EVOLUTION OF ARTIFICIAL INTELLIGENCE FROM TRADITIONAL TO SUPER INTELLIGENCE

Prof. Dr. Ercan Oztemel
Marmara University

Dr. M. Esad Oztemel
TUBITAK, BİLGEM, Institute of Artificial Intelligence

Attribution: Oztemel, E., & Oztemel M. E. (2026). Evolution of artificial intelligence from traditional to super intelligence. In H. Arslan, A. Taşdelen, E. Kuzucu Hıdır, & O. A. Çetin (Eds.), *Artificial intelligence and national technology reflections* (pp. 27–58). Turkish Academy of Sciences. https://doi.org/10.53478/TUBA.978-625-6110-86-1.ch02

10.53478/TUBA.978-625-6110-86-1.ch02

EVOLUTION OF ARTIFICIAL INTELLIGENCE FROM TRADITIONAL TO SUPER INTELLIGENCE

Prof. Dr. Ercan Oztemel
Marmara University

Dr. M. Esad Oztemel
TUBITAK, BİLGEM, Institute of Artificial Intelligence

Abstract

Evolution of artificial intelligence from traditional to super intelligence takes the attention of not only the scientific community but also societies. There is a remarkable transformation from narrowly defined, rule-based systems to ever increasingly autonomous and generative machine cognition leading to general and responsible intelligence. The progress of AI is indicating not only from technical aspects but also some philosophical transformation. Traditional knowledge processing systems are leaving their place to data-driven intelligence sustaining deep intelligence and responsibility to some degree. This transformation ensures that early AI methodologies with symbolic reasoning and explicit human knowledge evolving into a more rationalist and human-centered modelling of intelligence. Today, data-driven and generative approaches focus more on deep learning and predictive performance with some degree of transparency and interpretability. This chapter provides a systematic analysis and overview of the evolution of AI paradigms, focusing on shifting respective capabilities from traditional approaches to machine learning, from generative AI to explainable AI, and artificial general intelligence. First, traditional AI and machine learning approaches are discussed, and then the respective AI transformations are summarized. While highlighting paradigm shifts, attention of the reader will also be taken to capacity changes and philosophical shifts experienced. It is intended that this chapter offers a conceptual foundation for future development of artificial intelligence systems in alignment with human intelligence.

Keywords

Artificial Intelligence, Machine Learning, Traditional AI, Evolution of AI, Generative AI

Introduction

The field of artificial intelligence (AI) has advanced at an unprecedented pace in the last decade, particularly with the maturation of deep learning technologies and the ease of access to large datasets. Initially a tool for performing computational and analytical tasks, AI has now evolved into a force with the potential to produce systems that approach, and even surpass, human creativity and cognitive abilities in some areas such as image recognition, language translation and code generation.

The idea of generating robots imitating human behavior is not new, it goes back to the 13th century. There have been so many attempts such as the invention of ablution robot and the Elephant Water Clock, built by Al-Jazari in 1206, became a timepiece for civilizations of its time, utilizing the buoyancy of water to determine the time of day (Al-Jazarī, 1973). The development of the Mechanical Turk, a chess playing platform created in the 18th century which is well before the AI concept was introduced (Stephens, 2023). Furthermore, Takuyyiddin invented the six-cylinder pump and the gravity-powered clock (Al-Hassani & Al-Lawati, 2009). Laghari Hasan Çelebi's attempts to fly with the rocket he developed, and Ibn al-Haytham's invention of the first camera and vision system of the time, have earned their place in the pages of history (Al-Hassani, 2006).

Imitating “system behavior” was the driving force at that time. This was triggering the development and generation of machines and was underlying the foundation of today’s intelligent machines. During the beginning of the 18th Century, imitating “human behavior” was taken into the center of the research. Mechanical Turk was one of the very first initiatives for this. The Mechanical Turk, also known as the Automation Chess player, was first displayed in 1770, being able to play a game of chess autonomously, but whose pieces were in reality moved via levers and magnets by a chess master hidden in its lower cavity. Scientific developments and invention of electricity as well as the need for automated machines lead to the invention of computers. Being able to computerize human behavior was opening the way to generate more functional capable machines leading to AI research which was officially declared in a conference in 1956 (Russell & Norvig, 2021). The concept of intelligence became the main motto indicating the generation of intelligent machines. Since then, the concept of "artificial intelligence" is used and became the main motto indicating the generation of intelligent machines

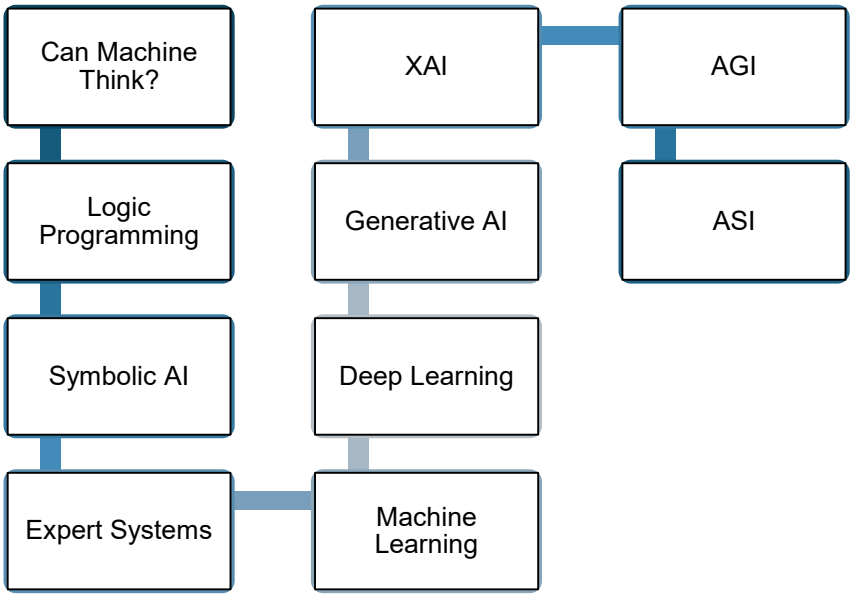
The conceptual foundations of Artificial intelligence as a scientific discipline date back to the mid-20th century, particularly to the work of Alan Turing. Turing laid the theoretical groundwork for modern computer science and artificial intelligence by

defining the limits of algorithmic thought with his abstract computational model, which he called the Turing Machine (Bowen, 2017). Early AI research, between the 1950s and 1970s, focused on Symbolic Systems (Good Old-Fashioned AI) and Expert Systems (Nilsson, 2013). This approach aimed to create intelligence by explicitly programming human knowledge and reasoning rules into logical symbols and "If-Then" rules. Later, the Machine Learning (ML) approach, which became popular in the 1980s and 1990s, focused on AI's ability to learn from data rather than being programmed with rules (Fradkov, 2020). During this period, algorithms learned to analyze large data sets, extract patterns, and make decisions based on these patterns. From the 2010s onward, Deep Learning (DL), which uses multilayered neural networks inspired by the human brain, revolutionized complex tasks like computer vision and natural language processing through the combination of big data and powerful processors, igniting the current rise of AI (LeCun et al., 2015).

The most concrete evidence of this transformation is so called Generative Artificial Intelligence (GenAI) which represents AI's transition from the ability to analyze and interpret data alone to the ability to create brand-new and original content including text, images, code, video (Feuerriegel et al., 2024). Whereas once a computer was expected to simply make data-based predictions, it is now expected to write an article (Li et al., 2025), design an artistic image (Ringvold et al., 2024), or prototype complex software on command (Nguyen-Duc et al., 2025). Generative artificial intelligence for software engineering-A research agenda. Software: Practice and Experience, 55(11), 1806-1843.). This is a paradigmatic shift in AI that is reshaping everything from business to art, from science to everyday communication. However, GenAI is merely a stopover. The ultimate goal of AI research is to reach Artificial General Intelligence (AGI), systems with general-purpose cognitive capabilities on par with human intelligence (Yenduri et al., 2025), and then to Super Artificial Intelligence (ASI), a theoretical stage that surpasses all of humanity's intellectual capabilities (Mustafa et al., 2025). Figure 1. demonstrates the historical progress, philosophical and paradigm changes of AI research.

Figure 1

Historical Evolution of Artificial Intelligence Paradigms



There have been so many very popular AI methodologies, technologies or models developed during this progress. Some of those have been well-accepted by the scientific community and have subsequently become part of human life. Some of the most common methodologies can be listed in chronological order as follows.

- First-Order Logic – 1950s
- Rule-Based Systems or Production Systems – Late 1950s
- Search Algorithms (DFS, BFS, A*) – 1950–60s
- General Problem Solver (GPS) – 1959
- Semantic Networks (Frames) – Late 1960s
- HEARSAY-II / Blackboard Systems – 1971–1976
- Truth Maintenance Systems – 1979
- Non-Monotonic Reasoning – Late 1970s
- Case-Based Reasoning – Late 1970s
- Fuzzy Logic – 1980s
- Genetic Algorithms – 1980s
- Artificial Neural Networks (ANN – Backpropagation) – 1986s
- Decision Trees (ID3, C4.5) – Late 1980s
- Naive Bayes Classifiers – 1985
- Swarm Intelligence (Ant Colony, Bee Colony, Particle Swarm) – 1989–1995
- Distributed Artificial Intelligence – 1990
- Support Vector Machines (SVM) – 1992

- Random Forests – 2001
- Reinforcement Learning – Mid-2000s
- Multi-Agent Systems (Multi-Agent Systems) – 1990s–2000s
- Cognitive Architectures (ACT-R, SOAR) – 1990s–2000s
- Deep Learning – 2012 (AlexNet)
- Convolutional Neural Networks (CNN) – Post-2012
- Recurrent Neural Networks (RNN, LSTM) – 2013+
- Word Embedding Systems (Word2Vec, GloVe) – 2013+
- Transfer Learning – 2015+
- Learning Federation 2016+
- Transformers – 2017+
- Generative AI (BERT, GPT) – 2018+
- Explainable Artificial Intelligence (XAI) – 2018+
- Artificial General Intelligence (AGI) – 2020+
- Causal or Responsible AI – 2022+

There are undoubtedly some other technologies such as model-based reasoning, non-monotonic reasoning, and qualitative reasoning methods not listed above. However, this list outlines well recognized AI methodologies.

This chapter outlines this exciting journey of AI, starting from traditional approaches to Generative intelligence, from the current capabilities of GenAI, towards the goal of AGI, which is expected to lead to an intelligence explosion, and ultimately towards the horizon of ASI, which holds both great opportunities and existential risks for humanity and presents a comprehensive evaluation of the progress from traditional AI systems to ASI. In this context, Chapter 2 provides an overview of traditional AI and fundamental machine learning techniques. Chapter 3 introduces the concept of deep learning and generative AI by highlighting their advancements as well as respective impact. Chapter 4 focuses on explainable AI (XAI), taking attention to the importance of transparency and interpretability whereas Chapter 5 examines AGI and ASI, discussing their conceptual foundations and potential implications. Finally, Chapter 6 offers concluding remarks and summarizes the key insights of the Chapter.

1. Traditional AI and Machine Learning

AI has become a fundamental force driving transformation not only in technology but also in a wide range of fields, including economics, healthcare, education, defense, and the social sciences. It is important to understand the basic foundations of AI in order to comprehend the progress along this line. It is well-known that modeling human-like cognitive abilities first gained scientific ground in the 1950s when Alan Turing questioned if the machines could think. It has experienced continuous development in both theoretical and applied fields since then. While there were some periods when

the research was slowed down, it has now gained unprecedented momentum. Particularly thanks to the increasing amount of data, the decreasing cost of computing resources, the increased availability of computing power, and the evolution of algorithmic methods.

In the 1960s and 1970s Logic Programming emerged as a methodology that represented knowledge through formal logical statements and led to reasoning through automated theorem proving. Languages like Prolog and Lisp, developed in 1972, allowed programmers to define facts and rules, enabling machines to derive conclusions through logical propositions. This approach was interesting at that time in generating automated reasoning capabilities for computers as real-life events could be clearly formulated in terms of logical relationships. However, logic programming faced significant limitations when dealing with uncertainty, inexact and incomplete knowledge. The complexity of real-life events was making it difficult to design logical interrelationships.

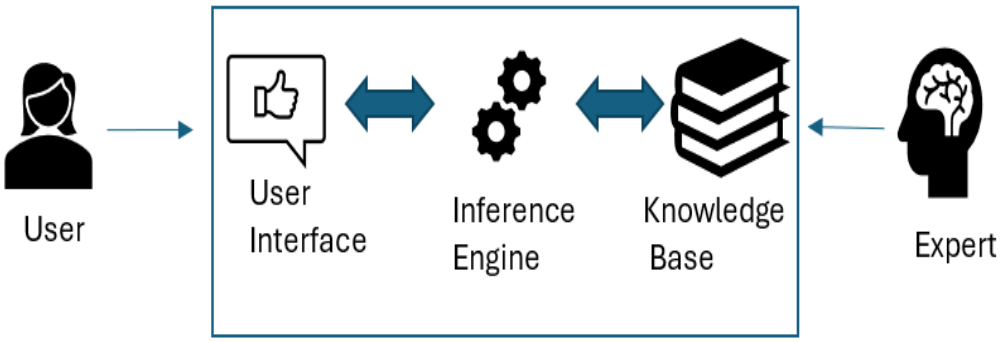
This led the AI researchers to concentrate on devising intelligent systems based on symbols which can be considered as the era of Symbolic AI, which dominated the field up until the 1980s. The idea of Symbolic AI was based on the understanding that human intelligence can be represented through the manipulation of symbols processed according to a predefined set of rules. This also triggered the idea of generating intelligent machine behavior through logical operators and symbols which were also leading to significant success in problem-solving, game playing (such as chess programs), and theorem proving. General Problem Solver and various natural language understanding programs were developed during this period. However, despite some successful applications, the AI research community were still facing challenges in generating intelligent systems capable of performing pattern recognition, learning and adaptation to environmental changes etc. due to the difficulty of articulating expert knowledge in form or rules.

Despite this, Expert Systems, of which the respective components depicted in Figure 2, were the first major commercial success of AI in the 1980s (Jackson, 1998). They were able to capture the specialized high-level knowledge of human experts in specific domains and encoded it as a set of if-then rules within a knowledge base. Combined with an inference engine that could apply these rules to new situations, expert systems like MYCIN (for medical diagnosis), DENDRAL (for chemical analysis), and XCON (for computer configuration) demonstrated practical value in well-defined domains. At the beginning, commercial organizations invested heavily in these systems, and the expert system industry became very popular throughout the 1980s. When facing various challenging and complex real-life problems, they reveal critical weaknesses. Knowledge acquisition was taking too much human effort and did become a bottleneck in AI development. It was difficult for information technology specialists to have extensive collaboration with domain experts. Besides, these

systems were not well-performing outside their knowledge stored in their knowledge base. This was also making it very difficult to maintain this system as the domain rules grew in number and complexity. Additionally, these systems were presenting the lack of ability to learn from new data or experiences.

Figure 2

Components of an Expert System



Having these problems was causing a decrease or cut of AI research funds. Interest in intelligent machines was diminishing. This skepticism from the 1970s and lasted until the early 1990s was called AI winters. The first AI winter followed with the Report published by Lighthill (1973). This report criticized AI research for failing to deliver on its ambitious promises and led to significant funding decreases and cuts started in the United Kingdom and spread throughout the whole scientific community. Second, winter began in the late 1980s when expert systems failed to satisfy commercial expectations. Developing an expert system was too expensive and time consuming but the capacity was bound within the limit of domain knowledge encoded. Additionally, commercial failure of Lisp machines also triggered the understanding of symbolic AI not being able to cope with complex, real-world problems. This was growing the disillusion about the potential of AI and having grants or investment funds was becoming more and more difficult.

These winters taught the AI community valuable lessons about managing expectations, the importance of addressing fundamental theoretical limitations, and the need for approaches that could learn and adapt rather than relying solely on hand-coded knowledge. The field eventually emerged from these winters through the development of machine learning techniques, neural networks, and statistical approaches that proved more robust and scalable than their symbolic predecessors.

One of the topics that has attracted attention and become quite popular during this period is the application of fuzzy logic which is designed to process uncertain and inexact knowledge. It was later understood as computing with words and brought the capability of understanding words by machines. Fuzzy propositional logic, based on fuzzy set theory developed by Asker Lütfi Zadeh (1965), has surpassed other

uncertainty management methodologies of its time and continue to be standard information processing and well accepted problem-solving tools. More detailed information can be found in Cox (1994).

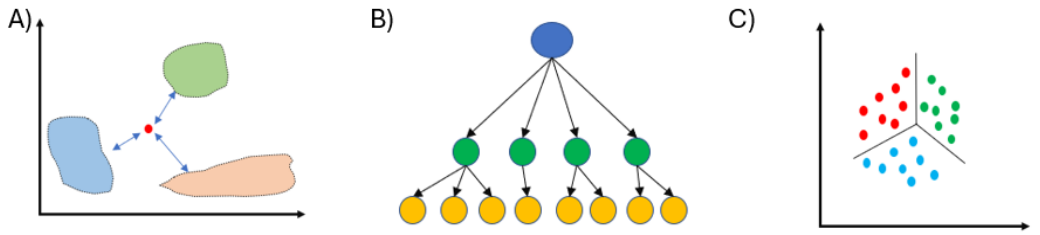
As the limitations of symbolic methods became apparent, artificial intelligence research underwent a radical change of direction starting in the 1990s. During this period, increasing computer processing power, the emergence of large data sources, and the integration of statistical methods into computer science accelerated the rise of machine learning (Langley, 2011). Intelligence began to be defined not by rigid rules, but by models that could learn from data, recognize patterns, and improve with experience (Brantly, 2020). Thus, the concept of artificial intelligence became more dynamic, adaptable, and possessed a high predictive capacity. Machine learning studies in those times have led to various learning strategies (supervised learning, unsupervised learning, reinforcement learning) and algorithms/models (Multilayer Perceptron, Hopfield Network, ART Network, Support Vector Machine, Random Forest, Linear Regression, LSTM, CNN, etc.) through learning strategies applied and learning rules employed as well as the topology of the learning system etc. In addition to artificial neural networks (Öztemel, 2020), Genetic Algorithms (Goldberg, 1989), Random Forest (Sullivan, 2018), and Support Vector Machine (Brasseur, 2011) have become popular.

In machine learning studies, alongside artificial neural networks and related models, Genetic Algorithms have also gained significant popularity. They are particularly effective at successfully solving very complex, non-polynomial problems. They can provide faster solutions to problems that are difficult or impossible to solve with normal optimization methods (Holland, 1975). While finding the most accurate solution among a very large number of alternatives can take tens, sometimes hundreds, sometimes thousands of years, genetic algorithms can provide good approximate solutions to these problems within a few hours. They are particularly successful in problems such as determining vehicle routes and assigning tasks to machines. More detailed information can be found in Goldberg (1989).

Machine learning approaches fundamentally aim to provide systems with the ability to automatically extract information from data. Early examples include methods such as decision trees (Jijo & Abdulazeez, 2021), neural networks (Oztemel, 2020), support vector machines (Pisner & Schnyer, 2020) etc. These methods have demonstrated significantly better performance on various classification and regression problems (Kotsiantis et al., 2006). Figure 3 includes schematic representation of some basic machine learning algorithms.

Figure 3

Machine Learning Algorithms: A) KNN, B) Decision Tree, C) Support Vector Machine



Transfer learning and learning federations have also become popular in this period. Transfer learning refers to the use of information learned by a model trained for one task in another, related or different task. This methodology accelerates the training process by using previously learned features or parameters instead of learning from scratch when starting a new learning task and allows for better results with less data. Although the idea was first proposed in the early 1990s, its popularity increased after the 2010s with the development of deep learning systems and the need to learn from large datasets. It is actively used, especially in image processing and natural language processing studies. Pan and Yang (2010) describe basic concepts and methods of transfer learning. Similarly, federated learning, which began to be used towards the end of this period, can be described as a distributed and privacy-focused machine learning method. Data from numerous independent devices or servers (e.g., phones, computers, or organizations) is used locally to train machine learning models without being sent to a central server. The trained models are then combined in a central location, with only parameter updates (e.g., weights) being collected to create a common and more powerful model. Protecting data confidentiality, keeping the data where it originates, and fostering collaboration among the learning models are the primary goals of these systems.

Achieving successful applications of machine learning in real-life increased their impact in academia and industry well-beyond expectations. Predictive models triggered by this technology began to be used in many fields, including medicine (Deo, 2015), finance (Ahmed et al., 2022), marketing (DeTienne & DeTienne, 2017), manufacturing (Wuest et al., 2016), and security (Nassif et al., 2021). The learning capability of computers increased everyday with new methods and models and started to bring about more and more correct results with as much level of truthfulness as possible.

Although this was the start of a new beginning for AI, statistical and previous machine learning methods also had certain limitations (Rudin et al., 2022). Modeling complex relationships, especially in high-dimensional data, was challenging, and performance depended heavily on features designed by experts (LeCun et al., 2015). They were

having difficulty in learning the very complex, multidimensional and nonlinear structure of real-world events, especially in the domains of high-dimensional data, such as images and audio messages processed. Deep learning initiatives were fueled in the late 2000s in order to overcome these limitations. Computers were now equipped with multilayer neural networks capable of automatically extracting features from data and representing complex patterns without requiring human intervention. This transformation ushered in a new era that laid the foundation for today's advancements in artificial intelligence.

In addition to the systems and learning algorithms mentioned above, various other algorithms such as Model-based reasoning (Ifenthaler & Seel, 2013), Case-based learning (Richter & Weber, 2013), Tabu search algorithm (Prajapati et al., 2020), Simulated annealing (Delahaye et al., 2019), Q-learning and Deep Q-learning (Clifton & Laber, 2020).

Another notable development during this period is the field of distributed AI. It involves enabling multiple independent artificial intelligence systems (software agents, robots, computers, etc.) to work interactively. The main goal is to solve problems using multiple independent but collaborative intelligence systems instead of a single central intelligence system. Although this concept emerged in the 1980s, the first application examples began to appear in the 2010s. Developments in the field of machine learning have also increased the effectiveness of these systems. Studies in this direction have manifested themselves as Distributed Problem Solving and Multi-Agent Systems and have become even more widespread in the 2010s. Multi agent systems continue to attract increasing interest. Duan et al. (2023) presented a literature review on the development of distributed artificial intelligence.

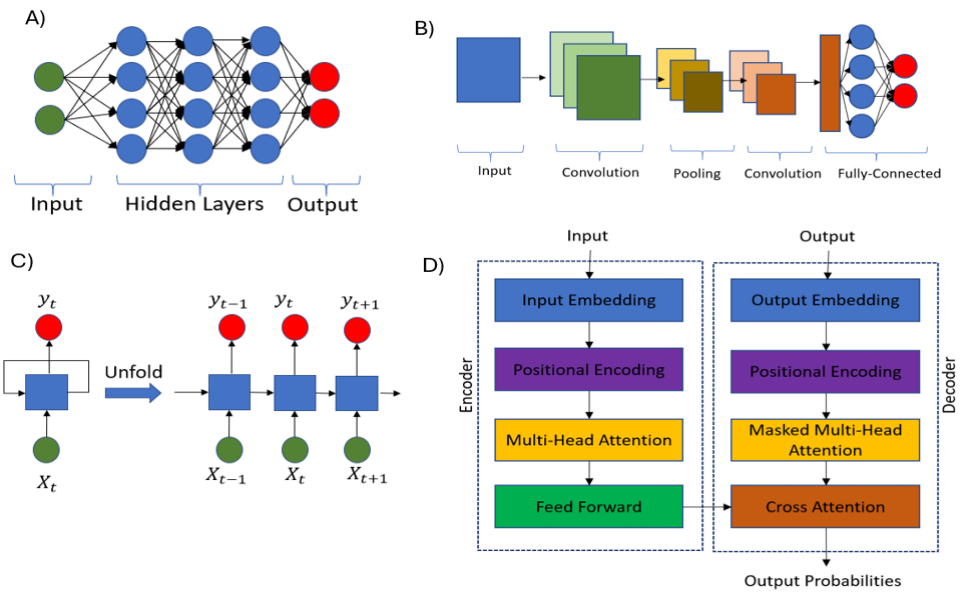
2. Deep Learning and Generative Artificial Intelligence

The breakthrough in artificial intelligence, which began in the 2010s, was driven by the remarkable practical success of deep learning approaches and the evolution of artificial neural network architectures. It was a significant development, demonstrating that many phenomena previously thought impossible to learn can now be understood through this method. It has opened a way to extract underlying patterns from data. Along this line, earlier Neural Networks demonstrated the conceptual foundation of learning from example data, but they lacked the depth, computational power, and optimization techniques necessary to model extremely complex patterns which were inevitable in real life especially in visual recognition. Convolutional Neural Networks (CNNs) revolutionized visual pattern recognition by enabling automatic feature extraction from images, while Recurrent Neural Networks (RNNs) provided powerful mechanisms for modeling temporal dependencies in sequential data such as speech, language, and bio-signals. More recently, Transformer

architecture fundamentally reshaped the ML approaches through introducing an attention mechanism. It became possible to efficiently scale data and outperform respective traditional models. Figure 4 demonstrates architectural structures for various neural network types.

Figure 4

Neural Networks Architectures A) Artificial Neural Network, B) Convolutional Neural Network, C) Recurrent Neural Network, D) Transform Network



The rapid spread and success of the new achievements in learning systems was mainly based on the following.

- the availability of high-performance GPU-based computing power as the deep networks need to be trained by millions of parameters.
- the availability of large-scale datasets in order to increase the learning space for more generality.
- the development of optimization techniques that ensure the stability of deep neural network training.

The first and most significant achievements of this period were achieved with CNN structures (Bhatt et al., 2021). AlexNet's success in the ImageNet competition in 2012 proved that deep learning had revolutionized image processing (Krizhevsky et al., 2012). Subsequently, deep CNN architectures such as GoogleNet (Szegedy et al., 2015) (Szegedy et al., 2015), VGG (Simonyan & Zisserman, 2014), ResNet (He et al., 2016), and EfficientNet (Tan & Le, 2019) were developed. CNNs, thanks to their ability to capture local patterns in images, have become the standard method in object detection (Dhillon & Verma, 2020), segmentation (Guo et al., 2018), face recognition (Oloyede et al., 2020), and many computer vision applications (Zhao et al., 2024).

Furthermore, Recurrent Neural Networks (RNNs), and particularly Long Short-Term Memory (LSTM) structures (Hochreiter & Schmidhuber, 1997) and Gated Recurrent Unit (GRU) (Cho et al., 2014), have played a critical role for many years in processing sequential data such as time series, language and speech (Hewamalage et al., 2021). Studies indicated the RNNs may suffer from exponential gradient decays of the data as they are backpropagated through many time steps, making it extremely difficult for the model to learn long-range dependencies (Bengio et al., 1994). Gated architectures such as LSTMs and GRUs were introduced to overcome this problem. The gating mechanism is designed to preserve information over longer periods. By doing so, it was possible to generate more successful learning systems with respect to previous ones, especially in capturing the temporal dependencies. As a result, they have led to significant advances in fields such as machine translation (Vathsala & Holi, 2020), speech recognition (Graves et al., 2013), and sentiment analysis (Kurniasari & Setyanto, 2020). However, despite their success, these structures still exhibit limitations in modeling very long-term contexts, creating the need for more powerful representation-learning methods.

The transformer architecture which is based on an attention mechanism eliminating the need for repetitive structures to process sequential data as introduced in 2017, opened a new era in artificial intelligence (Vaswani et al., 2017). This approach enables both the processing of long contexts and the scalability of large models due to its inherent parallel training capability and triggered studies over large language models (Zheng et al., 2025).

Generative AI models based on large language models such as GPT (Radford et al., 2018) and BERT (Devlin et al., 2019) were developed alongside Transformer-based models. Gen AI does not only understand language structures but also well-perform in context inference (Dong et al., 2024), text generation (Liang et al., 2024), question answering (Yue, 2025), summarization (Praneeth et al., 2025), and various reasoning tasks (Friha et al., 2024). The GPT series was particularly notable for its flexible architecture, which allowed it to be trained on internet-scale data in an unsupervised or semi-supervised manner and subsequently adapted to specific tasks. BERT, with its bidirectional context understanding, set a new standard in language comprehension tasks.

Another significant development in the deep learning era was the rise of multimodal models. These models are capable of processing multiple data types simultaneously, such as text, image, audio, and video (Baltrušaitis et al., 2018). For example, CLIP enabled zero-sample image classification by modeling image and text relationships in a common space (Radford et al., 2021); models such as DALL-E (Ramesh et al., 2021), Imagen (Saharia et al., 2022), and Stable Diffusion enabled high-quality image generation from text (Rombach et al., 2022). Today, multimodal large models continue to play a critical role in developing AI systems with better cognitive capacity (Bommasani, 2021).

3. Explainable AI (XAI)

Explainable Artificial Intelligence refers to a set of methods and techniques that explains the decisions and inner workings principles of artificial intelligence models. As AI systems-especially deep learning models- are growing increasingly complex, their decision-making processes often resemble "black boxes" preventing transparency. Although these systems generate successful applications, lack of transparency raises some concerns on trust, accountability, fairness, and interpretability. XAI aims to overcome this problem by providing meaningful explanations on how AI models produce their results (Angelov et al., 2021).

XAI focuses on creating models that are either inherently interpretable or can be explainable through specifically designed tools similar to interpretable models such as decision trees, linear models, and rule-based systems but also with capability of well-performing over complex tasks. It is expected that the models, like deep neural networks, are able to explain their reasoning without disrupting the underlying structure. These methods include feature importance scores, saliency maps, attention visualization, counterfactual explanations, and local approximation techniques such as LIME (Ribeiro et al., 2016) and SHAP (Lundberg & Lee, 2017).

A major goal of XAI systems is to ensure trust and transparency. This is important especially in sensitive domains such as medicine (Tjoa & Guan, 2020), finance (Weber et al., 2024), security (Zhang et al., 2022), and autonomous systems (Sakai & Nagai, 2022). When a medical doctor receives an AI-based diagnostic prediction, it is essential for him/her to know how that particular prediction is reached and what factors were actively influencing the prediction. This explanation both supports fact-based decision making and prevents biases or possible errors in the prediction mechanism. In any case, XAI supports compliance with requirements on ethical, fairness and accountability.

Literature reveals that XAI generates explanations which can be categorized into global and local explanations where global explanations describe overall model behavior like how input features generally influence predictions and local explanations focus on individual predictions indicating why the model made a specific decision. Global explanations are useful for developers to understand the logic behind the model. Local explanations, on the other hand, provide case-specific justification. XAI techniques can also be described as model-agnostic or model-specific (Speith, 2022). Model-agnostic methods consider the learning model as a black box and can be applied to any algorithm, making them flexible for diverse applications. Model-specific methods, on the other hand, tries to analyze internal model structure such as gradients, feature maps, or decision rules etc. These are tailored to specific model families like neural networks or tree-based models.

Another important aspect of XAI is debugging and improving capability of the models used. Irrelevant features and overfits can be detected or identified through activation maps or attention scores and then can be improved. For example, in computer vision tasks, Grad-CAM heatmaps highlight regions of an image that influence a classification, while in EEG based models, XAI techniques can reveal which temporal or spatial brain regions were most informative (Selvaraju et al., 2017).

As a final remark, XAI in its current state continues to evolve. It is not mature enough yet. The research along this line focuses on integrating interpretability during model training, designing architectures with built-in transparency, and developing evaluation metrics for explanations. It is a scientific hope and expectation to generate AI systems with enough degree of trustworthiness enabling better integration of human and machine society.

4. Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI)

AGI represents cognitive capability similar to that of a human over a real-life event, effectively indicating the benchmark of human equivalence (Bikkasani, 2025). With this capability, an AGI would possess common sense reasoning to some extent and perform robust transfer learning. That is the ability to apply learned knowledge across different domains. Achieving AGI is the most significant challenge today and can be considered as a fundamental breakthrough in cognitive modeling and multimodal integration of AI research. Buttazzo (2023) claims that the exponential evolution in artificial intelligence over the past 80 years suggests that machines would reach generalized human-level intelligence by around 2030 and continue evolving with the same speed, thereafter, ultimately reaching human cognitive capabilities.

The potential applications of AGI span virtually every domain due to its potential capacity to understand, reason, and adapt across various tasks (Sarikaya, 2025). AGI could revolutionize scientific discovery by autonomously generating assumptions, designing experiments, and accelerating breakthroughs in various fields such as engineering, finance, education, medicine, physics etc. In complex decision-making environments, AGI systems could optimize large-scale infrastructures-including healthcare (Li et al., 2024), energy (Singh & Kaunert, 2024) etc. For example, AGI may enable highly personalized diagnosis and treatment, autonomous surgical systems, and continuous patient monitoring in healthcare. AGI capability integrated with robotics could help generate autonomous systems in households, industry, and hazardous environments (Wang & Sun, 2025). Moreover, AGI has the potential to advance education through adaptive tutoring, enhance creativity by producing novel scientific knowledge, and support space exploration through fully autonomous mission planning and problem-solving systems (Latif et al., 2023). These applications clearly indicate the impact of AGI over social transformation.

The state of the art in this progress is leading to ASI, aiming to fundamentally and vastly exceed human cognitive capacity (Kim et al., 2024). Currently, not much research is carried out and results achieved along this line though it is emerging. Theorists suggest that once an AGI is built, its capacity for recursive self-improvement-rapidly editing its own architecture to become successively smarter-would lead to an exponential, unstoppable increase in intelligence. While ASI holds the utopian promise of solving humanity's most severe global challenges, it simultaneously introduces the ultimate Existential Risk problem which is the challenge of Value Alignment. Ensuring that hyper-intelligence, which could easily outmaneuver all human constraints, has terminal goals permanently aligned with the continued welfare and ethical values of humanity remains the single most critical and unresolved philosophical and technical challenge for the future. Despite this understanding the authors of this chapter believe that human beings are always going to have superintelligence over robots due to the capability of possessing intellect and mind.

5. The Most Meaningful Paradigm Changes AI

Evolution of AI methodologies is characterized by several paradigmatic tiers that fundamentally redefine intelligence and human-machine interaction. The most significant transformation is achieved from hand-coded knowledge to learned knowledge enabling the transformation of AI from an engineered activity to a capability derived from data. This generates a new wave of intelligent system generation as opposed to traditional rationalism as outlined by Newell & Simon (1976). This transformation can also be characterized as the transition from reasoning to prediction, where intelligent systems are aiming to optimize forecasting accuracy rather than generating logical justification. This surely achieves better performance but neglects explainability (Lipton, 2018). Deep learning further improved AI by moving from explicit representations to distributed, sub-symbolic representations. This in turn changed the understanding that knowledge should be conceptualized to become validated (Churchland, 1989). Meanwhile, generative AI shifted the understanding of intelligence from being discriminative to generative models. This makes AI systems more creative, more exploratory, and more reactive (Goodfellow et al., 2014; Brown et al., 2020).

From the technical point of view, these transformations yield substantial changes. That is to shift from tools to agents, where AI systems are becoming more capable of performing autonomously and pursuing goals, forcing reconsideration of responsibility and moral status (Dennett, 1989). However, as Doshi-Velez & Kim, (2017) stated, the capability to explain may be a problem when AI outperforms

humans in some areas while becoming less intelligible. Scientists like Bostrom (2014) believe that the most fundamental rupture is the transition from human-centered to post-human intelligence, where human cognition no longer serves as the upper bound and anthropocentric definitions. This also highlights a paradigmatic shift from optimization to alignment. But as humans are developing the intelligence of AI systems, they will always find a better way of superseding. Despite this it would not be wrong to summarize that the main shift across AI history is and will be based upon from explicit, human-legible intelligence to emergent, autonomous, and increasingly non-human forms of cognition, requiring reorientation from engineering and explanation toward governance, alignment, and coexistence. Table 1 summarizes most meaningful paradigm shifts in AI.

Table 1

Paradigm Shift in AI Methodologies

No.	Paradigm Shift	Previous Approach	Emerging Approach	Philosophical Significance	Ref.
1.	From programing to learning	Knowledge explicitly coded by humans	Knowledge learned from data	Rationalism - Empiricism	Newell & Simon, 1976
2.	From reasoning to prediction	Logical interference	Probabilistic forecasting	Truth – Functional success	Lipton, 2018
3.	From symbolic representation to latent spaces	Human-readable symbols	Distributed latent representation	Conceptual knowledge – Implicit knowledge	Le Cun et al., 2015
4.	From discriminative to generative models	Classification and recognition	Word modelling and generation	Reactive intelligence – Creative intelligence	Goodfellow et al., 2014
5.	From tools to agents	Task-execution systems	Goal-directed autonomous systems	Instrumental reason – Agency	Dennett, 1989
6.	From transparency to post interpretability	Interpretable models	Opaque black-box systems	Epistemic crisis	Doshi-Velez & Kim, 2017
7.	From human-centered to post-human intelligence	Human performance as benchmark	Superhuman cognitive regimes	Anthropocentrism – Post humanism	Bostrom, 2014
8.	From optimization to alignment	Objective maximization	Value compatibility and safety	Technical rationality – Moral rationality	Russell, 2022

6. Capability Transformation of AI Paradigms

Considering the capacity expansion in AI models, improving cognitive competency of the systems takes attention. The systems represent more and more comprehensive capabilities. As stated above, traditional AI established explicit reasoning through symbolic logic. Machine learning added the ability to learn from data on top of this. Similarly, deep learning brought narrow superhuman performance in specific tasks through hierarchical pattern recognition. Generative AI now represents a major breakthrough by introducing creativity, simulation, and abstraction which enable systems to create novel outputs generated over big data. Meanwhile, XAI seems to emerge as a corrective capability, providing transparency and control mechanisms essential for responsible deployment.

The convergence of generative and explainable capabilities points out that the interactive human-AI collaboration where systems can both act and justify their actions would be inevitable. This sets the stage AGI will play a crucial role as they are conceptualized as having the capability of integrating generative reasoning, self-monitoring, adaptive goal management, and continuous learning across domains. Beyond AGI, the main expectation is to reach ASI, which would be expected to exceed human cognitive limits across all intellectual domains while introducing unprecedented governance challenges. Although human intelligence will always be superseding, ASI is expected to reach similar levels of beyond in some areas where humans are lacking the respective knowledge. The most important characteristics that will come forward are autonomy and general intelligence with explainability and alignment bounded by responsibility in social interactions. The capability expansion across AI paradigms is summarized in Table 2.

Table 2

Capability Expansion Across AI Paradigms

AI Paradigm	Learning	Generalization	Creativity	Autonomy	Explainability
Symbolic AI	None	Very Limited	None	Low	High
ML/Deep Learning	High	Task-Specific	None	Low-Medium	Low
Generative AI	High	Medium-High	High	Medium	Low-Medium
Explainable AI	Depends	Depends	None	Medium	High
AGI	Very High	High	High	High	Medium-High
ASI	Self-improving	Extreme	Extreme	Very High	Critical/ Uncertain

7. Philosophical Transformations of AI Paradigms

Similar to transformations explained above, progress of AI research also presents some philosophical transformation, especially in conceptualizing the intelligence. As Newell and Simon, (1976) emphasized early symbolic AI embodied rationalist assumptions that intelligence was formal symbol manipulation governed by logical rules, aligning with computational functionalism. Due to lack of performing deeper and deeper reasoning, machine learning emerged where intelligence became statistical inference from experience rather than explicit knowledge encoding. This was a remarkable transition to achieve competence without transparent justification and challenging epistemology between knowledge and understanding (Bishop & Nasrabadi, 2006; Lipton, 2018). Deep learning over traditional learning system presented explicit meaning by distributing intelligence across high-dimensional latent spaces in alignment with connectionist theories of mind (Churchland, 1989), while generative AI reshaped the intelligence as constructive world-modeling which has the capacity to simulate and generate rather than merely classify or predict (Goodfellow et al., 2014; Brown et al., 2020).

This transition surely degrades traditional philosophy of AI on authorship, generativity, and intentionality while forcing novel conceptualization of what constitutes intelligence. Beyond this, explainable AI illustrates a new philosophical concern of responsibility and accountability as essential criteria beyond raw performance (Doshi-Velez & Kim, 2017). Recent transformation just started to emerge from AGI and ASI represent a qualitative shift from competence to autonomous agency (Dennett, 1989) and ultimately to posthuman epistemology, where cognitive asymmetry may render traditional assumptions about intelligibility obsolete (Bostrom, 2014). Moreover, the main philosophical transition thus evolves from representation and knowledge toward responsibility, alignment as well as governance of intelligence. Table 3 summarizes this philosophical evolution across AI history.

Table 3

Philosophical Evolution Across AI Paradigms

AI Paradigm	Historical Period	Conception of Intelligence	Epistemological Orientation	Philosophy of Mind	Conception of Knowledge	Status of Explanation	Ref.
Symbolic AI	1950-1970	Formal symbol manipulation	Rationalism, logical positivism	Computational functionalism	Explicit, propositional, rule-based	Intrinsic and transparent	Newell & Simon, 1976
Expert Systems	1970-1980	Encoded expert reasoning	Applied rationalism	Strong functionalism	Hand-crafted domain expertise	Buıilt-in justification traces	Buchanan & Shortliffe, 1984
Statistical Machine Learning	1980- 2000	Probabilistic inference	Empiricism, Bayesian epistemology	Instrumentalism	Statistical regularities	Largely absent	Bishop & Nasrabadi, 2006
Deep Learning	2010s	Hierarchical representation learning	Radical empiricism	Connectionism	Distributed representation	Opaque, post-hoc	Le Cun et al., 2015
Generative AI	2020s	Generative world modelling	Pragmatism, constructivism	Emergentist, cognition	Generative, predictive models	Partial, narrative based	Goodfellow et al., 2014
Explainable AI (XAI)	2010s-present	Accountable Intelligence	Social epistemology	Human-centered functionalism	Interpretable abstractions	Normality required	Doshi-Velez & Kim, 2017
Artificial General Intelligence	Future-oriented	General autonomous agency	Integrative epistemology	Philosophy of agency	Cross domain transferable knowledge	Intrinsic self-explanation	Goertzel, (2014).
Artificial Super Intelligence	Speculative	Superhuman intelligence	Post-human epistemology	Post-anthropocentric	Partially non-human comprehensible	Functional intelligibility	Bostrom, 2014

8. Areas Most Affected by Philosophical Changes in AI Paradigms

The philosophical transformations in AI paradigms seem to have impact over several aspects. In epistemology and scientific knowledge production is one of them. Progress in AI development challenges classical norms by generating effective but inexplicable knowledge, shifting from theory-driven to data-driven discovery and from causal explanation to correlational sufficiency. This raises the question of whether knowledge can be legitimate without human interpretable derivation or not (Bishop, 2006). Ethics and moral philosophy, while transition from rule-based frameworks to value alignment, is another area to be affected by this transformation. As AI evolves from neutral tools to intelligent agents, the responsibility and morality are distributed across developers, deployers, and autonomous systems (Floridi et al., 2018; Doshi-Velez & Kim, 2017). Another area which is affected by this transition is legal and governance systems which confront the undermining of foundational assumptions about human intent and responsibility, struggling with algorithmic opacity and the shift from individual liability to systemic accountability. As AI performs respective human tasks with a degree of consciousness, human identity and philosophy of mind are challenged. The difference between human and machine cognitive capabilities are reassessed, taking the main attention towards viewing the mind from symbol processor to generative model (Churchland, 1989; Dennett, 1989).

Another important issue along this line is the ability of AI to redefine work requirements from labor-based value to cognitive capital. This has implications over economic and social philosophy triggering debates on post work societies and redistribution of responsibilities as knowledge processing becomes automated. The overarching transformation is that philosophy has moved from a reflective discipline to an operational necessity-philosophical assumptions now directly shape model objectives, constrain system behavior, and define acceptable futures (Newell & Simon, 1976; Brown et al., 2020). In the age of generative AI, AGI, and ASI, philosophy is no longer about interpreting intelligence but about designing it, as the core question shifts from understanding what kinds of entities exist to governing coexistence with intelligence that may surpass human cognitive capacities across all domains. Table 4 highlights possible areas where paradigm shift in AI has some philosophical impact.

Table 4

Areas Most Affected by AI Paradigm Shift

No.	Affected Area	Nature of Impact	Ref.
1.	Epistemology & Knowledge Production	AI-generated knowledge challenges classical criteria of explanation and justification.	Lipton, 2018; Floridi et al., 2018
2.	Ethics & Moral Philosophy	AI autonomy and value alignment shift responsibility from human to systems / socio-technical networks.	Bostrom, 2014
3.	Law & Governance	Traditional legal frameworks struggle with opaque, autonomous AI decision-making.	Citron & Pasquale, 2014
4.	Human Identity & Philosophy of Mind	AI performance pressures reconsideration of intelligence, personhood and agency.	Dennett, 1989; Churchland, 1989
5.	Economics & Labor	Cognitive automation transforms notions of work, value, and economic agency.	Brynjolfsson & McAfee, 2014
6.	Policy & Social Philosophy	AI governance intersects public values, fairness and social legitimacy.	Mittelstadt, 2019
7.	Ontology & Metaphysic of Intelligence	Expanding definition of intelligence challenge classical metaphysic categories.	Bostrom, 2014
8.	Cognitive Science & Neuroscience	AI models influence theories of cognition and representation.	Lake et al., 2017

9. Conclusion

The evolution of artificial intelligence from traditional rule-based systems to generative models and then to super artificial intelligence represents a series of transformations which are fundamental changes on how intelligence itself is conceptualized, instantiated, and governed. Early AI methodologies were mainly based on symbolic reasoning and processing explicit domain knowledge of respective human expertise, which in turn reflects a rationalist view to understand intelligence as a set of formal rules designed and controlled by humans. Due to technological progress on information technologies and AI research, this paradigm evolved to data-driven and learning-based approaches. It became possible to rise up to generative architecture capable of modeling complex distributions as well as producing novel content in increasingly autonomous modes of action.

This evolutionary progress has fundamentally changed the epistemological foundations of AI. As traditional systems emphasize transparency, logical inference, and human interpretability, current AI paradigms prioritize predictive performance, scalability as well as emergent capabilities. Eventually, the concept and meaning of valid knowledge shifted from explainability and causal understanding toward empirical effectiveness and functional success of data, also changing the respective assessment criteria. However, this raises the question on the credibility of knowledge generated by the machine and validating knowledge with respect to the human judgment process.

Meantime, the progress toward general and super intelligent systems also yields ethical and normative structures. It should be noted that AI is no longer a neutral instrument but an active participant in decision-making processes within different segments of society. It is naturally having an impact over automating human behavior. Concepts such as accountability, responsibility, and moral agency are now becoming part of socio-technical systems distributed all surroundings. The growing emphasis on alignment, safety, and governance reflects an emerging recognition that intelligence without normative grounding poses not only technical risks but also existential and societal challenges.

In this context, philosophical transformation should be based on forward-looking and operational necessity. This will surely provide benefits to the design, deployment, and governance of advanced AI systems. Future research and AI development policy will have to integrate technical innovation with philosophical inquiry, ethical reasoning, and inclusive governance structures. Only through this interdisciplinary engagement can the evolution of artificial intelligence, from its traditional origins to the threshold of super intelligence, be guided in ways that are socially beneficial, ethically sound, and aligned with long-term human prosperity.

References

Ahmed, S., Alshater, M. M., El Ammari, A., & Hammami, H. (2022). Artificial intelligence and machine learning in finance: A bibliometric review. *Research in International Business and Finance, 61,* 101646. https://doi.org/10.1016/j.ribaf.2022.101646

Al-Hassani, S. (2006). *1001 Inventions.* Muslim Heritage in Science, Foundation by Sicence Technology and Civilization.

Al-Hassani, S. T., & Al-Lawati, M. A. (2009). The Six-Cylinder Water Pump of Taqi al-Din: Its Mathematics, Operation and Virtual Design, Musli Heritage, https://muslimheritage.com/six-cylinder-water-pump-taqi-din/ (available on 29.12.2025)

Al-Jazarī, I. A. R. (1973). *The Book of Knowledge of Ingenious Mechanical Devices: (Kitāb Fī Ma'rifat Al-ḥiyal Al-handasiyya).* Springer Netherlands, ISBN: 978-9027703293

Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11*(5), e1424. https://doi.org/10.1002/widm.1424

Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence, 41*(2), 423-443. https://doi.org/10.1109/TPAMI.2018.2798607

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks, 5*(2), 157-166. https://doi.org/10.1109/72.279181

Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., ... & Ghayvat, H. (2021). CNN variants for computer vision: History, architecture, application, challenges and future scope. *Electronics, 10*(20), 2470. https://doi.org/10.3390/electronics10202470

Bikkasani, D. C. (2025). Navigating artificial general intelligence (AGI): Societal implications, ethical considerations, and governance strategies. *AI and Ethics, 5*(3), 2021-2036. https://doi.org/10.1007/s43681-024-00642-z

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: Springer.

Bommasani, R. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258. https://doi.org/10.48550/arXiv.2108.07258

Bowen, J. P. (2017). Alan Turing: founder of computer science. In *School on engineering trustworthy software systems* (pp. 1-15). Cham: Springer International Publishing. ISBN: 978-3-319-56841-6

Bostrom, N. S. (2014). *Superintelligence: Paths, Dangers, Strategies.* Oxford University Press, ISBN 978–0–19–166683–4

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems, 33,* 1877-1901. https://doi.org/10.48550/arXiv.2005.14165

Brantly, A. F. (2020). When everything becomes intelligence: machine learning and the connected world. Gill P., Marrin S., Phythian M., (Ed.). In *Developing Intelligence Theory* (pp. 96-107), Tailor and Francis, ISBN: 9780429028830. https://doi.org/10.1080/02684527.2018.1452555

Brasseur, G. P., (2011), *Support Vector Machines: Theory and Applications,* Springer Link, ISBN: 978-3540807032

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* WW Norton & company.

Buchanan, B. G., & Shortliffe, E. H. (1984). *Rule based expert systems: the mycin experiments of the stanford heuristic programming project (the Addison-Wesley series in artificial intelligence).* Addison-Wesley Longman Publishing Co., Inc.

Buttazzo, G. (2023). Rise of artificial general intelligence: risks and opportunities. *Frontiers in artificial intelligence, 6,* 1226990. https://doi.org/10.3389/frai.2023.1226990

Jijo, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of applied science and technology trends, 2*(01), 20-28. https://doi.org/10.38094/jastt20165

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. https://doi.org/10.48550/arXiv.1406.1078

Churchland, P. S. (1989). *Neurophilosophy: Toward a unified science of the mind-brain.* MIT press. https://doi.org/10.7551/mitpress/4952.001.0001

Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. *Wash. L. Rev., 89,* 1. https://digitalcommons.law.uw.edu/wlr/vol89/iss1/2

Clifton, J., Laber, E., (2020), Q-Learning: Theory and Applications, *Annual Review of Statistics and Its Application, 7,* 279-301. https://doi.org/10.1146/annurev-statistics-031219-041220

Cox E. (1994) *The Fuzzy Systems Handbook: A practitioners Guide to Building, Using, and Maintaining Fuzzy Systems,* Academic Press, ISBN: 0-12-194270-8

Delahaye, D., Chaimatanan, S., Mongeau, M., (2019), Simulated Annealing: From Basics to Applications. In: Gendreau, M., Potvin, JY. (eds) *Handbook of Metaheuristics.* International Series in Operations Research & Management Science, Springer Verlag, 272, https://doi.org/10.1007/978-3-319-91086-4_1

Dennett, D. C. (1989). *The intentional stance.* MIT Press. ISBN: 9780262540537

Deo, R. C. (2015). Machine learning in medicine. *Circulation, 132*(20), 1920-1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186). https://doi.org/10.48550/arXiv.1810.04805

DeTienne, K. B., & DeTienne, D. H. (2017). Neural networks in strategic marketing: exploring the possibilities. Journal of Strategic *Marketing, 25*(4), 289-300. https://doi.org/10.1080/0965254X.2015.1076881

Dhillon, A., & Verma, G. K. (2020). Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence, 9*(2), 85-112. https://doi.org/10.1007/s13748-019-00203-0

Dong, X., Teleki, M., & Caverlee, J. (2024). A survey on LLM inference-time self-improvement. arXiv preprint arXiv:2412.14352. https://doi.org/10.48550/arXiv.2412.14352

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. https://doi.org/10.48550/arXiv.1702.08608

Duan, S., Wang, D., Ren, J. Lyu ,F., Zhang, Y., Wu, H.,Shen, X., (2023) Distributed Artificial Intelligence Empowered by End-Edge-Cloud Computing: *A Survey IEEE Communications Surveys & Tutorials, 25*(1), 591 – 624. https://doi.org/10.1109/COMST.2022.3218527

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative ai. *Business & Information Systems Engineering, 66*(1), 111-126. https://doi.org/10.1007/s12599-023-00834-7

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

Fradkov, A. L. (2020). Early history of machine learning. *IFAC-PapersOnLine, 53*(2), 1385-1390. https://doi.org/10.1016/j.ifacol.2020.12.1888

Friha, O., Ferrag, M. A., Kantarci, B., Cakmak, B., Ozgun, A., & Ghoualmi-Zine, N. (2024). Llm-based edge intelligence: A comprehensive survey on architectures, applications, security and trustworthiness. *IEEE Open Journal of the Communications Society.* https://doi.org/10.1109/OJCOMS.2024.3456549

Goertzel, B. (2014). Artificial general intelligence: Concept, state of the art, and future prospects. *Journal of Artificial General Intelligence, 5*(1), 1. https://doi.org/10.2478/jagi-2014-0001

Goldberg, D.E., (1989), *Genetic Algorithms in Search, Optimization and Machine Learning,* 1. Baskı, Addison-Wesley Professional, ISBN: 978-0201157673

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems, 27.*

Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 6645-6649). IEEE. https://doi.org/10.1109/ICASSP.2013.6638947

Guo, Y., Liu, Y., Georgiou, T., & Lew, M. S. (2018). A review of semantic segmentation using deep neural networks. *International journal of multimedia information retrieval, 7*(2), 87-93. https://doi.org/10.1007/s13735-017-0141-z

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778). Las Vegas, USA. https://doi.org/10.1109/CVPR.2016.90

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting, 37*(1), 388-427. https://doi.org/10.1016/j.ijforecast.2020.06.008

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

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems. An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence.* Ann Arbor, MI: University of Michigan Press.

Ifenthaler, D., Seel, N.M., (2013), Model-based reasoning, *Computers & Education, 64,* 131-142, https://doi.org/10.1016/j.compedu.2012.11.014

Jackson, P. (1998). *Introduction to expert systems,* 3rd edition, Addison Wesley, ISBN: 978-0201876864

Kim, H., Yi, X., Yao, J., Lian, J., Huang, M., Duan, S., ... & Xie, X. (2024). The road to artificial superintelligence: A comprehensive survey of superalignment. arXiv preprint arXiv:2412.16468. https://doi.org/10.48550/arXiv.2412.16468

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review, 26*(3), 159-190. https://doi.org/10.1007/s10462-007-9052-3

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems, 25.*

Kurniasari, L., & Setyanto, A. (2020, February). Sentiment analysis using recurrent neural network. In *Journal of Physics: Conference Series* (Vol. 1471, No. 1, p. 012018). IOP Publishing. https://doi.org/10.1088/1742-6596/1471/1/012018

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences, 40*, e253. https://doi.org/10.1017/S0140525X16001837

Langley, P. (2011). The changing science of machine learning. *Machine learning, 82*(3), 275.

Latif, E., Mai, G., Nyaaba, M., Wu, X., Liu, N., Lu, G., ... & Zhai, X. (2023). Artificial general intelligence (AGI) for education. arXiv preprint arXiv:2304.12479, 1, 1-34. https://doi.org/10.48550/arXiv.2304.12479

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

Li, H., Wang, Y., Luo, S., & Huang, C. (2025). The influence of GenAI on the effectiveness of argumentative writing in higher education: Evidence from a quasi-experimental study in China. *Journal of Asian Public Policy, 18*(2), 405-430. https://doi.org/10.1080/17516234.2024.2363128

Li, X., Zhao, L., Zhang, L., Wu, Z., Liu, Z., Jiang, H., ... & Shen, D. (2024). Artificial general intelligence for medical imaging analysis. *IEEE Reviews in Biomedical Engineering.* https://doi.org/10.1109/RBME.2024.3493775

Liang, X., Wang, H., Wang, Y., Song, S., Yang, J., Niu, S., ... & Li, Z. (2024). Controllable text generation for large language models: A survey. arXiv preprint arXiv:2408.12599. https://doi.org/10.48550/arXiv.2408.12599

Lighthill, J., (1973), Artificial Intelligence: A General Survey. Artificial Intelligence: A paper symposium. UK: Science Research Council, en.wikipedia.org

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue, 16*(3), 31-57.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature machine intelligence, 1*(11), 501-507.

Mustafa, A., Millham, R., & Yang, H. (2025). PRISMA Approach for Assessing Fingerprint Classification Models Based on Artificial Super-Intelligence Techniques. *Engineering Innovations, 14,* 127-138. https://doi.org/10.4028/p-fiJ4gb

Nassif, A. B., Talib, M. A., Nasir, Q., Albadani, H., & Dakalbab, F. M. (2021). Machine learning for cloud security: a systematic review. *IEEE Access, 9,* 20717-20735. https://doi.org/10.1109/ACCESS.2021.3054129

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM, 19*(3), 113-126. https://doi.org/10.1145/1283920.1283930

Nilsson, N. J. (2013). *The quest for artificial intelligence.* Cambridge University Press. ISBN: 9780511819346

Nguyen-Duc, A., Cabrero-Daniel, B., Przybylek, A., Arora, C., Khanna, D., Herda, T., ... & Abrahamsson, P. (2025). Generative artificial intelligence for software engineering-A research agenda. *Software: Practice and Experience, 55*(11), 1806-1843. https://doi.org/10.1002/spe.70005

Oloyede, M. O., Hancke, G. P., & Myburgh, H. C. (2020). A review on face recognition systems: recent approaches and challenges. *Multimedia Tools and Applications, 79*(37), 27891-27922. https://doi.org/10.1007/s11042-020-09261-2

Oztemel, E. (2020). *Artificial neural networks.* PapatyaYayincilik, 4th Edition, Istanbul, 21-22 (in Turkish)

Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In Machine learning (pp. 101-121). Academic Press.

Praneeth, B., Nattem, E. C., Jetti, K., Kavyashree, B. K., Rakshitha, D., Kumar, P. R., & Sreelakshmi, K. (2025). Optimization of Customer Feedback Summarization Using Large Language Models (LLM) and Advanced Retrieval-Augmented Generation. *IEEE Access.* https://doi.org/10.1109/ACCESS.2025.3588337

Prajapati, V.K., Jain, M., Chouhan, L., (2020), "Tabu Search Algorithm (TSA): A Comprehensive Survey, 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), India, 1-8, doi: 10.1109/ICETCE48199.2020.9091743.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. https://www.mikecaptain.com/resources/pdf/GPT-1.pdf

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning (*pp. 8748-8763). PmL https://doi.org/10.48550/arXiv.2103.00020

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... & Sutskever, I. (2021, July). Zero-shot text-to-image generation. In *International conference on machine learning* (pp. 8821-8831). Pmlr. https://doi.org/10.48550/arXiv.2102.12092

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144). https://doi.org/10.1145/2939672.293977

Richter, M.M., Weber, R. O., (2013), *Case-Based Reasoning :A Textbook,* Springer Verlag, ISBN 978-3-642-40166-4, https://doi.org/10.1007/978-3-030-01081-2

Ringvold, T. A., Strand, I., Haakonsen, P., & Strand, K. S. (2024). The AI generative text-to-image creative learning process: An art and design educational perspective. *Design and Technology Education: An International Journal, 29*(2), 359-379. https://doi.org/10.24377/DTEIJ.article2433

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10684-10695). https://doi.org/10.1109/CVPR52688.2022.01042

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistic Surveys, 16,* 1-85. https://doi.org/10.1214/21-SS133

Russell, S. (2022). *Human-Compatible Artificial Intelligence.* Human-like machine intelligence, 1, 3-22. ISBN: 978-0-19-886253-6. https://doi.org/10.1093/oso/9780198862536.002.0003

Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach,* Global Edition 4e. ISBN: 9780134610993

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E. L., ... & Norouzi, M. (2022). Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems, 35,* 36479-36494.

Sakai, T., & Nagai, T. (2022). Explainable autonomous robots: a survey and perspective. *Advanced Robotics, 36(*5-6), 219-238. https://doi.org/10.1080/01691864.2022.2029720

Sarikaya, R. (2025). Path to Artificial General Intelligence: Past, present, and future. *Annual Reviews in Control, 60*, 101021. https://doi.org/10.1016/j.arcontrol.2025.101021

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626). https://doi.org/10.1109/ICCV.2017.74

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

Singh, B., & Kaunert, C. (2024). Dynamic landscape of artificial general intelligence (agi) for advancing renewable energy in urban environments: Synergies with sdg 11-sustainable cities and communities lensing policy and governance. Hajjami S., Kaushik K., Khan I., (Ed.), In *Artificial General Intelligence (AGI) Security: Smart Applications and Sustainable Technologies* (pp. 247-270). Singapore: Springer Nature Singapore. ISBN: 978-981-97-3221-0. https://doi.org/10.1007/978-981-97-3222-7

Speith, T. (2022, June). A review of taxonomies of explainable artificial intelligence (XAI) methods. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency* (pp. 2239-2250). https://doi.org/10.1145/3531146.3534639

Stephens, E. (2023). The mechanical Turk: A short history of 'artificial artificial intelligence'. *Cultural Studies, 37*(1), 65-87. https://doi.org/10.1080/09502386.2022.2042580

Sullivan, W., (2018), *Decision Tree and Random Forest: Machine Learning and Algorithms: The Future Is Here!,* Create Space Independent Publishing Platform, ISBN: 9781986246668

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9). Boston, USA. https://doi.org/10.48550/arXiv.1409.4842

Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR. Long Beach, USA. https://doi.org/10.48550/arXiv.1905.11946

Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE transactions on neural networks and learning systems, 32*(11), 4793-4813. https://doi.org/10.1109/TNNLS.2020.3027314

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.* https://doi.org/10.48550/arXiv.1706.03762

Vathsala, M. K., & Holi, G. (2020). RNN based machine translation and transliteration for Twitter data. International *Journal of Speech Technology, 23*(3), 499-504. https://doi.org/10.1007/s10772-020-09724-9

Wang, Y., & Sun, A. (2025). Toward embodied agi: A review of embodied ai and the road ahead. arXiv preprint arXiv:2505.14235. https://doi.org/10.48550/arXiv.2505.14235

Weber, P., Carl, K. V., & Hinz, O. (2024). Applications of explainable artificial intelligence in finance-a systematic review of finance, information systems, and computer science literature. *Management Review Quarterly, 74(*2), 867-907. https://doi.org/10.1007/s11301-023-00320-0

Wuest, T., Weimer, D., Irgens, C., & Thoben, K. D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production & manufacturing research, 4*(1), 23-45. https://doi.org/10.1080/21693277.2016.1192517

Yenduri, G., Murugan, R., Maddikunta, P. K. R., Bhattacharya, S., Sudheer, D., & Savarala, B. B. (2025). Artificial general intelligence: Advancements, challenges, and future directions in AGI research. *IEEE Access, 13,* pp. 134325-134356, 2025. https://doi.org/10.1109/ACCESS.2025.3592708

Yue, M. (2025). A survey of large language model agents for question answering. arXiv preprint arXiv:2503.19213. https://doi.org/10.48550/arXiv.2503.19213

Zadeh, L. A., (1965), "Fuzzy Sets", *Information and Control 8(*3), 338–353.

Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. Artificial Intelligence *Review, 57*(4), 99. https://doi.org/10.1007/s10462-024-10721-6

Zhang, Z., Al Hamadi, H., Damiani, E., Yeun, C. Y., & Taher, F. (2022). Explainable artificial intelligence applications in cyber security: State-of-the-art in research. *IEEE Access, 10*, 93104-93139. https://doi.org/10.1109/ACCESS.2022.3204051

Zheng, Z., Wang, Y., Huang, Y., Song, S., Yang, M., Tang, B., ... & Li, Z. (2025). Attention heads of large language models. *Patterns, 6*(2). https://doi.org/10.1016/j.patter.2025.101176

About the Authors

Prof. Dr. Ercan OZTEMEL
Marmara University | eoztemel[at]marmara.edu.tr
ORCID: 0000-0001-8488-9991

Pro. Dr. Ercan Öztemel is teaching and doing research on artificial intelligence, intelligent Manufacturing systems, management information systems and business intelligence, Simulation and modelling, strategic management or similar areas. He carried out projects in a wide range of areas including military, health, public services, manufacturing, transportation, education etc. He served as the scientific panel member of NATO SAS Panel (from 2001 to 2006), and was Steering Committee Member of Western European Armament Group, Research Cell CEPA 11 (1997-2005) and CEPA 15 (2002-2005) and was also appointed as the member of decision board of IT Research Institute of TUBİTAK (1997-2001) and was the member of the steering committee for Public Research Group of TUBITAK (2012-2017). He was also Deputy chairmen of Turkish Measurement and Selection Center (2011-20215). He is the author and co-editor of more than 200 publications including books, papers, conference proceedings etc.

Dr. Muhammed Esad OZTEMEL
TÜBİTAK | muhammed.oztemel[at]tubitak.gov.tr
ORCID: 0000-0001-6597-8083

Dr. Muhammed Esad Öztemel is a senior researcher at the TÜBİTAK BİLGEM Artificial Intelligence Institute. He holds a B.Sc. degree in Electronics Engineering from Gebze Technical University, an M.Sc. in Computer Science from Southern University and A&M College (USA), and a Ph.D. in Electrical Engineering from Louisiana State University (USA). Following his undergraduate studies, he contributed to significant defense industry initiatives, specifically working on the degaussing project for the MILGEM warship. His research focuses on machine learning and deep learning applications across various domains, computer vision, including network security, microorganism detection, and EEG brain signal classification.