are informed about data applications and are given control over their personal data (Tang, 2024). Furthermore, organisations developing AI solutions must adopt design principles that comply with ethical and legal regulations and act in accordance with human rights and privacy standards as stipulated by legislation such as the European General Data Protection Regulation (Banu, 2024). Only then can the social legitimacy and sustainability of AI use in education be ensured (Huang, 2023).

Bias, Fairness and Transparency

AI systems can produce unfair and exclusionary outcomes for certain student groups due to biased modelling processes based on limited representative data sets (Baker & Hawn, 2021). This reinforces existing inequalities in education and undermines fairness in assessment and evaluation (Simbeck, 2023). It also carries the risk of negatively affecting inclusivity through "allocative" or "representational" harms (Eden et al., 2024).

In this context, transparency and explainability form the basis of fair and reliable AI systems. Making AI algorithms understandable to teachers and students in terms of what data they use, what criteria they apply, and how they make decisions is critical for establishing pedagogical trust, reducing algorithmic bias, and ensuring accountability (Shafik, 2024).

Trust in AI Systems in Education

Trust emerges as a fundamental determinant for the social acceptance and sustainable use of AI systems in education. Research shows that trust in AI systems is directly related to qualities such as explainability, accountability, accuracy, and consistency (Lim et al., 2023). Explainability enables users to understand why AI produces certain outputs (Nguyen, 2025). Accountability reduces the "black box" perception by providing clarity on who is responsible, thereby strengthening pedagogical trust (Shafik, 2024).

However, the risks of excessive automation and "algorithmic authority" necessitate a critical assessment of trust in the educational context. The delegation of decision-making processes to automated systems can lead to accountability issues and tendencies towards "overconfidence," particularly when users lack sufficient knowledge about how the system operates (Selwyn et al., 2021). Therefore, AI systems must be designed and implemented with an approach that centres human oversight and balances AI autonomy with pedagogical control (Karran et al., 2025).

Copyright and Content Ownership

The ownership of content generated by GenAI in text, visual, and other media formats raises new pedagogical and legal uncertainties in education. The proliferation of large language models has led to a fundamental paradigm shift in academic writing, plagiarism, and intellectual property (Hutson, 2024). The question of how works generated by AI should be assessed in terms of human authorship and copyright remains a matter of debate (Bozkurt, 2023). This situation blurs the boundaries between teacher effort, student production and AI contribution, necessitating a re-examination of the concepts of authorship, originality and legal responsibility.

In particular, the assessment of work produced by students using AI-assisted tools has brought academic integrity and plagiarism debates to the forefront. Presenting AI outputs as student work without scrutiny risks undermining the learning process. The use of GenAI as a "supportive tool" in creative processes brings to the fore the redefinition of authorship within a collaborative framework (Creely & Blannin, 2023). In this context, it is critically important for educational institutions to develop clear policies on the acceptable use of AI, provide guidance on proper attribution and AI ethics, and adopt a holistic approach that protects both student rights and copyright.

Responsibility and Regulations in AI Use

Responsibility in the use of AI in education has a multi-layered structure shared among teachers, institutions, developers, and policymakers throughout the life cycle of AI systems. All of these actors must assume ethical and legal obligations based on protecting human rights and human dignity. Clear frameworks of responsibility should be established between developers, implementers, and regulators (Nguyen et al., 2023). In this context, it is emphasised that education stakeholders are directly responsible for the use of AI in education and that this responsibility is non-transferable.

AI regulations should aim to provide ethical assurance without restricting pedagogical innovation. They should also focus on scaling the benefits of AI in a safe, responsible and sustainable manner. AI policies in education should consider equality, accountability, the right to education and democratic values, in addition to data protection (Ellis, 2024). International human rights-based frameworks should guide this process (Rotenberg, 2025). Institutional ethical guidelines and policies should establish clear standards for the use of AI in educational institutions. Regulations must develop balanced policy approaches between innovation and risks.

Future Perspectives

The impact of AI on education points to a long-term transformation process that reshapes learning models, pedagogical approaches and education policies, differing from current practices. The future of AI in education paves the way for new pedagogical paradigms in line with trends such as automation, hybrid learning models, multimodal approaches, and human-AI collaboration (Vorobyeva et al., 2025). This section aims to address this transformation in light of evidence-based research findings and pedagogical theoretical frameworks.

Autonomous and Hybrid Learning Models

With the integration of AI into education systems, autonomous and hybrid learning models are expected to become increasingly widespread. Autonomous learning enables students to independently manage their own learning processes through AI-supported systems, while hybrid learning models offer flexible structures that combine the strengths of face-to-face and online environments (Wang & Li, 2024). AI-based personalisation, adaptive guidance, and intelligent teaching systems have the potential to enhance learner autonomy by optimising students' learning speeds and processes (Gotavade, 2024).

However, realising this potential in a pedagogically meaningful way depends on maintaining a balance between AI autonomy and teacher guidance. Research shows that teachers' roles in guidance, motivation, facilitation, and emotional support play a complementary function in hybrid models and that the human–AI co-learning approach strengthens this balance (Karran et al., 2024). Considering the risk that excessive automation may undermine student self-efficacy and critical thinking, it is considered essential that AI systems be developed with human-centred designs that support human autonomy and ensure equal learning opportunities (Karran et al., 2025).

Multimodal AI-Supported Learning Ecosystems

Future learning ecosystems will be supported by multimodal AI systems capable of integrating different data types such as text, sound, image, movement, and context. These systems have the potential to make learning experiences more holistic, interactive, and context-sensitive by responding to individual differences in how students process information. Advanced multimodal language models enable personalised learning experiences by integrating natural language processing with audiovisual content production (Zerkouk et al., 2025). This approach has positive effects on learning motivation, participation, and performance (Heilala et al., 2024).

However, multimodal AI-supported learning ecosystems also bring new design responsibilities in terms of accessibility, inclusivity, and cognitive load. Adaptations such as annotated content and subtitle support for visually and hearing-impaired students increase the inclusivity of education (Elgohary & Zhu-Tien, 2025). However, the simultaneous use of multiple input channels carries the risk of increasing cognitive load (Kamal & Cooper-Stachowsky, 2024). Therefore, adhering to the principles of cognitive load theory in the design of multimodal AI systems, avoiding excessive information presentation, and adopting inclusive pedagogical approaches are fundamental requirements (Wells, 2025).

Human–AI Co-Creation

AI will go beyond being merely an automation tool in future educational environments, taking on a co-creative role alongside teachers and students. In this approach, AI is positioned as a creative partner in processes such as content design, problem-solving scenarios, and learning material production. By combining teachers' pedagogical expertise with AI's big data processing and learner profile analysis capabilities, it contributes to the emergence of richer and more effective learning experiences (Zhong & Cochrane, 2024). Research shows that such human–AI co-creativity models have positive effects on students' cognitive development, problem-solving, and critical thinking skills (Jose et al., 2025).

However, the limits and ethical dimensions of human–AI collaboration must be carefully considered, as much as its pedagogical value. Issues of originality, creative responsibility, and ethical ownership are critical in co-creation processes. Risks such as plagiarism, copyright infringement, and the excessive externalisation of creativity are present in AI-generated content (Zerkouk et al., 2025). Therefore, while harnessing the creative potential of AI, it is necessary to develop human-centred approaches that focus on human creativity, critical thinking, and pedagogical guidance (Jose et al., 2025).

Super Personalisation and New Pedagogical Approaches

In the future of education, the approach of super personalisation, as the highest level of individualised learning, will become increasingly decisive. Super personalisation refers to a framework that holistically analyses students' prior knowledge, learning objectives, cognitive states, and motivation levels through AI-supported micro-adaptations and real-time data analytics. This framework aims to dynamically optimise learning content and

learning activities based on the multidimensional learner data obtained (Merino Campos, 2025). This approach promotes the development of new pedagogical paradigms that support student agency and autonomous learning by placing individual learning profiles at the centre of the learning ecosystem (Aggrey, 2018).

However, the effects of hyper-personalisation on learning equity and pedagogical justice must be critically evaluated. Despite the potential to offer learning opportunities tailored to each student, algorithmic biases, data privacy risks, and digital inequalities may limit the fair and inclusive implementation of these approaches (Roshanaei et al., 2023). Furthermore, the lack of transparency in micro-adaptation processes carries the risk of weakening students' control over their learning processes. Therefore, it is essential that super-personalisation-based pedagogical approaches are designed in an integrated manner with ethical principles, data security, and fairness mechanisms.

AI Vision in Education Policies

The transformative impact of AI on education systems necessitates the development of a comprehensive vision that encompasses not only classroom applications but also national and international education policies. Future-oriented education policies should maximise the pedagogical potential offered by AI while adopting inclusive strategies to manage ethical, legal and societal risks. In this context, teacher training, AI literacy, digital competencies and pedagogical leadership occupy a central position among policy priorities (Mbambo & Plessis, 2024).

The success of education policies depends on sustainable practices supported by ethical usage guidelines and infrastructure investments. International frameworks emphasise that human-centred design, data privacy, algorithmic fairness, and equal access principles should form the basis of AI policies (Zerkouk et al., 2025). At the same time, providing a robust technological infrastructure and reducing digital inequalities are emerging as prerequisites for a fair and inclusive transformation to enable the effective use of AI in education (Khan et al., 2025).

The educational policy perspective discussed in this section demonstrates that the transformative power of AIin education can only be meaningful within a human-centred, ethical and sustainable vision. Robust policy frameworks, continuous adaptation and multi-stakeholder governance approaches will be decisive in positioning AI as a transformative tool that promotes the common good in education.

Conclusion and Recommendations

This section has comprehensively assessed the current state of AI in education and its future research and application perspectives. The existing literature in the field of AI in Education highlights the potential of AIto transform teaching processes and contribute to the sustainable development goals of education systems. It also shows that structural problems persisting in ethical, political, and social dimensions related to pedagogical

applications need to be addressed (Mustafa et al., 2024). This book chapter addresses the impact of AIin education through a multidimensional approach, ranging from theoretical foundations to pedagogical transformation, from application examples to ethical and legal dimensions, and future perspectives. It demonstrates that AI represents a new phase in human-technology interaction and that this transformation can only produce meaningful and lasting gains for the future of education if it is managed in accordance with human-centred, ethical, and sustainable principles.

General Assessment

AI offers significant contributions in areas such as personalising learning processes in education, adapting teaching materials, and optimising teaching processes (Wang et al., 2024). Systematic literature reviews and bibliometric studies reveal that the impact of AI is not limited to technological innovations; it is also directly related to goals of access equality, inclusivity, and sustainable education. In this context, there is a growing academic focus on AI's potential to support sustainable education practices, particularly in relation to the fourth Sustainable Development Goal (Nikolopoulou, 2025).

This chapter demonstrates that AI is not merely a supportive tool in education, but a transformative element grounded in cognitive, pedagogical, and epistemological foundations. The educational implications of artificial intelligence, the similarities and differences between human and machine learning, and the roles AI assumes in education as a "tool," "environment," and "pedagogical actor" have been addressed within a theoretical framework. This framework has made it possible to understand AI-supported learning environments in terms of pedagogical rationale. However, the literature emphasises that the pedagogical contributions of AI applications must be considered alongside their ethical, reliability, and accessibility dimensions (Mustafa et al., 2024). Issues such as data privacy, algorithmic bias, trust, copyright, and responsibility reveal that AI's potential in education can only become sustainable and meaningful when balanced with multidisciplinary, ethical, and pedagogical approaches (Wang et al., 2024).

The Sustainable Potential of AI in Education

The sustainable potential of AI in education extends beyond improving individual learning experiences to encompass the capacity of education systems to contribute to sustainable development goals. The literature reveals that AI has significant potential in terms of increasing inclusive and accessible educational opportunities, supporting sustainable development principles, and integrating education policies with sustainability strategies (Nikolopoulou, 2025). In this respect, AI is considered a tool that can strengthen the transformative role of education, particularly in the context of the Sustainable Development Goals.

However, it is emphasised that AI-based applications can also contribute to environmental and social sustainability through energy efficiency, resource optimisation, and the digitisation of educational materials. However, realising this potential depends on educational institutions making strong infrastructure investments, promoting digital

literacy, and developing inclusive policy frameworks. Otherwise, there is a risk that AI-supported transformation will deepen digital inequalities (Abulibdeh, 2025).

The sustainable use of AI in education is possible not only through technological capacity but also through a holistic approach to pedagogical, ethical, and institutional conditions. A human-centred design approach, ethical guidelines, the development of teacher competencies, and the strengthening of critical AI literacy stand out as key components of this process. In the long term, AI should be positioned not merely as a tool focused on improving learning outcomes, but as a pedagogical element that supports the development of higher-level skills such as critical thinking, problem solving, and creativity.
This approach will enable AI to become a sustainable tool for transformation that makes education more equitable, accessible, and high-quality.

Recommendations for Research and Application

The theoretical, pedagogical, and political discussions presented in this section of the book highlight the need for research and applications concerning the future of AI in education to be approached in a more systematic, ethical, and sustainable manner. In this context, the recommendations presented below aim to provide a guiding framework for both academic research and educational practice.

In terms of research, there is a need for longitudinal and interdisciplinary studies that holistically address the dimensions of sustainability, equity, and pedagogical impact in the field of AI in education. Research programmes that examine the relationship between AI and the Sustainable Development Goals and include pedagogical and system-level impact assessments should be prioritised (Nikolopoulou, 2025). In addition, it is important to increase evidence-based studies addressing the effects of AI applications on access equality, inclusiveness, and the digital divide in education (Abulibdeh, 2025). Future research should critically focus on theoretical and empirical approaches that address the relationship between AI and learning theories, cognitive load in human-AI interaction, and learner autonomy. Furthermore, it is necessary to examine the pedagogical and cognitive effects of super-personalisation and multimodal AI applications and to develop hybrid and innovative methodologies in this process (Wang et al., 2024).

From the perspective of education practitioners and policymakers, clear policy and guideline frameworks must be established to ensure that AI is adopted in education in an ethical, reliable, and sustainable manner. Continuous professional development programmes aimed at strengthening teachers' AI literacy and pedagogical use competencies are emerging as a key element in the effective implementation of AI-supported learning environments (Wang et al., 2024). Education policies should include regulations based on the principles of transparency, accountability, and sustainability, covering ethical, legal, and social dimensions. Data privacy, ethical guidelines, and human-centred design approaches should be explicitly structured (Mustafa et al., 2024). Furthermore, infrastructure investments aimed at reducing digital inequalities, along with interdisciplinary

collaborations between education scientists, IT specialists, and policymakers, will support the fair and effective use of AI in education.

As a general assessment, these recommendations highlight that the transformation of AI in education must be addressed not only through technological innovation but also through ethical sensitivity, pedagogical justification, and sustainability principles. This approach will contribute to positioning AIas a transformative tool that enhances human potential and prioritises social benefit.

Acknowledgements or Notes

Authors' Disclosures Language Translation and Localization: For the translation and localization of content, [DeepL and ChatGPT 5,2, December, 2025] was employed. Human translators subsequently reviewed and adjusted the translations to ensure accuracy, cultural appropriateness, and contextual relevance.

References

Abulibdeh, A. (2025). A systematic and bibliometric review of artificial intelligence in sustainable education: Current trends and future research directions. *Sustainable Futures*, *10*, 101033.

Aggrey, E., Bhagat, K. K., Chang, M., Chen, N. S., Chew, S., Doniyorbek, A., ... & Zheng, Y. (2018). Smart learning futures: a report from the 3rd US-China smart education conference. *Smart Learning Environments*, *5*(1).

Almulla, M., & Ali, S. I. (2024). The Changing Educational Landscape for Sustainable Online Experiences: Implications of ChatGPT in Arab Students' Learning Experience. *International Journal of Learning Teaching and Educational Research*, *23*(9), 285. https://doi.org/10.26803/ijlter.23.9.15

Asoodar, M., Janesarvatan, F., Yu, H., & Jong, N. de. (2024). Theoretical foundations and implications of augmented reality, virtual reality, and mixed reality for immersive learning in health professions education. *Advances in Simulation*, *9*(1), 36. https://doi.org/10.1186/s41077-024-00311-5

Baker, R. S., & Hawn, A. (2021). Algorithmic Bias in Education. *International Journal of Artificial Intelligence in Education*, *32*(4), 1052. https://doi.org/10.1007/s40593-021-00285-9

Balzan, F., Santos, P. P. de A., Gabbrielli, M., Albarracin, M., & Lopes, M. (2025). A Computational Model of Inclusive Pedagogy: From Understanding to Application. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2505.02853

Banu, J. T. (2024). AI in Learning: Designing the Future. *Technical Communication Quarterly*, 1. https://doi.org/10.1080/10572252.2024.2428400

Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K., & Sailer, M. (2025). Looking Beyond the Hype: Understanding the Effects of AI on Learning. *Educational Psychology Review*, *37*(2). https://doi.org/10.1007/s10648-025-10020-8

Bellas, F., Guerreiro-Santalla, S., Naya-Varela, M., & Duro, R. J. (2022). AI Curriculum for European High Schools: An Embedded Intelligence Approach. *International Journal of Artificial Intelligence in Education*, *33*(2), 399. https://doi.org/10.1007/s40593-022-00315-0

Bozkurt, A. (2023). Generative AI, Synthetic Contents, Open Educational Resources (OER), and Open Educational Practices (OEP): A New Front in the Openness Landscape. *Open Praxis*, *15*(3), 178. https://doi.org/10.55982/openpraxis.15.3.579

Broisin, J. (2020). Towards intelligent technology-enhanced learning solutions for transforming higher education: contributions and future directions. *HAL (Le Centre Pour La Communication Scientifique Directe)*. https://hal.science/tel-03684452

Castro, G. P. B., Chiappe, A., Soledad, M., & Nieblas, C. A. (2025). Key Barriers to Personalized Learning in Times of Artificial Intelligence: A Literature Review. *Applied Sciences*, *15*(6), 3103. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/app15063103

Chen, X., Xie, H., & Hwang, G. (2020). A multi-perspective study on Artificial Intelligence in Education: grants, conferences, journals, software tools, institutions, and researchers. *Computers and Education Artificial Intelligence*, *1*, 100005. https://doi.org/10.1016/j.caeai.2020.100005

Ciaschi, M., & Barone, M. (2025). Examining the Increasing Use of Artificial Intelligence in Education, A step Closer to Personalized Learning. *Annals of Computer Science and Information Systems*, *44*, 9. https://doi.org/10.15439/2025f7856

Córdova-Esparza, D. (2025). AI-Powered Educational Agents: Opportunities, Innovations, and Ethical Challenges. *Information*, *16*(6), 469. https://doi.org/10.3390/info16060469

Creely, E., & Blannin, J. (2023). The implications of generative AI for creative composition in higher education and initial teacher education. *ASCILITE Publications*, 357. https://doi.org/10.14742/apubs.2023.618

Deng, X., Zhang, Y., Guo, T., Zhang, Y., Kang, Z.-B., & Yang, H. (2025). ChallengeMe: An Adversarial Learning-enabled Text Summarization Framework. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2502.05084

Dhaas, A. (2024). Augmented Reality in Education: A Review of Learning Outcomes and Pedagogical Implications. *American Journal of Computing and Engineering*, *7*(3), 1. https://doi.org/10.47672/ajce.2028

Eden, C. A., Chisom, O. N., & Adeniyi, I. S. (2024). Integrating AI in education: Opportunities, challenges, and ethical considerations. *Magna Scientia Advanced Research and Reviews*, *10*(2), 6. https://doi.org/10.30574/msarr.2024.10.2.0039

Elgohary, O., & Zhu-Tien. (2025). Attention, Action, and Memory: How Multi-modal Interfaces and Cognitive Load Alter Information Retention. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2509.05898

Ellis, R. A. (2024). The Education Leadership Challenges for Universities in a Postdigital Age. *Postdigital Science and Education*. https://doi.org/10.1007/s42438-024-00461-9

Fabiana, M. G., & Fabian, G. G. (2025). Connecting education in the new emerging trends: Educators' artificial intelligence awareness and readiness for integration in teaching and learning. *International Journal of Research Publication and Reviews*, *6*(6), 2416. https://doi.org/10.55248/gengpi.6.0625.2058

Funa, A., & Gabay, R. A. E. (2024). Policy guidelines and recommendations on AI use in teaching and learning: A meta-synthesis study. *Social Sciences & Humanities Open*, *11*, 101221. https://doi.org/10.1016/j.ssaho.2024.101221

Gao, L. (2024). A Literature Review: Which, How and What for the Use of Artificial Intelligence in Gamification. *European Conference on Games Based Learning*, *18*(1), 281. https://doi.org/10.34190/ecgbl.18.1.2627

Garzón, J., Patiño, E., & Marulanda, C. (2025). Systematic review of artificial intelligence in education: Trends, benefits, and challenges. *Multimodal Technologies and Interaction*, *9*(8), 84.

Gennari, R., Melonio, A., Pellegrino, M. A., & D'Angelo, M. (2023). *How to Playfully Teach AI to Young Learners: a Systematic Literature Review*. 1. https://doi.org/10.1145/3605390.3605393

Gibson, D. (2023). Learning theories for artificial intelligence promoting learning. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13341

Gotavade, T. S. (2024). Artificial Intelligence Ecosystem for Automating Self-Directed Teaching. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2411.07300

Gràcia, X. G., & Gil, J. M. S. (2021). Artificial Intelligence in Education. *Seminar Net*, *17*(2). https://doi.org/10.7577/seminar.4281

Guan, X., Feng, X., & Islam, A. Y. M. A. (2023). The dilemma and countermeasures of educational data ethics in the age of intelligence. *Humanities and Social Sciences Communications*, *10*(1). https://doi.org/10.1057/s41599-023-01633-x

Gupta, T. R. (2024). Adaptive Learning Systems: Harnessing AI to Personalize Educational Outcomes. *International Journal for Research in Applied Science and Engineering Technology*, *12*(11), 458. https://doi.org/10.22214/ijraset.2024.65088

Hameed, H. A. (2020). Artificial Intelligence: What it Was, and What it Should Be? *International Journal of Advanced Computer Science and Applications*, *11*(6). https://doi.org/10.14569/ijacsa.2020.0110609

Hariharan, A. (2025). Artificial Intelligence for Optimal Learning: A Comparative Approach towards AI-Enhanced Learning Environments. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2510.11755

Heilala, V., Araya, R., & Hämäläinen, R. (2025, March). Beyond text-to-text: An overview of multimodal and generative artificial intelligence for education using topic modeling. In *Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing* (pp. 54-63).

Huang, L. (2023). Ethics of Artificial Intelligence in Education: Student Privacy and Data Protection. *Science Insights Education Frontiers*, *16*(2), 2577. https://doi.org/10.15354/sief.23.re202

Hutson, J. (2024). Rethinking Plagiarism in the Era of Generative AI. *Journal of Intelligent Communication*, *4*(1). https://doi.org/10.54963/jic.v4i1.220

Jose, B., Cherian, J., Verghis, A. M., Varghise, S. M., Mumthas, S., & Joseph, S. (2025). The cognitive paradox of AI in education: between enhancement and erosion. *Frontiers in Psychology*, *16*. https://doi.org/10.3389/fpsyg.2025.1550621

Kamal, Z. H., & Cooper-Stachowsky, M. (2024, December 20). First Impressions on Supporting Students with a Course-Aligned, AI-powered Virtual Teaching Assistant (TA), Oliver. *Proceedings of the Canadian Engineering Education Association (CEEA)*. https://doi.org/10.24908/pceea.2024.18548

Kamarudin, S. (2024). Strategic 4.0 in Academia: A Comprehensive Review of Digital Transformation and Future Gaps Agenda [Review of *Strategic 4.0 in Academia: A Comprehensive Review of Digital Transformation and Future Gaps Agenda*]. *Studies in Systems, Decision and Control*, 31. Springer International Publishing. https://doi.org/10.1007/978-3-031-62102-4_3

Karran, A. J., Charland, P., Martineau, J.-L., Guinea, A. O. de, Lesage, AM., Sénécal, S., & Leger, P.-M. (2024). Multi-stakeholder Perspective on Responsible Artificial Intelligence and Acceptability in Education. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2402.15027

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, *15*(1), 17458. https://doi.org/10.1038/s41598-025-97652-6

Khan, S., Mazhar, T., Shahzad, T., Khan, M. A., Rehman, A. U., Saeed, M. M., & Hamam, H. (2025). Harnessing AI for sustainable higher education: ethical considerations, operational efficiency, and future directions. *Discover Sustainability*, *6*(1). https://doi.org/10.1007/s43621-025-00809-6

Kohnke, L., & Foung, D. (2024). Deconstructing the Normalization of Data Colonialism in Educational Technology. *Education Sciences*, *14*(1), 57. https://doi.org/10.3390/educsci14010057

Labadze, L., & Grigolia, M. (2023). Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00426-1

Lake, B. M., & Baroni, M. (2023). Human-like systematic generalization through a meta-learning neural network. *Nature*, *623*(7985), 115. https://doi.org/10.1038/s41586-023-06668-3

Lan, M., & Zhou, X. (2025). A qualitative systematic review on AI empowered self-regulated learning in higher education. *Npj Science of Learning*, *10*(1), 21. https://doi.org/10.1038/s41539-025-00319-0

Langley, P. (2019). An Integrative Framework for Artificial Intelligence Education. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(1), 9670. https://doi.org/10.1609/aaai.v33i01.33019670

Létourneau, A., Deslandes-Martineau, M., Charland, P., Karran, A.-J., Boasen, J., & Léger, P. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education [Review of *A systematic review of AI-driven intelligent tutoring systems (ITS) in K-12 education*]. *Npj Science of Learning*, *10*(1), 29. Nature Portfolio. https://doi.org/10.1038/s41539-025-00320-7

Limonova, V., Santos, A., Mamede, H. S., & Filipe, V. (2024). Maximising Attendance in Higher Education: How AI and Gamification Strategies Can Boost Student Engagement and Participation. In *Lecture notes in networks and systems* (p. 64). Springer International Publishing. https://doi.org/10.1007/978-3-031-60224-5_7

Mallik, S., & Gangopadhyay, A. (2023). Proactive and reactive engagement of artificial intelligence methods for education: a review [Review of *Proactive and reactive engagement of artificial intelligence methods for education: a review*]. *Frontiers in Artificial Intelligence*, *6*. Frontiers Media. https://doi.org/10.3389/frai.2023.1151391

Mbambo, G. P., & Plessis, D. (2024). Impact of Artificial Intelligence on Teacher Training in Open Distance and Electronic Learning. *International Journal of Learning Teaching and Educational Research*, *23*(5), 370. https://doi.org/10.26803/ijlter.23.5.19

McCarthy, A., Maor, D., McConney, A., & Cavanaugh, C. (2023). Digital transformation in education: Critical components for leaders of system change. *Social Sciences & Humanities Open*, *8*(1), 100479. https://doi.org/10.1016/j.ssaho.2023.100479

Merino-Campos, C. (2025). The impact of artificial intelligence on personalized learning in higher education: A systematic review. *Trends in Higher Education*, *4*(2), 17.

Mustafa, M. Y., Tlili, A., Lampropoulos, G., Huang, R., Jandrić, P., Zhao, J., ... & Saqr, M. (2024). A systematic review of literature reviews on artificial intelligence in education (AIED): a roadmap to a future research agenda. *Smart Learning Environments*, *11*(1), 59.

Navarrete, E., Nehring, A., Schanze, S., Ewerth, R., & Hoppe, A. (2025). A Closer Look into Recent Video-based Learning Research: A Comprehensive Review of Video Characteristics, Tools, Technologies, and Learning Effectiveness [Review of *A Closer Look into Recent Video-based Learning Research: A Comprehensive Review of Video Characteristics, Tools, Technologies, and Learning Effectiveness*]. *International Journal of Artificial Intelligence in Education*, *35*(4), 1631. Springer Science+Business Media. https://doi.org/10.1007/s40593-025-00481-x

Neji, W., Boughattas, N., & Ziadi, F. (2023). Exploring New AI-Based Technologies to Enhance Students' Motivation. *Issues in Informing Science and Information Technology*, *20*, 95. https://doi.org/10.28945/5149

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. P. T. (2023). Ethical principles for artificial intelligence in education. *Education and information technologies*, *28*(4), 4221-4241.

Nguyen, V. (2025). The Use of Generative AI Tools in Higher Education: Ethical and Pedagogical Principles. *Journal of Academic Ethics*. https://doi.org/10.1007/s10805-025-09607-1

Nikolopoulou, K. (2025). Generative artificial intelligence and sustainable higher education: Mapping the potential. *Journal of Digital Educational Technology*, *5*(1), ep2506.

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education Artificial Intelligence*, *2*, 100020. https://doi.org/10.1016/j.caeai.2021.100020

Pratiwi, N., Efendy, A. G., Rini, H. C., & Ahmed, N. A. (2024). Speaking Practice using ChatGPT's Voice Conversation: A Review on Potentials and Concerns [Review of *Speaking Practice using ChatGPT's Voice Conversation: A Review on Potentials and Concerns*]. *Journal of Language Intelligence and Culture*, *6*(1), 59. https://doi.org/10.35719/jlic.v6i1.149

Rismanchian, S., & Doroudi, S. (2025). The Evolution of Research on AI and Education Across Four Decades: Insights from the AIxEd Framework. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-025-00483-9

Roshanaei, M., Olivares, H., & Lopez, R. R. (2023). Harnessing AI to Foster Equity in Education: Opportunities, Challenges, and Emerging Strategies. *Journal of Intelligent Learning Systems and Applications*, *15*(4), 123. https://doi.org/10.4236/jilsa.2023.154009

Rotenberg, M. (2025). Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (Council Eur.). *International Legal Materials*, *64*(3), 859-902.

Saxena, R., Narang, S., & Ahuja, H. (2023). Improving the Effectiveness of E-learning Videos by leveraging Eye-gaze Data. *Engineering Technology & Applied Science Research*, *13*(6), 12354. https://doi.org/10.48084/etasr.6368

Selwyn, N., Hillman, T., Rensfeldt, A. B., & Perrotta, C. (2021). Digital Technologies and the Automation of Education — Key Questions and Concerns. *Postdigital Science and Education*, *5*(1), 15. https://doi.org/10.1007/s42438-021-00263-3

Shafik, W. (2024). *Towards Trustworthy and Explainable AI Educational Systems* (p. 17). https://doi.org/10.1007/978-3-031-72410-7_2

Shaikh, G., & Kıranlı Güngör, S. (2025). Artificial Intelligence in Education: Insights from a Bibliometric Study (2010–2025) Based on Scopus and Web of Science. *International Journal of Modern Education Studies*, *9*(2).

Simbeck, K. (2023). They shall be fair, transparent, and robust: auditing learning analytics systems. *AI and Ethics*, *4*(2), 555. https://doi.org/10.1007/s43681-023-00292-7

Singh, A., Kiriti, M. K., Singh, H., & Shrivastava, A. (2025). Education AI: exploring the impact of artificial intelligence on education in the digital age. *International Journal of Systems Assurance Engineering and Management*, *16*(4), 1424. https://doi.org/10.1007/s13198-025-02755-y

Song, D. (2024). Artificial intelligence for human learning: A review of machine learning techniques used in education research and a suggestion of a learning design model [Review of *Artificial intelligence for human learning: A review of machine learning techniques used in education research and a suggestion of a learning design model*]. *American Journal of Education and Learning*, *9*(1), 1. https://doi.org/10.55284/ajel.v9i1.1024

Tang, K. H. D. (2024). Implications of Artificial Intelligence for Teaching and Learning. *Acta Pedagogia Asiana*, *3*(2), 65. https://doi.org/10.53623/apga.v3i2.404

Twabu, K. (2025). Enhancing the cognitive load theory and multimedia learning framework with AI insight. *Discover Education*, *4*(1). https://doi.org/10.1007/s44217-025-00592-6

Vorobyeva, K. I., Belous, S., Savchenko, N. V., Smirnova, L. M., Nikitina, S. A., & Zhdanov, S. P. (2025). *Personalized learning through AI: Pedagogical approaches and critical insights*. *Contemporary Educational Technology, 17*(2), ep574.

Voskoglou, M. Gr. (2023). Artificial Intelligence and Digital Technologies in the Future Education. *Qeios.* https://doi.org/10.32388/07ve29

Wadhwa, R., Rabby, F., Bansal, R., & Hundekari, S. (2024). Drivers and Impact of Artificial Intelligence on Student Engagement. In *Advances in computational intelligence and robotics book series* (p. 161). IGI Global. https://doi.org/10.4018/979-8-3693-4268-8.ch011

Wang, L., & Li, W. (2024). The Impact of AI Usage on University Students' Willingness for Autonomous Learning. *Behavioral Sciences*, *14*(10), 956. https://doi.org/10.3390/bs14100956

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, *252*, 124167.

Wells, M. B. (2025). Empowering non-verbal individuals through AI-driven symbolic text prediction: a metaliteracy approach to communication and inclusion. *Discover Education*, *4*(1). https://doi.org/10.1007/s44217-025-00809-8

Willet, K. B. S., & Na, H. (2024). Generative AI Generating Buzz: Volume, Engagement, and Content of Initial Reactions to ChatGPT in Discussions Across Education-Related Subreddits. *Online Learning*, *28*(2). https://doi.org/10.24059/olj.v28i2.4434

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, *8*(10), 1839. Nature Portfolio. https://doi.org/10.1038/s41562-024-02004-5

Zerkouk, M., Mihoubi, M., & Chikhaoui, B. (2025b). A Comprehensive Review of AI-based Intelligent Tutoring Systems: Applications and Challenges. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2507.18882

Zhong, J., & Cochrane, T. (2024). Heutagogy-based Human-AI Co-creation Practice. *ASCILITE Publications*, 351. https://doi.org/10.14742/apubs.2024.1083

Zhu, H., Sun, Y., & Yang, J. (2025). Towards responsible artificial intelligence in education: a systematic review on identifying and mitigating ethical risks. *Humanities and Social Sciences Communications, 12*(1). 1-14.

Author Information

Bünyami Kayalı

https://orcid.org/0000-0001-6419-9088

Bayburt University

Bayburt,

Turkey

Contact e-mail: *bunyamikayali@bayburt.edu.tr*

Citation

Kayalı, B. (2025). Artificial Intelligence in Education: Transformation, Applications, and Future Perspective. In A. Kaban, & T. Demirel (Eds.), *Interdisciplinary Applications of Artificial Intelligence - 1* (pp. 185-215). ISTES.

Chapter 10- Artificial Intelligence in Programming and Software Development

Muhammed Coşkun IRMAK

Chapter Highlights

- Artificial intelligence is integrated across the entire software development lifecycle, from requirements analysis to maintenance, enabling a holistic transformation of software engineering processes.
- Large language model–based programming assistants enhance developer productivity in code generation and testing while reinforcing the necessity of human oversight.
- AI-driven software maintenance and quality management approaches enable predictive handling of technical debt and sustainable improvement of code quality.
- Modern software development paradigms such as low-code/no-code platforms, DevOps, and AIOps are significantly strengthened through artificial intelligence–based automation and analytics.
- The adoption of AI in software development introduces new ethical, security, and legal challenges, highlighting the importance of responsible and explainable AI practices.
- Evidence from both academic research and industrial applications indicates a clear evolution toward semi-autonomous and autonomous software engineering systems.

Introduction

Software development has undergone a continuous process of evolution since the emergence of computer science. In its early stages, software systems exhibited relatively simple structures due to limited hardware resources and narrow application domains. Today, however, software plays a critical role in almost every field, including healthcare, education, finance, industry, defense, and social life. This expansion has significantly increased both the technical complexity and the societal impact of software systems. The growing complexity of software systems has brought requirements such as speed, accuracy, reliability, and sustainability to the forefront of software development processes. Although traditional software engineering approaches—such as waterfall, agile, and DevOps methodologies—have sought to address these requirements, they often prove insufficient in the face of increasing data volumes, rising user expectations, and rapidly changing requirements. Consequently, the need for automation and intelligent decision-support mechanisms has become increasingly evident.

For many years, automation in software development has been implemented through compilers, integrated development environments (IDEs), testing tools, and version control systems. However, these tools are largely based on rule-driven and deterministic structures. In the presence of uncertainties, complex dependencies, and unpredictable error patterns encountered during software development, such tools offer limited flexibility. Artificial intelligence (AI), particularly with advances in machine learning and deep learning, provides a powerful paradigm for overcoming these limitations. Today, AI is no longer merely a component of software systems; rather, it has become an active support mechanism throughout the entire software lifecycle, including requirements analysis, design, coding, testing, deployment, and maintenance. This development transforms not only the technical aspects of software development processes but also their cognitive dimensions.

In recent years, the emergence of generative AI systems based on large language models (LLMs) has marked a significant turning point in software development. The increasingly blurred boundaries between natural language and programming languages are fundamentally reshaping the nature of software development. Software requirements are no longer defined solely through technical documentation but can also be articulated through natural language expressions. While this shift enhances the accessibility of software development, it simultaneously introduces new responsibilities and areas of debate within software engineering.

AI-driven software development approaches have the potential to improve developer productivity, reduce error rates, and enhance overall software quality. Nevertheless, issues such as automated code generation, excessive reliance on AI, security vulnerabilities, and uncertainties related to licensing and intellectual property necessitate a cautious and responsible adoption of these technologies. Therefore, the role of AI in software development processes should be evaluated not only as a technical advancement but also in conjunction with its ethical, legal, and societal dimensions.

The purpose of this chapter is to examine the role of AI in programming and software development from a holistic perspective and to evaluate it in light of current academic literature and industrial practices. Throughout the

chapter, the contributions of AI to software development processes, its interaction with modern software paradigms, and potential future directions are discussed. In this way, the chapter aims to provide readers with a comprehensive and balanced framework for understanding the current state and future trajectory of AI-driven software development.

AI in Software Development Processes

Software development processes are traditionally addressed within the framework of a lifecycle consisting of requirements analysis, design, coding, testing, deployment, and maintenance (Fig. 1). For many years, these processes have been conducted based on human expertise and rule-based automation tools. However, the increasing scale of software systems, along with the adoption of microservice architectures, cloud-based infrastructures, and continuous delivery requirements, has led software development processes to become significantly more complex and dynamic.

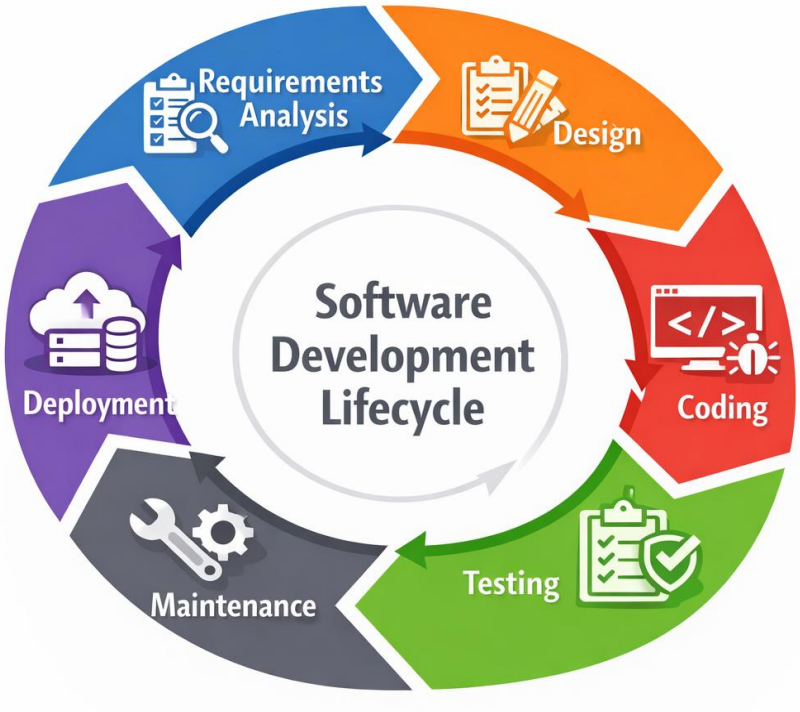


Figure 1. Software Development Lifecycle

This transformation has increased the need for data-driven, learning, and adaptive systems within software development processes. AI has emerged as one of the key technologies addressing this need and has begun to be integrated into nearly every stage of the software development lifecycle. In the literature, this approach is commonly referred to as “AI-assisted” or “AI-augmented software engineering” (Storey et al., 2020; Amershi et al., 2019a). AI-driven software development does not conceptualize the software lifecycle as a sequence of individually automated steps; rather, it treats it as an interconnected and continuously learning system. In this context, AI plays a significant role not only during the coding phase but also at much earlier stages of the development process. During the requirements analysis phase, natural language processing (NLP) techniques and LLMs are employed to automatically analyze, classify, and prioritize user requirements. User stories, requirement

documents, and feedback data can be processed by AI systems to identify potential inconsistencies and omissions (Zhang et al., 2023; Arora et al., 2024). This capability contributes to mitigating common challenges in software projects, such as misunderstandings and scope creep. At the design stage, AI can provide architectural recommendations based on patterns learned from past projects. Machine learning–based models analyze component dependencies, design patterns, and performance risks to offer decision support to developers (Amershi et al., 2019). This approach enables software architectures to be shaped through data-driven analyses rather than relying solely on intuitive or experience-based decisions.

AI Support in Coding Processes

The coding phase represents the stage of software development where AI applications are most visibly manifested. LLMs are capable of generating executable code from natural language inputs, analyzing existing codebases, and providing context-aware code completion suggestions. In particular, the emergence of Transformer-based models has established a strong bridge between programming languages and natural language (Chen et al., 2021; Li et al., 2022). AI-powered programming assistants automate repetitive coding tasks, allowing developers to focus on more complex problem-solving activities. Empirical studies have demonstrated that such tools increase developer productivity and significantly reduce coding time for specific tasks (Kim et al., 2024). Nevertheless, the literature emphasizes that human oversight remains indispensable with regard to the correctness and security of the generated code.

AI in Testing, Bug Detection, and Quality Assurance

Software testing processes constitute one of the most time- and cost-intensive phases of the software development lifecycle. AI offers effective solutions in areas such as the automatic generation of test cases, bug prediction, and the optimization of regression testing. Deep learning–based models can analyze historical defect reports and code changes to proactively identify code regions with a high likelihood of faults (Gao et al., 2019). These approaches enable testing activities to be conducted in a more targeted and efficient manner while contributing to the sustainable improvement of software quality. The literature highlights that AI-driven testing approaches provide significant advantages, particularly in large-scale and continuously evolving software systems (Tufano et al., 2024).

Process-Oriented Integration of AI

The true strength of AI in software development processes emerges when it is employed not at isolated stages but through a holistic, process-oriented approach. Leveraging AI across all phases—from requirements engineering to maintenance—enables software development processes to become more flexible, predictable, and adaptive. This approach transforms software development from a static production activity into a continuously learning and self-improving system. In the literature, this transformation is regarded as a fundamental paradigm shift for the future of software engineering (Storey et al., 2020; Suri et al., 2023).

AI in Software Maintenance and Quality Management

An examination of the software lifecycle reveals that maintenance activities account for the largest share of total cost and labor effort. Studies in the literature indicate that approximately 60–80% of the total cost of software projects arises from maintenance processes (Bennett & Rajlich, 2000; Sommerville, 2016). This highlights that software maintenance is not merely a secondary activity but a strategic process situated at the core of software engineering. Software maintenance serves multiple purposes, including defect correction, adaptation of systems to new requirements, enhancement of performance and reliability, and prevention of potential future issues. However, the scale, architectural complexity, and continuously evolving nature of modern software systems have made maintenance activities increasingly challenging and costly. In this context, AI offers significant potential to transform software maintenance into a more predictable, systematic, and proactive process.

Traditional software maintenance approaches largely rely on developer expertise and manual analyses. Code reviews, static analysis tools, and a limited set of quality metrics typically form the basis of maintenance-related decisions. Nevertheless, such approaches often struggle to adequately capture the semantic structure of software systems and their evolution over time. In contrast, AI-based approaches conceptualize software not as a static artifact but as a system that evolves dynamically. Machine learning models can analyze historical code changes, defect reports, and performance metrics to predict maintenance requirements in advance. This paradigm enables the adoption of proactive and predictive strategies in software maintenance, rather than reactive interventions (Tsoukalas et al., 2021).

Technical Debt, Code Quality, and AI

The concept of technical debt refers to the negative impact that short-term solutions may have on software quality and sustainability in the long term. The accumulation of technical debt leads to increased complexity in software systems, higher defect rates, and reduced development velocity (Cunningham, 1992). AI is increasingly utilized in the measurement and management of technical debt. In particular, supervised learning and regression-based models are capable of predicting the likelihood that specific code components will cause maintenance-related issues in the future. The literature reports that machine learning–based technical debt prediction models offer higher accuracy compared to traditional metric-based approaches (Tsoukalas et al., 2021; Digkas et al., 2021).

From a code quality perspective, AI-based systems focus not only on surface-level metrics such as lines of code or cyclomatic complexity but also on the semantic structure and contextual usage of code. This enables a more holistic assessment of software quality.

Code Smells and Automated Refactoring

Code smells are structural issues that do not directly cause faults but hinder the maintainability of software systems. Common examples of code smells include long methods, highly coupled classes, and duplicated code blocks (Beck, 1999). In recent years, deep learning and graph-based approaches have been widely adopted for the

automatic detection of code smells. Abstract syntax trees (ASTs) and graph neural networks (GNNs) enable the modeling of structural relationships within source code and facilitate the learning of complex patterns (Liu et al., 2019; Šikić et al., 2022).

Following the detection of code smells, AI-driven systems can provide automated or semi-automated refactoring recommendations. Search-based software engineering (SBSE) and reinforcement learning approaches aim to apply refactoring steps with minimal risk while preserving software functionality (Mansoor et al., 2017; Naik et al., 2024).

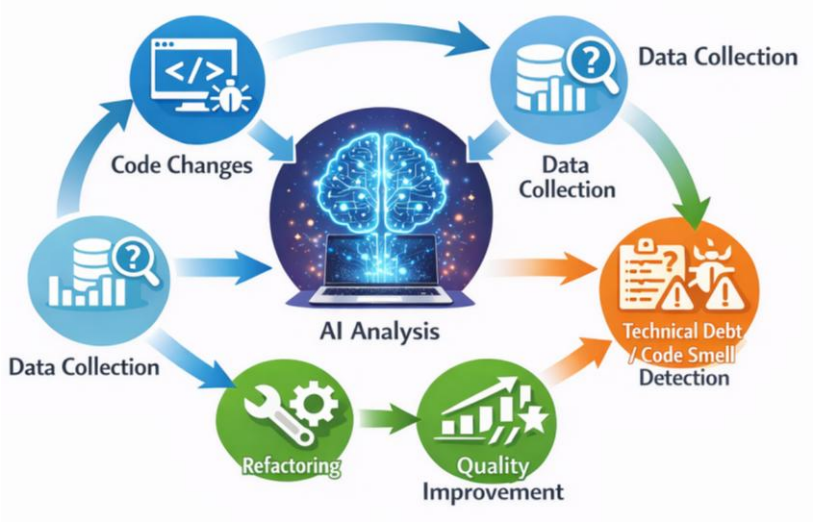


Figure 2. AI-Powered Software Maintenance Process

Predictive Maintenance and Anomaly Detection

One of the most significant contributions of AI to software maintenance lies in the application of predictive maintenance approaches. Log data, performance metrics, and user behavior can be analyzed by machine learning models to anticipate potential system failures in advance. Deep learning–based time-series models and anomaly detection algorithms enable the early identification of abnormal system behavior, thereby providing opportunities for timely intervention. These approaches are particularly critical for preventing service disruptions in large-scale and continuously operating software systems (Amgothu & Kankanala, 2024; Goswami et al., 2022).

AI-Driven Modern Software Development Approaches

In recent years, software development processes have undergone a significant transformation driven by cloud computing, microservice architectures, and continuous delivery requirements. This transformation has necessitated faster development cycles, more frequent updates, and high scalability in software systems. Traditional software development approaches often prove inadequate in managing the complexity encountered in such dynamic environments. At this point, AI has become a central component of modern software paradigms, offering intelligent automation capabilities across both development and operational processes (Fig. 3).

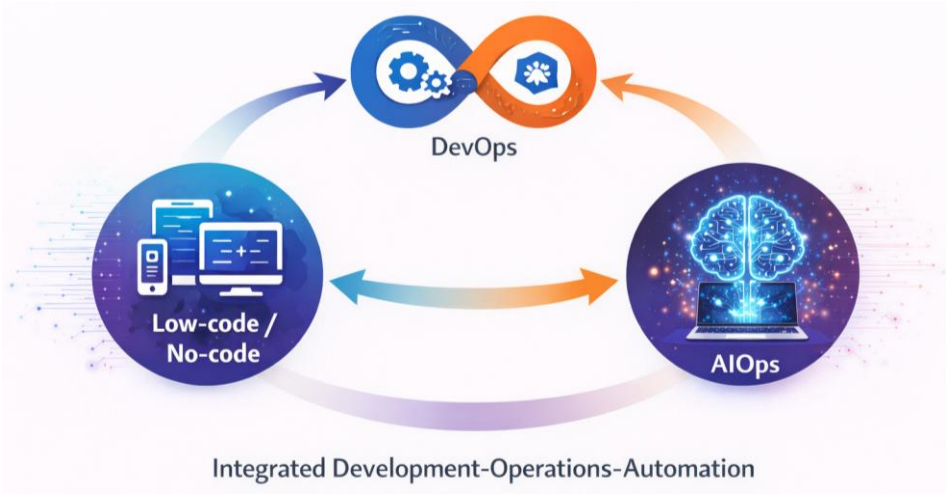


Figure 3. AI-Powered Modern Software Approaches

Within the scope of modern software development approaches, concepts such as low-code/no-code platforms, DevOps, and AIOps—when combined with AI support—enable the software development lifecycle to become more agile, predictable, and sustainable. In the literature, this integrated approach is regarded as a new paradigm in software engineering, in which the boundaries between development and operations are increasingly blurred (Fitzgerald & Stol, 2017; Mittal, 2025).

Low-code and no-code platforms enhance the accessibility of software development processes by enabling non-technical users to develop applications. These platforms provide features such as visual modeling tools, pre-built components, and automatic code generation. With the integration of AI, low-code and no-code platforms have evolved beyond mere visual programming environments into intelligent software development platforms. AI plays an active role in transforming requirements expressed in natural language into application components within low-code/no-code environments. LLMs can analyze user inputs and recommend appropriate data structures, workflows, and user interface designs. The literature reports that this approach supports the democratization of software development and significantly reduces prototyping time (Sahay et al., 2020; Binzer & Winkler, 2022). Nevertheless, it is also emphasized that low-code and no-code platforms exhibit limitations in handling complex business logic, high-performance requirements, and security concerns. Therefore, AI-enabled low-code and no-code approaches are regarded not as replacements for professional software development processes, but rather as complementary tools that augment them.

The DevOps approach represents a software development culture aimed at enhancing collaboration between development and operations teams. Continuous integration (CI) and continuous deployment (CD) processes have become indispensable components of modern software projects. However, the complexity and data-intensive nature of these processes make manual management increasingly challenging. AI provides effective solutions in DevOps processes across areas such as log analysis, performance monitoring, root cause analysis of failures, and deployment risk prediction. Machine learning–based models can analyze historical deployment data to proactively identify releases with a high likelihood of failure. This capability enables more informed decision-making in software deployment processes (Amgothu & Kankanala 2024; Patchamatla & Owolabi, 2025).

AIOps (Artificial Intelligence for IT Operations) refers to an approach that integrates AI into IT operations processes. AIOps systems analyze large volumes of operational data to provide capabilities such as anomaly detection, event correlation, and automated remediation (self-healing) mechanisms. AIOps solutions based on deep learning and time-series analysis can detect abnormal system behavior at an early stage, thereby preventing service disruptions. The literature indicates that AIOps approaches significantly reduce operational overhead, particularly in cloud-based and microservice architectures (Dang et al., 2019; Nedelkoski et al., 2019). By transforming software systems from being merely monitored into systems that are self-monitoring and adaptive, AIOps increases the level of autonomy in software engineering. This shift reduces reliance on human intervention in software operations while offering substantial advantages in terms of system reliability and service continuity.

The common characteristic of low-code/no-code platforms, DevOps, and AIOps approaches lies in the integrated incorporation of AI into software development and operational processes. This integrated structure transforms software systems from static products into dynamic and learning ecosystems. AI-driven modern software development approaches enhance speed, quality, and flexibility within software development processes while simultaneously redefining the role of software engineers. By reducing manual effort, developers are able to focus on higher-level responsibilities such as system design, architectural decision-making, and ethical considerations. In the literature, this transformation is regarded as a critical stage in the evolution of software engineering (Storey et al., 2020; Suri et al., 2023).

Ethical, Security, and Legal Considerations

The integration of AI into programming and software development processes brings not only technical gains but also significant ethical, security, and legal challenges. In particular, automated code generation, decision-support systems, and autonomous software components necessitate a re-examination of concepts such as responsibility, accountability, and trust within software engineering. In this context, AI-driven software development should be regarded not merely as a technological advancement but also as a socio-technical transformation. The literature emphasizes that the widespread adoption of AI-based systems in software engineering processes must be addressed within an ethical and trustworthy framework (Floridi et al., 2018; Amershi et al., 2019a). Otherwise, short-term productivity gains may evolve into serious security and legal risks in the long term.

One of the most significant ethical issues associated with AI-driven software development tools concerns the question of responsibility for the generated code. When automatically produced code behaves incorrectly, introduces security vulnerabilities, or leads to unethical outcomes, it remains unclear whether responsibility lies with the developer, the AI system itself, or the platform provider. LLMs may propagate biases and errors present in their training datasets into the code generation process. In the literature, this phenomenon is described as the reflective nature of AI systems and is considered one of the primary sources of ethical risk (Bender et al., 2021). Consequently, the human-in-the-loop approach is widely regarded as an indispensable ethical principle. Transparency and explainability also occupy a central position in ethical evaluations. Developers' ability to

understand the rationale behind code generated or recommended by AI is critical both for building trust and for effective debugging and validation processes.

AI-driven software development tools provide significant contributions to secure software development processes; however, they also introduce new attack surfaces. During automated code generation, the production of code fragments that do not comply with security standards may lead to the inadvertent introduction of vulnerabilities into software systems. Pearce et al. (2022) demonstrated that code generated by LLMs may contain common security flaws, such as SQL injection and buffer overflow vulnerabilities. This finding highlights the substantial risks associated with using AI-assisted development tools without adequate security review and validation mechanisms. In addition, AI models themselves can become targets of malicious attacks. Threats such as data poisoning, reverse engineering, and membership inference attacks pose serious risks to the reliability and trustworthiness of AI systems employed in software development processes (Shokri et al., 2017). Consequently, secure software development principles must be extended to explicitly encompass AI–based systems.

The fact that AI-powered programming assistants are largely trained on open-source code repositories gives rise to significant uncertainties regarding copyright and license compliance. When automatically generated code exhibits a high degree of similarity to licensed or copyrighted code, legal issues may arise. Recent studies have reported that, in certain cases, code generated by AI can closely resemble—or even nearly replicate—code contained in the training datasets (Negri-Ribalta et al., 2024; German, 2024). This situation may lead developers to unintentionally violate software licenses. From a legal perspective, the ownership of code generated by AI remains an issue that has not yet been clearly defined. Furthermore, the processing of personal data within AI-driven software development workflows must be carefully considered in light of data protection regulations such as KVKK (Turkish Law No. 6698 on the Protection of Personal Data), GDPR, and HIPAA. In particular, the use of cloud-based AI services introduces additional risks related to the sharing of sensitive data with third parties.

Another important dimension of ethical and security discussions concerns developers' overreliance on AI. This phenomenon, referred to in the literature as "automation bias," describes the tendency of users to accept outputs generated by AI without critical evaluation (Parasuraman & Riley, 1997). When AI-driven software development tools produce incorrect or incomplete outputs, developers may fail to detect these errors, potentially leading to serious consequences. Therefore, it is essential to position AI systems not as replacements for developers, but as tools that support and augment human expertise.

In recent years, the concept of "responsible AI" (responsible AI) has gained increasing attention in the software engineering literature. This approach is grounded in core principles such as fairness, transparency, security, privacy, and accountability. Designing and deploying AI systems used in software development processes in accordance with these principles is essential for producing solutions that are sustainable from both technical and societal perspectives (Floridi et al., 2018; Jobin et al., 2019).

Academic and Industrial Applications

The maturity of AI-driven software development approaches can be assessed not only through theoretical models and conceptual frameworks but also through tangible outcomes derived from academic research and industrial applications. In this context, jointly considering applications originating from both academia and industry is of critical importance for demonstrating the effectiveness of AI in software engineering. While academic research enables the testing and validation of novel methods and algorithms under controlled environments, industrial applications reveal the scalability, reliability, and sustainability of these approaches under real-world conditions. The literature emphasizes that progress in AI-driven software development is largely fueled by the interaction between these two ecosystems (Storey et al., 2020).

AI-Driven Software Development in Academic Research

In the academic literature, the integration of AI into software engineering has been a subject of research for many years. While early studies primarily focused on rule-based systems and statistical models, recent research has increasingly emphasized approaches based on deep learning and LLMs. In particular, the emergence of the concept of "big code" has enabled large-scale source code repositories to be utilized as rich data sources for machine learning applications (Allamanis et al., 2018).

The success of LLMs in code generation and code analysis has been extensively investigated in academic studies. Chen et al. (2021) demonstrated that Transformer-based models are capable of generating executable code from natural language inputs and achieving performance comparable to that of human developers on specific tasks. Similarly, models such as CodeBERT, GraphCodeBERT, and related architectures have shown strong performance in tasks including code search, code summarization, and bug detection (Feng et al., 2020; Guo et al., 2020).

Academic research has also increasingly highlighted AI-driven applications related to software maintenance, such as automated test case generation, code smell detection, technical debt analysis, and automated refactoring. These studies demonstrate that the role of AI in software engineering extends well beyond code generation, encompassing the entire software development lifecycle.

AI-Driven Software Development Practices in Industry

In industrial software development environments, AI has become an integral part of daily development practices. Large technology companies, in particular, extensively employ AI-based tools to enhance productivity and reduce error rates throughout software development processes.

Tools such as GitHub Copilot, Amazon CodeWhisperer, and Google's AI-powered development solutions represent some of the most widely adopted examples in industry. These tools provide developers with context-

aware code suggestions, automate repetitive tasks, and significantly shorten development time. An empirical study conducted by Kim et al. (2024) reported that AI-assisted code completion tools substantially improve developer productivity.

In industry, AI is utilized not only during the coding phase but also across areas such as test automation, deployment pipelines, and operational monitoring. AIOps solutions enhance service continuity by proactively predicting performance degradations and system failures in large-scale software systems (Dang et al., 2019; Nedelkoski et al., 2019).

Academia–Industry Interaction and Case Studies

The interaction between academic research and industrial applications plays a critical role in the maturation of AI-driven software development. Methods developed in academic environments are tested in industry using real-world data and usage scenarios, while industrial needs generate new problem domains for academic research. For example, anomaly detection systems employed in large-scale microservice architectures have emerged through the adaptation of deep learning models originally developed in the academic literature to industrial settings. Similarly, automated testing and defect prediction systems, informed by academic research, have delivered substantial improvements in software quality assurance practices (Patchamatla & Owolabi, 2025). Such case analyses demonstrate that AI-driven software development has moved beyond an experimental approach and now offers scalable and reliable solutions at an industrial level.,

While academic research provides methodological innovations and theoretical depth in the field of AI-driven software development, industrial applications reveal the practical value and limitations of these approaches. Considering these two perspectives together enables a more comprehensive understanding of the true potential of AI in software engineering.

The literature anticipates that future advancements in software engineering will be largely shaped through collaborations between academia and industry (Storey et al., 2020; Suri et al., 2023). In this context, AI-driven software development practices should be evaluated by jointly considering their scientific contributions and practical implications.

Future Perspectives and Conclusion

The integration of AI into programming and software development processes is regarded not merely as an instrumental improvement, but as a paradigmatic transformation in software engineering. The increasing automation of activities such as coding, testing, maintenance, and operations is redefining both the nature of software development and the roles of software engineers. In the literature, this transformation is described as an evolution of software engineering from a “human-centered” discipline toward a structure based on human–machine collaboration (Storey et al., 2020; Suri et al., 2023). In the future, software development processes are

expected to become more predictable, adaptive, and autonomous through AI-driven systems. This evolution is likely to bring about profound changes in software engineering at both technical and organizational levels.

Current AI-driven software development tools largely function as assistive systems that rely on developer input. However, research suggests that a substantial portion of software development may, in the future, be carried out by semi-autonomous or fully autonomous systems. This perspective is discussed within the framework of "autonomous software engineering" (Suri et al., 2023). Autonomous software development systems consist of AI agents that holistically address processes such as requirements analysis, architectural design, code generation, testing, deployment, and maintenance. These agents can continuously learn from environmental feedback and improve software systems over time. Such an approach has the potential to reduce reliance on human intervention, particularly in large-scale and continuously evolving software systems. Nevertheless, the literature emphasizes that significant open questions remain regarding the reliability, verifiability, and ethical accountability of fully autonomous systems. For this reason, future software development approaches are expected to be shaped not by complete autonomy, but rather by the principle of supervised autonomy, in which human oversight remains an integral component of the development process.

Transformation of Software Engineering Roles

AI-driven software development is also leading to a significant transformation in the roles of software engineers. While the traditional software developer profile has largely focused on coding and debugging activities, future software engineers are expected to possess higher-level competencies such as problem formulation, system design, AI orchestration, and result validation.

In the literature, this emerging role is referred to as the "AI-augmented software engineer" (Storey et al., 2020). Within this framework, software engineers evolve from merely using AI systems to becoming experts who design, supervise, and guide them. This transformation necessitates a restructuring of software engineering education to equip future professionals with the required interdisciplinary skills.

AI in Software Education and Emerging Competencies

The widespread adoption of AI-powered programming tools is exerting a significant influence on the content and pedagogical approaches of software education. In the future, software education is expected to extend beyond teaching the syntax of programming languages to encompass software development competencies in conjunction with AI.

Research indicates that AI-supported learning environments can enhance student productivity; however, excessive automation may weaken fundamental algorithmic thinking and problem-solving skills (Denny et al., 2024). Therefore, balanced educational approaches that position AI as a supportive tool rather than a replacement for foundational learning are of critical importance.

Trustworthy, Explainable, and Responsible AI

In the future, the success of AI-driven software development systems will be evaluated not only based on technical metrics such as performance and speed, but also according to criteria including reliability, explainability, and ethical compliance. Explainable AI (XAI) approaches aim to enhance transparency in software engineering by enabling an understanding of the rationale behind automatically generated code and decision-making processes (Arrieta et al., 2020).

In addition, the energy consumption and environmental impact of AI systems are emerging as important evaluation dimensions for future software systems. In this context, the importance of sustainable AI and green software engineering approaches is steadily increasing.

Overall Assessment and Contribution of the Chapter

This chapter has examined the role of AI in programming and software development processes from a holistic perspective, encompassing the software development lifecycle, maintenance and quality management, modern software paradigms, ethical and legal dimensions, as well as academic and industrial applications. The discussions presented demonstrate that AI-driven software development is not merely a technical innovation, but rather a comprehensive transformation that reshapes the fundamental principles of software engineering.

In conclusion, AI-driven software development represents an inevitable and irreversible trajectory within the software engineering discipline. This chapter aims to serve as a strong reference point for researchers, students, and practitioners seeking to understand the current state and future directions of AI in software development.

Disclosure of AI usage

During the preparation of this work, the author used ChatGPT in order to generate text and check grammar. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the published work.

References

Allamanis, M., Barr, E. T., Devanbu, P., & Sutton, C. (2018). A survey of machine learning for big code and naturalness. *ACM Computing Surveys (CSUR)*, *51*(4), 1-37. https://doi.org/10.1145/3212695

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., ... & Zimmermann, T. (2019a, May). Software engineering for machine learning: A case study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)* (pp. 291-300). IEEE. https://doi.org/10.1109/ICSE-SEIP.2019.00042

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019b, May). Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1-13). https://doi.org/10.1145/3290605.3300233

Amgothu, S., & Kankanala, G. (2024). AI/ML–DevOps Automation. *American Journal of Engineering Research (AJER)*, *13*(10), 111-117.

Arora, C., Grundy, J., & Abdelrazek, M. (2024). Advancing requirements engineering through generative ai: Assessing the role of llms. In *Generative AI for Effective Software Development* (pp. 129-148). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-55642-5_6

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, *58*, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Beck, K., Fowler, M., & Beck, G. (1999). Bad smells in code. *Refactoring: Improving the design of existing code*, *1*(1999), 75-88.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623). https://doi.org/10.1145/3442188.3445922

Bennett, K. H., & Rajlich, V. T. (2000, May). Software maintenance and evolution: a roadmap. In *Proceedings of the Conference on the Future of Software Engineering* (pp. 73-87). https://doi.org/10.1145/336512.336534

Binzer, B., & Winkler, T. J. (2022, October). Democratizing software development: a systematic multivocal literature review and research agenda on citizen development. In *International Conference on Software Business* (pp. 244-259). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-20706-8_17

Chen, M. (2021). Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*. https://doi.org/10.48550/arXiv.2107.03374

Chen, R., Zhang, S., Li, D., Zhang, Y., Guo, F., Meng, W., ... & Liu, Y. (2020, October). Logtransfer: Cross-system log anomaly detection for software systems with transfer learning. In *2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE)* (pp. 37-47). IEEE. https://doi.org/10.1109/ISSRE5003.2020.00013

Cunningham, W. (1992). The WyCash portfolio management system. *ACM Sigplan Oops Messenger*, *4*(2), 29-30.

Dang, Y., Lin, Q., & Huang, P. (2019, May). Aiops: real-world challenges and research innovations. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)* (pp. 4-5). IEEE. https://doi.org/10.1109/ICSE-Companion.2019.00023

Denny, P., MacNeil, S., Savelka, J., Porter, L., & Luxton-Reilly, A. (2024). Desirable characteristics for ai teaching assistants in programming education. In *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1* (pp. 408-414). https://doi.org/10.1145/3649217.3653574

Digkas, G., Ampatzoglou, A., Chatzigeorgiou, A., Avgeriou, P., Matei, O., & Heb, R. (2021). The risk of generating technical debt interest: a case study. *SN Computer Science*, *2*(1), 12. https://doi.org/10.1007/s42979-020-00406-6

Feng, Z. (2020). Codebert: A pre-trained model for program-ming and natural languages. *arXiv preprint arXiv:2002.08155*. https://doi.org/10.48550/arXiv.2002.08155

Fitzgerald, B., & Stol, K. J. (2017). Continuous software engineering: A roadmap and agenda. *Journal of Systems*

and Software, *123*, 176-189. https://doi.org/10.1016/j.jss.2015.06.063

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and machines*, *28*(4), 689-707. https://doi.org/10.1007/S11023-018-9482-5

Gao, J., Tao, C., Jie, D., & Lu, S. (2019, April). What is AI software testing? and why. In *2019 IEEE International Conference on Service-Oriented System Engineering (SOSE)* (pp. 27-2709). IEEE. https://doi.org/10.1109/SOSE.2019.00015

German, D. M. (2024). Copyright-Related Risks in the Creation and Use of ML/AI Systems. In *Large Language Models in Cybersecurity: Threats, Exposure and Mitigation* (pp. 145-152). Cham: Springer Nature Switzerland.

Goswami, M., & Bhatt, A. (2022). Study on Implementing AI for Predictive Maintenance in Software Releases. *International Journal of Research Radicals in Multidisciplinary Fields, ISSN*, 93-99.

Guo, D., Ren, S., Lu, S., Feng, Z., Tang, D., Liu, S., ... & Zhou, M. (2020). Graphcodebert: Pre-training code representations with data flow. *arXiv preprint arXiv:2009.08366*. https://doi.org/10.48550/arXiv.2009.08366

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, *1*(9), 389-399. https://doi.org/10.1038/s42256-019-0088-2

Kim, S. Y., Fan, Z., Noller, Y., & Roychoudhury, A. (2024). Codexity: secure AI-assisted code generation. *arXiv preprint arXiv:2405.03927*. https://doi.org/10.48550/arXiv.2405.03927

Liu, H., Jin, J., Xu, Z., Zou, Y., Bu, Y., & Zhang, L. (2019). Deep learning based code smell detection. *IEEE transactions on Software Engineering*, *47*(9), 1811-1837. https://doi.org/10.1109/TSE.2019.2936376

Mansoor, U., Kessentini, M., Maxim, B. R., & Deb, K. (2017). Multi-objective code-smells detection using good and bad design examples. *Software Quality Journal*, *25*(2), 529-552. https://doi.org/10.1007/s11219-016-9309-7

Mittal, A. (2025, April). AI-Driven DevOps Automation for Cloud-Native Application Modernization. In *International Conference of Global Innovations and Solutions* (pp. 130-147). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-02853-2_9

Naik, P., Nelaballi, S., Pusuluri, V. S., & Kim, D. K. (2024). Deep learning-based code refactoring: A review of current knowledge. *Journal of Computer Information Systems*, *64*(2), 314-328. https://doi.org/10.1080/08874417.2023.2203088

Nedelkoski, S., Cardoso, J., & Kao, O. (2019, May). Anomaly detection and classification using distributed tracing and deep learning. In *2019 19th IEEE/ACM international symposium on cluster, cloud and grid computing (CCGRID)* (pp. 241-250). IEEE. https://doi.org/10.1109/CCGRID.2019.00038

Negri-Ribalta, C., Geraud-Stewart, R., Sergeeva, A., & Lenzini, G. (2024). A systematic literature review on the impact of AI models on the security of code generation. *Frontiers in Big Data*, *7*, 1386720. https://doi.org/10.3389/fdata.2024.1386720

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors*, *39*(2), 230-253. https://doi.org/10.1518/001872097778543886

Patchamatla, P. S., & Owolabi, I. (2025). Comparative Study of Open-Source CI/CD Tools for Machine Learning Deployment. *CogNexus*, *1*(01), 239-250. https://doi.org/10.63084/cognexus.v1i01.34

Sahay, A., Indamutsa, A., Di Ruscio, D., & Pierantonio, A. (2020, August). Supporting the understanding and

comparison of low-code development platforms. In *2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)* (pp. 171-178). IEEE. https://doi.org/10.1109/SEAA51224.2020.00036

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017, May). Membership inference attacks against machine learning models. In *2017 IEEE symposium on security and privacy (SP)* (pp. 3-18). https://doi.org/IEEE.10.1109/SP.2017.41

Šikić, L., Kurdija, A. S., Vladimir, K., & Šilić, M. (2022). Graph neural network for source code defect prediction. *IEEE access*, *10*, 10402-10415. https://doi.org/10.1109/ACCESS.2022.3144598

Sommerville, I. (2016). Software Engineering (10th ed.). Pearson Higher. ISBN-10, 1292096144.

Storey, M. A., Ernst, N. A., Williams, C., & Kalliamvakou, E. (2020). The who, what, how of software engineering research: a socio-technical framework. *Empirical Software Engineering*, *25*(5), 4097-4129. https://doi.org/10.1007/s10664-020-09858-z

Suri, S., Das, S. N., Singi, K., Dey, K., Sharma, V. S., & Kaulgud, V. (2023, September). Software engineering using autonomous agents: Are we there yet?. In *2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)* (pp. 1855-1857). IEEE. https://doi.org/10.1109/ASE56229.2023.00174

Tsoukalas, D., Mittas, N., Chatzigeorgiou, A., Kehagias, D., Ampatzoglou, A., Amanatidis, T., & Angelis, L. (2021). Machine learning for technical debt identification. *IEEE Transactions on Software Engineering*, *48*(12), 4892-4906. https://doi.org/10.1109/TSE.2021.3129355

Tufano, M., Agarwal, A., Jang, J., Moghaddam, R. Z., & Sundaresan, N. (2024). AutoDev: Automated AI-driven development. *arXiv preprint arXiv:2403.08299*. https://doi.org/10.48550/arXiv.2403.08299

Zhang, Q., Fang, C., Xie, Y., Zhang, Y., Yang, Y., Sun, W., ... & Chen, Z. (2023). A survey on large language models for software engineering. *arXiv preprint arXiv:2312.15223*. https://doi.org/10.48550/arXiv.2312.15223

Author Information

Muhammed Coşkun IRMAK

https://orcid.org/0000-0003-3888-1328

Atatürk University

Faculty of Applied Sciences

Department of Data Science and Analytics

Erzurum, Türkiye

Contact e-mail: *irmakm@atauni.edu.tr*

Citation

Irmak, M.C. (2025). Artificial Intelligence in Programming and Software Development. In A. Kaban & T. Demirel (Eds.), *Interdisciplinary Applications of Artificial Intelligence - 1* (pp. 216-232). ISTES.

Chapter 11- Artificial Intelligence in the Internet of Things (IoT) and Intelligent Systems

Önder Yıldırım

Chapter Highlights

- This chapter examines the Artificial Intelligence of Things (AIoT) paradigm, which has emerged from the integration of the Internet of Things (IoT) and artificial intelligence, from conceptual, architectural, and application-oriented perspectives.
- The evolution of IoT systems from data-collection–oriented infrastructures to intelligent systems capable of learning, prediction, and autonomous decision-making is discussed in detail.
- Edge–fog–cloud architectures and federated learning approaches are comparatively analyzed in the context of latency, scalability, and privacy requirements in AIoT systems.
- The roles of supervised, unsupervised, deep, and reinforcement learning algorithms in shaping the cognitive capabilities of intelligent systems are explained within the IoT context.
- Interdisciplinary applications of AIoT in domains such as industry, healthcare, agriculture, and smart cities are examined through domain-specific examples.
- Security, privacy, and explainable artificial intelligence (XAI) are addressed as critical dimensions for the trustworthy and ethical design of AIoT systems.
- The chapter outlines future research directions within the frameworks of Industry 5.0, human-centered design, sustainability, and autonomous/agent-based AIoT systems.
- The holistic perspective presented emphasizes that AIoT is not merely a technical technology, but a transformative paradigm with interdisciplinary and societal implications.

Introduction

Internet of Things (IoT) can be defined as a concept that allows physical objects to take part in digital networks through the help of sensors, embedded systems and wireless communication-channels which produce data out of this connectivity and utilize the derived data as service (Mastorakis et al., 2020). Early IoT applications were largely limited to monitoring, remote control and rule-based automation as decision-making was usually performed by human operators or centralized servers (González García & Núñez-Valdez, 2018).

With the growth of IoT ecosystems, amount–velocity–variety of data that is generated grew significantly, being very hard to obtain valuable information and actionable insights from it (Jagatheesaperumal & Rahouti, 2022). This change made IoT infrastructures not be able to stay as the system of "data collection" but need developing into systems that are capable to learn, adapt and decision making (Zhang & Tao, 2020).

AI-enabled IoT systems therefore need to be designed as distributed and intelligent infrastructures rather than centralized data-processing platforms. Recent studies emphasize that intelligence should be coordinated across the entire edge–fog–cloud continuum in order to achieve low latency, scalability, context awareness, and trustworthy operation in AI-driven IoT environments (Firouzi et al., 2026).

In this perspective, the AI plays a key role in the development of IoT. As such, AI systems are software–hardware composite system that perceive their environment and take actions—by means of the requirement to process data—to achieve goals (and can learn from their environment) in uncertain conditions; like this, it complements an IoT "sense–interpret–act" loop (Ó Fathaigh, 2019). With the help of ML and DL techniques, IoT-enabled systems are able to train themselves with environment patterns to predict future situations and take decisions by themselves (Zhang & Tao, 2020).

The connections IoT+AI are more and more addressed in the literature under the term Artificial Intelligence of Things (AIoT). AIoT is not at all about (simple) AI-based algorithm deployments over IoT: it also involves design principles for example, distributing intelligence across the edge–fog–cloud layers to minimize latency and use bandwidth more efficiently; and addressing issues of privacy in a better way (Firouzi et al., 2022; Shi et al., 2016).

The objective of this chapter is to depict the IoT–AI convergence beyond mere technical intersection, which signifies an interdisciplinary revolution in architectural, algorithmic and design levels involving security–priv acy and sustainability aspects (Mastorakis et al., 2020; Siarry et al., 2021). Thus, the concepts of AIoT and its models, algorithms applications, and security–ethical needs are considered all together (Kök et al., 2023; Kuzlu et al., 2021).

Brainchild of the IoT and Artificial Intelligence Concept

Chapter 11: Artificial Intelligence in the Internet of Things (IoT) and Intelligent Systems

The rapid proliferation of the Internet of Things (IoT) and the parallel advancements in artificial intelligence (AI) have given rise to a new generation of intelligent cyber–physical systems. As IoT environments began producing massive volumes of heterogeneous and real-time data, the limitations of traditional rule-based and centralized processing approaches became increasingly evident. Artificial intelligence emerged as a natural and necessary complement to IoT, enabling systems to move beyond mere data acquisition toward learning, reasoning, and autonomous decision-making. The convergence of these two paradigms has led to the conceptualization of the Artificial Intelligence of Things (AIoT), which represents the brainchild of IoT's pervasive sensing capabilities and AI's cognitive and adaptive intelligence. AIoT embodies a shift from connected systems to intelligent, context-aware, and self-optimizing systems, forming the foundation for next-generation applications across industry, healthcare, smart cities, and beyond.

Internet of Things (IoT) Basics and Building Blocks

IoT is a collection of physical objects that connect to the internet with a unique identifier, generate data and facilitate sharing of this data across multiple systems (Mastorakis et al., 2020). An IoT system usually includes sensing/actuation (sensors–actuators), communication, and application layers; combined with cyber-physical systems, they make possible the integration of physical processes with digital counterparts (González García & Núñez-Valdez, 2018).

IoT scaling raises the need for connectivity, interoperability among heterogeneous devices, low power consumption and real-time operation. Thus, the full-cloud, centralized approach is not ideal in many cases when it comes to latency and network traffic (Shi et al., 2016; Yang et al., 2020). These challenges further emphasize the need to push intelligence toward the edge and leverage distributed architectures in AIoT environments (Firouzi et al., 2022).

Artificial Intelligence Application in IoT

AI techniques are key to understanding the massive and streaming data coming from IoT, for instance in pattern detection, classification, prediction and anomaly discovery tasks (Zhang & Tao, 2020). Although the methods of supervised/unsupervised learning are utilized for finding patterns in sensor data, deep learning enhances the processing of complex data types like images, audios and time series (González García & Núñez-Valdez, 2018).

Within this scheme of things, AI empowers IoT with three fundamental functionalities: (i) situational awareness (sensing + inference), (ii) prediction/forecasting (predictive analytics), and (iii) autonomous decision-making (control/actuation)" (Ó Fathaigh, 2019; Zhang & Tao 2020). Minimization of machine downtime in industry; prediction of risk using wearable sensor data in healthcare; and optimising urban traffic through predicting congestion are some problems that emerge as possible with this set of capabilities (Siarry et al., 2021; Batty et al., 2012).

The AIoT (Artificial Intelligence of Things) Paradigm

The AIoT is the concept that IoT and AI are architected together, and intelligence runs on not just in the cloud but also at edge and fog (Mastorakis et al., 2020). Benefits for this solution include low latency, less traffic and context sensitive, increased privacy (Shi et al., 2016; Firouzi et al., 2022).

In addition to simply distributing computation, the proposed architecture highlights a unified framework in which computing resources and AI capabilities collaborate across layers. This architectural view treats AIoT not as a collection of isolated components but as an integrated system capable of adaptive, scalable, and context-aware decision making (Firouzi et al., 2026).

The emergence of AIoT also raises the bar for trustworthiness and explicability. Because black-box decisions are unacceptable in the presence of risk (healthcare, transportation or industry), explainable AI (XAI) methodologies now stand out as key mechanisms for AIoT design (Kök et al., 2023).

AI are IoT Architectures (Edge–Fog–Cloud, Federated Learning)

But bringing all the data to the cloud in IoT is not feasible for every use case because of latency, bandwidth cost issues and privacy of data. Accordingly, AIoT is redefined on the network architectures with computation and intelligence distributed to each layer in the network (Shi et al., 2016; Yang et al.

The Edge–Fog–Cloud Continuum

The edge–fog–cloud spectrum suggests the need for closer-to-source preprocessing and inferences on lower-cost, lower-power nodes from where the data emanates at source (edge/fog) than provided by cloud while dreaming about training large, will long-term analytics or grand global optimization processes with ed by revolutionary computational resources (Shi et al., 2016). As a result, latency-sensitive inference is at the edge and large scale model training/big data analytics are in cloud (Firouzi et al., 2022).

Beyond performance optimization, the edge–fog–cloud continuum is also positioned as a foundation for security, privacy, and trust. Processing data closer to the source reduces unnecessary exposure of sensitive information and enables compliance-aware, resilient systems that balance performance and ethical constraints (Firouzi et al., 2026).

This structure reduces the communication overhead and energy consumption, as well as lowers the privacy leakage. From the perspective of health and industry, if you transfer "summaries/parameters but not raw data", this renders beneficial aspects regarding system performance and regulation issue in both domain (Siarry et al., 2021; Shi et al., 2016).

Edge AI and Real-Time Smart Systems

Edge AI is running AI inferencing on the IoT devices or nodes that are close to these. It reduces the latency since decisions can be made on-site and makes it less likely that data with privacy implications get leaked out of the device (Shi et al., 2016). Edge AI has significant advantages especially for computer vision, event detection, and anomaly detection (Zhang & Tao, 2020).

Owing to edge limitations, "lightweight" models that leverage strategies such as model compression, quantization, knowledge distillation are commonly built. Thus, the energy consumption of edge devices can be controlled while intelligent functions are sustainable with reasonable accuracy (Yang et al., 2020; Zhang & Tao, 2020).

Federated Learning and Distributed Learning

Federated Learning (FL), Fedegration learning aims to construct a shared model by allowing edge devices, where data exists, to learn locally and share only the learned model updates rather than centralizing raw data. This method is a good choice for AIoT based on privacy and bandwidth (Liu et al., 2023).

Within the edge–fog–cloud continuum, federated learning is not only a learning technique but also an architectural strategy that preserves data ownership while enabling collaborative intelligence. By exchanging model parameters instead of raw data, AIoT systems can scale while maintaining privacy and trust among participating nodes (Firouzi et al., 2026).

Security/robustness is an absolute concern in FL deployments; poisoning, malicious update and model manipulation are among the possible threats to hurt global models. As such, federated mechanism like malicious model update detection is introduced for the protection and creation of reliable FL on AIoTs gains popularity (Liu et al., 2023).

FL + Blockchain: The FLchain Method

FL FL's reliance on a central aggregator could cause single point of failure, trust and transparency problems. In a cyber–physical setting, it can addiscover to store model upgrades immutably in ledgers and provide more decentralized coordination (Nguyen et al., 2021).

The method of Nguyen et al. (2021) achieves a reduction in dependency on a central server focused not only in security, but also traceability and scalability. Nevertheless, the latency and energy consumption of blockchain demand performance optimization for edge applications (Nguyen et al., 2021; Shi et al., 2016).

Artificial Intelligence Algorithms in Intelligent Systems

Smart systems also involve learning and adaptive control in addition to acquiring (gathering) and interpreting the information from environment around them. These functions are realized by AI algorithms families, which are decided according to the character of IoT data and needs of decision-making (Zhang & Tao, 2020).

Supervised and Unsupervised Learning

In supervised learning, classification/regression models are generated from labeled data and it is widely used in AIoT for fault detection, event classification, and state recognition (González García & Núñez-Valdez, 2018). Unsupervised learning with clustering and anomaly detection on unlabeled data can capture the "unexpected behaviors," especially in heterogeneous IoT networks (Zhang & Tao, 2020).

In fact, AIoT systems commonly are of hybrid nature (unsupervised approach to pattern/anomaly detection followed by supervised models contributing quality decisions). This design aids in dealing with the cost of real-life labelling, and data variability (Gonzáles García & Núñez-Valdez, 2018).

Deep Learning and Time-Series Analysis

The vast majority of IoT data is time series (sensor readings, device events, environmental measurements). Recurrent networks, in particular the popular LSTM and GRU and their newer variants are commonly utilized for prediction and anomaly detection (Siarry et al., 2021). For the case that images/audio are collected by IoT devices, CNN-based methods achieve promising results in various domains including security, transportation and medical monitoring (Zhang & Tao, 2020).

t time, the deployment of deep learning on the edge will present challenges for model efficiency when energy is scarce. Whereas edge-friendly models and methodology at architectural level bear the decision of whether AIoT implementations are feasible (Shi et al., 2016; Yang et al., 2020).

Reinforcement Learning and Autonomous Decision-Making

RL is one approach that provides autonomous control through learning policies based on reward–penalty mechanisms in a changing environment. In AIoT, RL can be an essential tool for adaptive fields such as smart traffic management, energy optimization and robotic control (Firouzi et al., 2022).

But RL creates problems as well, such as sample efficiency, safe exploration and computational complexities. Therefore, the safe and light-weight RL on edge/fog is an important direction of research in the evolution AIoT towards autonomous systems (Shi et al., 2016; Yang et al., 2020).

Interdisciplinary AIoT Applications

Through the real-time perception and decision-making capabilities, AIoT promotes "data-based management" in numerous fields. These systems seek to achieve full efficiency, safety and sustainability at the same time (Mastorakis et al., 2020; Firouzi et al., 2022).

Industrial IoT (IIoT) and Smart Manufacturing

AIoT enables applications like predictive maintenance, process optimization, and digital twins in the Industry 4.0 scenario. Big data analytics and AI shift manufacturing processes from passive monitoring to predictive and autonomous management system (Jagatheesaperumal & Rahouti, 2022).

In IIoT use cases, edge/fog architectures facilitate latency-free real-time quality control and safety-critical decisions. Therefore, reliance on centralized clouds is reduced and the resilience of industrial systems becomes higher (Yang et al., 2020; Shi et al., 2016).

AIoT for Healthcare (Internet of Medical Things—IoMT)

In the healthcare services, AIoT ensures a continuous flow of data through wearable sensors and remote detectors for potential early warning systems and customized care opportunities (Siarry et al., 2021). Quality of data, reliability and explicability of the model indeed are crucial for clinical decision support (Kök et al., 2023).

Health data is extremely sensitive and distributed learning approaches like FL are popular for minimizing privacy concerns by keeping data on the edge. There is also an increasing trend in environmental/sustainability aspects of medical AI, including energy/carbon footprint of medical AI (Liu et al., 2023) and bio-waste management in the context of AI-driven medicine (Doo et al., 2024).

Unified edge–fog–cloud architectures support these scenarios by enabling real-time analytics at the edge, intermediate coordination at the fog, and long-term optimization in the cloud, creating a trustworthy foundation for mission-critical AIoT applications (Firouzi et al., 2026).

Smart Agriculture and Environmental Monitoring

The AIoT is helping in precision agriculture to increase yields and to save natural resources by tracking soil moisture, temperature and plant health (Hunter et al., 2016) and in making smart irrigation/fertilization decisions (Mastorakis et al., 2020). Analytics at the edge enables on-device decision-making with variable connections such as field/greenhouse environment_temp (Shi et al., 2016).

In the field of environmental monitoring, AIoT can contribute to sustainability by enabling real-time anomaly detected under various circumstances including air quality management and disaster early warning water contamination. Therefore, the "green AIoT" solution needs to tackle both monitoring applications and eco-impact of AI computation together (Doo et al., 2024).

Smart Cities and Transportation

In fact, the underlying objective of smart cities is to leverage data analytics for optimal management of city-scale infrastructures (Batty et al., 2012). By integrating the data from sensors with AI, AIoT or e-intelligence can enhance decision quality for traffic flow analysis, dynamic signal control and demand forecast (Zhang & Tao, 2020).

In such applications, we have to consider more privacy (e.g., location data) and fairness/ethics as well as explaintability; moreover security problems in large scale of city network cannot be ignored. Thus, urban-scale AIoT architecture must be studied in combination with XAI and secure architectural principles (Kök et al., 2023; Kuzlu et al., 2021).

Security, Privacy, and Trustworthiness

By connecting the physical world with digital intelligence, AIoT increases the attack surface: IoT devices and AI models are two kinds of targets. Thus, the security of AIoT has a multi-layer structure and collectively involves device, network, data as well as model layers (Kuzlu et al., 2021).

Cybersecurity Threats in IoT Systems

In the context of IoT, lackluster authentication, absence of patches and standardization problems result in attacks such as distributed denial of service (DDoS), unauthorized entry and malware. Meanwhile, in AIoT, other threats can be exploited directly against decision making mechanisms such as data poisoning (Zhang et al., 2019), model reverse engineering and adversarial sample attacks (Kuzlu et al., 2021; Liu et al., 2023).

In the context of federated learning, malicious clients can pollute a global model by sending poisoned updates. As a result, "model update security" and attack detection mechanisms are two significant research problems in AIoT (Liu et al., 2023; Nguyen et al., 2021).

AI-Based Defense Mechanisms

AI itself is not merely a threat to AIoT security; it serves as a protective shield. Instead of the signature-based solutions, ML/DL-based security analytics ceased to adheres to static methodology and provides more dynamic innovation anomaly detection, attack classification,and Behavioral analysis (Kuzlu et al., 2021).

Yet the models that serve as the basis for security can themselves be attacked, and therefore trustworthy AI principles (robustness, verifiability, traceability) are crucial in AIoT design (Kök et al., 2023; Ó Fathaigh, 2019).

Explainable AI (XAI) and Trustworthy AIoT

Explainability of AIoT decisions is important in safety critical applications for accountability, debugging and user's trust. XAI tries to give human-understandable rationales of the model outputs, and is an important part of the trustworthiness in AIoT (Kök et al., 2023).

In what we refer to as distributed and hetegeneous AIoT systems, transparency is not only an ethical issue but also and engineering one to ensure operating reliability. In particular, under FL/edge settings, explanations are beneficial for interpreting where and why a model goes wrong and avoiding risky decisions (Kök et al., 2023; Liu et al., 2023).

Future Directions: Intelligent and Autonomous Systems

AIoT is evolving the concept of smart IoT systems to more autonomous, adaptive and human-centric systems. This transformation entails aspects of governance, ethics and sustainability as architectural choices (Firouzi et al., 2022; Ó Fathaigh, 2019).

Agent-Based and Autonomous AIoT

In AIoT systems addressing such multi-agent collaboration, localized sensing, and decision-making for edge nodes will be increasingly the case; despite this emergent shift to an “agent-like” structure as found here in our approach is that of agent-based formations (Firouzi et al., 2022). This has the potential for scalable control mechanisms in robotics, autonomous transportation and industrial automation (Yang et al., 2020).

Industry 5.0 and Human-Centered Design

The Industry 5.0 point of view underlines not only efficiency but also human–machine cooperation and trust, ethics and sustainability. In this framework, the implementation of AIoT not only results from “intelligence to enhance humans” but also approaches with the premise that it must be a "collaborative combination with humans" (Kök et al., 2023; Ó Fathaigh, 2019).

Sustainability and Green AIoT

The large computational overhead and energy consumption of AI model have raised the attention to energy-caused AIoT environment issues. As a result, green AIoT emphasizes on-device inference at the edge, minimizing data

exchanges undesired and an energy-considerate model construction (Shi et al., 2016). AI's influence on environmental sustainability In areas that are computation-intensive, such as healthcare, AI's effect on the environment orthe ecosystem continues to be coined a "two-edged sword" (Doo et al., 2024).

Open Research Problems

A few of the white spots in AIoT research that are still open include device diversity, scalable secure FL system, adversarial robustness, edge-oriented XAI and sustainable computing (Kök et al., 2023; Liu et al., 2023). Furthermore, in the blockchain-embedded FLchain solutions, low latency and minimal energy costs are vital requirements as well for the practical implementation (Nguyen et al., 2021).

Conclusion

This chapter has systematically discussed the AIoT paradigm that is derived from the fusion of IoT with AI, from conceptual foundations and architectural recipes to algorithms, applications domains, and security-privacy-trustworthiness dimension (Mastorakis et al., 2020; Zhang & Tao, 2020). The transition of IoT from rule-based automation to learning and decision-making systems has been further accelerated by the ability of AI to process data and make inferences (Ó Fathaigh, 2019).

Even though the continuum from edge to fog to cloud fulfills AIoT requirements, regarding real-time quality of service and scalability, federated learning has significant benefits with respect to privacy-preserving and data ownership (Shi et al., 2016; Liu et al., 2023). Blockchain integration (FLchain) can solve the trust/coordination problems of FL with the promise of transparency, and higher-levels of decentralization, but at high cost in terms of energy consumption and latency (Nguyen et al., 2021).

In applications from cross-discipline, AIoT provides wide areas of impact such as predictive maintenance and industry optimization, remote monitoring with decision support in healthcare, resource efficiency in agriculture and data driven urban management in cities (Jagatheesaperumal & Rahouti, 2022; Siarry et al., 2021; Batty et al., 2012). However, security, privacy and explainability continue to be essential design pre-requirements for the social acceptance and reliable operation of AIoT in critical sectors (Kök et al., 2023; Kuzlu et al., 2021).

Perspectives in the future suggest that AIoT will also transition to system of systems (autonomous, agent-oriented & human centered) and sustainable/green AI approaches would be at heart of technology agenda (Firouzi et al., 2022; Doo et al., 2024). However, AIoT should be a social technological change domain that is influenced by trust (Kök et al., 2023; Ó Fathaigh, 2019), ethics and sustainability.

As large language models and agentic AI are increasingly integrated into IoT environments, architectural choices become tightly linked to energy efficiency, resilience, transparency, and trust. The edge–fog–cloud continuum offers a sustainable pathway for deploying these capabilities while addressing privacy, explainability, and system robustness (Firouzi et al., 2026).

Disclosure of AI usage

During the preparation of this work, the author(s) used [OpenAi-ChatGPT and Google Gemini] in order to [e.g., generate text, check grammar, analyze data, or assist with literature search]. After using this tool/service, the author reviewed and edited the content as needed and take full responsibility for the content of the published work.

References

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., & Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics, 214*(1), 481–518.

Doo, F. X., Vosshenrich, J., Cook, T. S., Moy, L., Durand, D. J., Harvey, H., … Hanneman, K. (2024). Environmental sustainability and AI in radiology: A double-edged sword. *Radiology, 310*(2), e232030.

Ó Fathaigh, R. (2019). *European Commission: High-Level Expert Group on Artificial Intelligence publishes Ethics Guidelines for Trustworthy AI.* In IRIS (Vol. 2019, No. 3).

Firouzi, F., Farahani, B., & Marinšek, A. (2022). The convergence and interplay of edge, fog, and cloud in the AI-driven Internet of Things (IoT). *Information Systems, 107*, 101840. https://doi.org/10.1016/j.is.2021.101840

Firouzi, F., Farahani, B., & Marinšek, A. (2026). Reflection on the convergence and interplay of edge, fog, and cloud in the AI-driven Internet of Things (IoT). *Information Systems, 138*, 102662. https://doi.org/10.1016/j.is.2025.102662

González García, C., Núñez-Valdez, E. R., García-Díaz, V., Pelayo G-Bustelo, B. C., & Cueva-Lovelle, J. M. (2019). A Review of Artificial İntelligence in The Internet of Things. International Journal of Interactive Multimedia and Artificial Intelligence, 5(4), 64–70. https://doi.org/10.9781/ijimai.2018.03.004

Jagatheesaperumal, S. K., & Rahouti, M. (2022). The duo of artificial intelligence and big data for Industry 4.0: Applications, challenges, and opportunities. *IEEE Internet of Things Journal, 9*(15), 12857–12885. https://doi.org/10.1109/JIOT.2021.3139827

Kök, İ., Okay, F. Y., & Muyanlı, Ö. (2023). Explainable artificial intelligence (XAI) for Internet of Things: A survey. *IEEE Internet of Things Journal, 10*(16), 14764–14779. https://doi.org/10.1109/JIOT.2023.3287660

Kuzlu, M., Fair, C., & Guler, O. (2021). Role of artificial intelligence in the Internet of Things (IoT) cybersecurity. *Discover Internet of Things, 1*(1). https://doi.org/10.1007/s43926-020-00001-4

Liu, W., Lin, H., Wang, X., Xu, H., Zhang, L., & Choo, K.-K. R. (2023). D2MIF: A malicious model detection mechanism for federated-learning-empowered AIoT systems. IEEE Internet of Things Journal, 10(3), 2603–2613. https://doi.org/10.1109/JIOT.2021.3128919

Mastorakis, G., Mavromoustakis, C. X., & Mongay Batalla, J. (Eds.). (2020). Convergence of artificial intelligence and the Internet of Things. Springer. https://doi.org/10.1007/978-3-030-44907-0

Nguyen, D. C., Ding, M., Pham, Q.-V., Pathirana, P. N., Le, L. B., Seneviratne, A., Li, J., Niyato, D., & Poor, H. V. (2021). Federated learning meets blockchain in edge computing: Opportunities and challenges. IEEE Internet of Things Journal, 8(16), 12806–12825. https://doi.org/10.1109/JIOT.2021.3072611

Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637–646. https://doi.org/10.1109/JIOT.2016.2579198

Siarry, P., Jabbar, M. A., Aluvalu, R., & Abraham, A. (Eds.). (2021). The fusion of Internet of Things, artificial intelligence, and cloud computing in health care. Springer. https://doi.org/10.1007/978-3-030-75220-0

Yang, L., Chen, X., Perlaza, S. M., & Zhang, J. (2020). Special issue on artificial-intelligence-powered edge computing for Internet of Things. *IEEE Internet of Things Journal, 7*(10).

Zhang, J., & Tao, D. (2021). Empowering things with intelligence: A survey of the progress, challenges, and opportunities in artificial intelligence of things. *IEEE Internet of Things Journal, 8*(10), 7789–7817. https://doi.org/10.1109/JIOT.2020.3039359

Author Information

Önder Yıldırım

https://orcid.org/0000-0003-4333-2323

Atatürk University

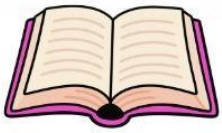
Oltu, Erzurum

Türkiye

Contact e-mail: *ondery@atauni.edu.tr*

Citation

Yıldırım, Ö. (2025). Artificial intelligence in the Internet of Things (IoT) and intelligent systems. In A. Kaban & T. Demirel (Eds.), *Interdisciplinary Applications of Artificial Intelligence - 1* (pp. 233–245). ISTES.



Artificial intelligence is no longer a future concept —it is an integral part of today's research, education, and software ecosystems. This book provides a clear and accessible introduction to artificial intelligence, from fundamental concepts to cutting-edge developments such as generative AI, large language models, and AI-driven research platforms. Each chapter bridges theory and practice, helping readers understand how AI technologies are designed, applied, and evaluated across disciplines.

Beyond technical foundations, the book critically addresses key issues including data ethics, privacy, security, and human–AI interaction. By emphasizing responsible use and human-centered design, it offers a balanced perspective on both the potential and the limitations of artificial intelligence in real-world contexts.

Written for researchers, educators, graduate students, and professionals, this volume serves as a practical guide and a conceptual reference for anyone seeking to engage with artificial intelligence in an informed and interdisciplinary manner.

ISTES