



REVIEW

Fundamentals of artificial intelligence

Rafael Guzmán Robles

Airbus Spain, Madrid, Spain

Received 9 January 2025; accepted 10 March 2025

Available online 8 April 2025



KEYWORDS

Artificial intelligence;
Machine learning;
Deep learning;
Reinforcement
learning;
Generative AI

Abstract Artificial Intelligence (AI) is revolutionizing various industries, with healthcare being one of the most impacted sectors. This article explores the fundamentals of AI, with a specific focus on machine learning, deep learning, and generative AI. Machine learning, a subset of AI, enables systems to identify patterns in large data sets, improving over time without being explicitly programmed. Deep learning, a more advanced subfield, uses multi-layered neural networks to process complex information. The advent of generative AI, such as GPT and GANs, has expanded the potential of AI to create new content autonomously, transforming areas like drug discovery and personalized medicine. The article also addresses the ethical considerations surrounding the use of AI, particularly concerning data privacy, algorithmic bias, and equitable access to AI-driven technologies. These considerations are essential for ensuring the responsible development and implementation of AI in healthcare.

© 2025 SESPM. Published by Elsevier España, S.L.U. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

PALABRAS CLAVE

Inteligencia artificial;
Aprendizaje
automático;
Aprendizaje profundo;
Aprendizaje por
refuerzo;
IA generativa

Fundamentos de inteligencia artificial

Resumen La Inteligencia Artificial (IA) está revolucionando varias industrias, siendo la sanidad uno de los sectores más afectados. Este artículo explora los fundamentos de la IA, centrándose específicamente en el aprendizaje automático, el aprendizaje profundo y la IA generativa. El aprendizaje automático, un subconjunto de la IA, permite a los sistemas identificar patrones en grandes conjuntos de datos, mejorando con el tiempo sin ser programados explícitamente. El aprendizaje profundo, un subcampo más avanzado, utiliza redes neuronales multicapa para procesar información compleja. La llegada de la IA generativa, como las GPT y las GAN, ha ampliado el potencial de la IA para crear nuevos contenidos de forma autónoma, transformando ámbitos como el descubrimiento de fármacos y la medicina personalizada. El artículo también aborda las consideraciones éticas que rodean el uso de la IA, en particular las relativas a la privacidad de los datos, el sesgo algorítmico y el acceso equitativo a las tecnologías impulsadas

E-mail address: rafael.guzman@airbus.com.

<https://doi.org/10.1016/j.senol.2025.100685>

0214-1582/© 2025 SESPM. Published by Elsevier España, S.L.U. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

por la IA. Estas consideraciones son esenciales para garantizar el desarrollo y la aplicación responsables de la IA en la atención sanitaria.

© 2025 SESPM. Publicado por Elsevier España, S.L.U. Se reservan todos los derechos, incluidos los de minería de texto y datos, entrenamiento de IA y tecnologías similares.

Artificial intelligence

Medical expertise is acquired through exposure to clinical cases, thanks to which a physician refines his or her ability to diagnose and treat disease. Just like humans acquire knowledge and skills through practice, machines also “learn” from data and experience by identifying patterns and improving their predictive accuracy.

The gap between human Intelligence and machine intelligence is closing, thanks to technological breakthroughs occurring in the field of computer science known as Artificial Intelligence (AI). It focuses on creating systems to perform tasks that would normally require human intelligence, such as perception, learning and reasoning.

Artificial Intelligence began in the 1950s, where the first specific programming languages such as LISP were developed. From 1970, we saw the rise of systems capable of emulating human expertise in what is considered the second wave. The third wave took place between 2000 to 2020, with the explosion of machine learning and deep learning, enabled by the increase in computational power and the availability of large volumes of data. Today, we are experiencing the fourth wave, with generative AI, where Large Language Models (LLMs) such as GPT are giving AI the power and “creativity” to generate high-quality text, images, audio and other types of data (Fig. 1).

In this article, we will explore some of the subfields of AI, as illustrated above. Each of these fields has been a major

breakthrough with its technology affecting our everyday lives. This trend will continue to accelerate in the future, there is no going back to a world without AI.

Machine learning

The first subfield of artificial intelligence that appeared approximately 20 years ago is Machine Learning. As its name states, the algorithms driving the machines are no longer deterministically programmed with fixed rules, instead, the machines are incentivized to learn from data samples and improve their performance over time. Machine learning models are excellent at identifying patterns in large volumes of data and generalizing and extracting knowledge from them.

Machine Learning models need to be “trained” on data. The training process of a Machine Learning model consists of feeding it with a set of input data paired with its labels, which are the correct answers so that the machine can learn from these examples. Throughout the training process, the model adjusts its internal parameters to minimize the delta between what it predicts and the results it was given to compare within these labels. The goal of training is for the model not only to memorize the specific examples from the training data set but also to learn general principles that will allow it to adapt to similar situations in the future. Once the model has been trained and evaluated on a sufficient quantity of data, it can be asked to make predictions or decisions autonomously on new data it has never yet been exposed to. This continuous cycle of learning and adaptation is what allows machine learning models to improve their performance over time.

There are three main categories inside Machine Learning:

- **Supervised Learning:** as explained, the model is trained on labeled data. It is often used to predict or classify an output. A model could be given a dataset containing thousands of chest X-rays labeled with the positive or negative diagnosis of pneumonia. The algorithm then learns to identify patterns in the images that distinguish lungs affected by pneumonia from those that are healthy.
- **Unsupervised Learning:** this approach is used when the data is not labeled. There is no output to predict or classify. Instead, these algorithms attempt to identify patterns or hidden structures in the data without any prior instruction on what to look for. An example would be to provide a large dataset with patient data and ask it to classify the individuals according to their genetic characteristics.
- **Reinforcement Learning:** This method teaches algorithms to make decisions through experimentation and feedback. The model is rewarded every time a task is correctly

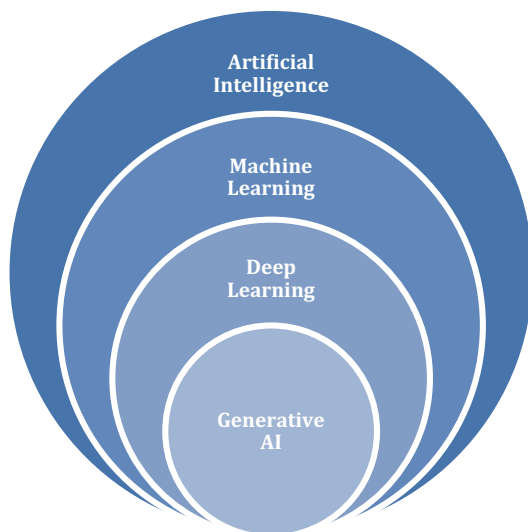


Figure 1 Hierarchical relationship of technologies in artificial intelligence.

performed, therefore inciting it to repeat the behavior and maximize its gain. The key components of reinforcement learning are the following:

- **Agent:** the apprentice who interacts with the environment.
- **Environment:** everything surrounding the agent and with which it interacts.
- **Actions:** the decisions taken by the agent.
- **Rewards:** the feedback the agent receives after performing an action, indicating if it succeeded or failed.
- **Policy:** the strategy the agent applies to make decisions based on the feedback received.

For example, a robot could be given the goal to find the exit of a maze full of obstacles. The robot (agent) moves in different directions (actions) within the maze (environment). Each time the robot avoids an obstacle and gets closer to the exit, it receives a positive reward. If it collides with an obstacle or moves away from the exit, it receives negative feedback. Over time, the robot constructs a policy that allows it to move efficiently around the room and reach its goal more frequently.¹

Deep learning

A subset inside Machine Learning is Deep learning which uses deep Artificial Neural Networks. These networks consist of multiple layers of artificial neurons allowing the model to understand and analyze the data at different levels of complexity (Fig. 2).

In image recognition, for example, a deep neural network processes information progressively through its multiple layers. In the first layers, basic features such as edges and contours are detected, identifying horizontal, vertical or diagonal lines in the image. These lines are then combined in the following layers to identify simple shapes such as circles, squares or corners. As the information moves through the network, these simple shapes are integrated to recognize more complex features, such as parts of an object or a pair of eyes. Finally, in the upper layers, the neural network is able to combine all these features to recognize the complete object and thus identify the image as a "human face".

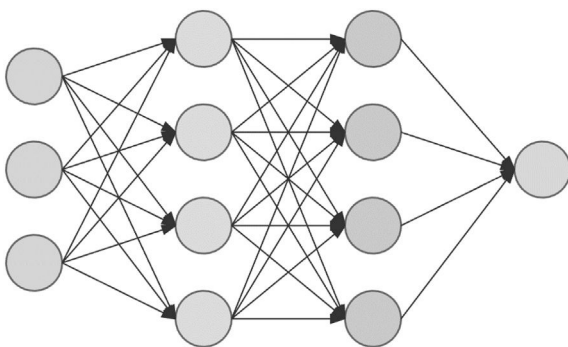


Figure 2 Four-layer neural network.

Artificial neural networks (ANNs)

Artificial neural networks (ANNs) attempt to reproduce the structure and function of the human brain where neurons are interconnected and communicate with each other through synapses, processing information in a distributed and parallel manner. Similarly, an artificial neural network is composed of interconnected layered nodes (artificial neurons). Each node performs simple computations and transmits the information to the nodes of the next layer.

Therefore, as biological neurons have three main parts: the dendrites, the cell body (soma), and the axon, artificial neurons emulate this with inputs (each with its associated weight) an activation function, and an output. Dendrites receive signals from other neurons, for the soma to process and then finally the axon transmits the processed signal to other neurons through the axon terminals. Analogously in an artificial neuron, the weights and activation function process the inputs, analogously to how the soma processes signals in a biological neuron. Then, the activation function determines whether the neuron should "fire", thus controlling the flow of information. Finally, the output transmits the signal to other neurons in the network acting like an axon.

Convolutional neural networks (CNNs)

Convolutional Neural Networks (CNNs) are another type of neural network specifically designed to process and analyze visual data, such as images and videos. They have proven to be highly effective for computer vision and image processing. They automatically detect important features in images, such as edges, shapes and textures, which make them especially effective in image recognition and classification tasks. Their structure is composed of two main types of layers that work sequentially: "convolution" layers and "pooling" layers. The convolution layers act as filters to highlight specific patterns in the image, and the pooling layers reduce the amount of data to be processed, discarding redundant information and keeping only the most relevant.

An example of the use of these convolutional neural networks is the classification of different types of skin wounds. A CNN can be trained to differentiate between benign and malignant images, therefore helping dermatologists to accurately identify possible cases of skin cancer or diseases. This facilitates clinical decision-making and improves diagnosis.

Generative AI (Generative AI)

The latest AI revolution is Generative AI, which is a sub-discipline of Deep Learning. Generative AI represents a significant breakthrough in the field of artificial intelligence, as it is not only limited to analyzing and classifying existing data, but has the ability to create new and original content from learned patterns.

This technology has exploded thanks to developments in advanced neural network models, such as generative antagonistic networks (GANs) and language models such as GPT (Generative Pre-trained Transformer). Models are trained on huge amounts of data and learn to generate content that mimics the style and characteristics of the information on which they were trained.

The development of these models and their application in mass-use tools such as ChatGPT or Midjourney has highlighted the impressive progress of generative artificial intelligence and its ability to create content autonomously. ChatGPT is a conversational assistant based on the GPT language models developed by OpenAI that uses advanced neural networks to generate answers. It responds to simple instructions, known as prompts, with coherent and contextualized text, thus demonstrating its ability to understand and reproduce human interactions. Platforms such as Midjourney, which use neural networks to generate images based on textual descriptions, open up a world of possibilities in visual creation from artistic illustrations to complex designs.

These developments not only demonstrate the power of generative AI in specific tasks, such as text or image generation, but also pave the way towards the integration of “multimodal” models which combine different types of data, such as text, images and audio, enabling richer and more versatile interaction. Multimodal models will transform the way we interact with technology, enabling more immersive and contextual experiences across multiple applications.

Generative pre-trained transformers (GTPs)

GTPs are a type of AI model designed to process and generate human-like text and have become central to the progress of generative AI. They are built on a deep learning architecture called a Transformer,² which allows them to effectively understand and produce natural language. Unlike traditional models that require labeled data for each specific task, GTPs are pre-trained on vast amounts of unstructured text data, enabling them to learn the nuances of language, including grammar, context, and even some level of reasoning. This pre-training phase is followed by fine-tuning, where the model is adjusted for specific applications such as translation, summarization, or conversation.

GTPs can generate coherent and contextually appropriate responses based on a given prompt, making them highly versatile and effective for a range of natural language processing tasks. Their ability to generalize knowledge across different domains has made them a foundational technology for various applications, from conversational AI and content creation to coding assistance and beyond.

Generative antagonistic networks (GANs)

Generative antagonistic networks (GANs) are one of the most advanced and widely used techniques in generative AI. GANs consist of two neural networks: a generator and a discriminator. These two components compete with each other in a process that iteratively improves the quality of the generated data.

- **Generator:** creates synthetic data from a random input. For example, it could generate images of human faces that mimic the characteristics of real faces.
- **Discriminator:** evaluates the data generated by the generator and compares it with real data, determining whether it is true or false.

Feedback from the discriminator helps the generator to improve the quality of its data, by allowing it to adjust its parameters to create more realistic images. This cycle repeats until the generator produces images that the discriminator cannot distinguish from the real ones.

The training process of a GAN can be summarized as follows⁴:

1. The generator creates a synthetic image.
2. The discriminator evaluates whether the image is real or generated.
3. The discriminator provides feedback to the generator.
4. The generator adjusts its parameters to create more realistic images.
5. This cycle repeats until the generator produces images that the discriminator cannot distinguish from the real ones.

Artificial intelligence application areas

Artificial intelligence techniques apply to a large variety of fields, enabling machines to perform complex tasks that previously required human intervention.

Natural Language Processing (NLP) is a branch of AI that focuses on the interaction between computers and human language. The goal is for machines to understand, interpret and respond to language effectively. A very common application is sentiment analysis, where NLP models evaluate text to identify its emotional tone, whether an opinion is positive or negative for example. Analyzing large volumes of text to extract relevant data, such as identifying medical diagnoses in electronic clinical records is another common use of this technology. Machines are also nowadays widely used to translate text from one language to another and facilitate global communication in real time.

In the field of Computer Vision, tasks such as object detection or facial recognition can be performed by machines. In medicine, these applications are particularly useful. Computer vision models can identify and locate specific objects within a medical image. They can also classify images into different categories, differentiating between healthy and unhealthy tissues. Image segmentation, where an image is divided into different areas such as the outline of an organ, facilitates detailed and accurate analysis of medical imagery.

The many applications of artificial intelligence extend to areas such as speech recognition, where AI enables machines to transcribe and understand human speech, facilitating interaction through virtual assistants and dictation systems. Robotic Process Automation (RPA) is another field where repetitive, rule-based tasks traditionally require human intervention are automated. Expert systems apply AI to medical diagnosis or financial advice to surpass the human expert decision-making process in these areas.

Ethical considerations in the use of AI

The use of artificial intelligence (AI) raises several ethical challenges that must be addressed to ensure fair, safe, and effective implementation. One of the main challenges is the protection of privacy and security of data. AI systems require large volumes of data for training and operation, and it is

important to make sure this information is handled in compliance with data protection regulations, such as the General Data Protection Regulation (GDPR) in Europe. It is essential to ensure that data are anonymized and encrypted to prevent unauthorized access and protect individuals' privacy. Furthermore, the AI Act (European Union) establishes a framework for the development and use of AI, with the aim of regulating its application and ensuring that the fundamental rights of European citizens are respected. Additionally, the technological infrastructure must have robust cybersecurity measures in place to prevent cyberattacks that could compromise the integrity and confidentiality of information during storage and processing.⁶

Another critical issue is bias and fairness in the use of AI. Algorithms can perpetuate or even amplify biases already present in the training data, which could lead to unfair or inaccurate decisions for certain demographic groups. Moreover, it is crucial to ensure that access to AI-based technologies is equitable and does not accentuate inequities. The implementation of these technologies must be fair, ensuring that the AI benefits all individuals, regardless of their socioeconomic background or geographic location. Likewise, transparency in the operation and a clear definition of responsibilities in the event of failures in the system are essential to build public trust.

Conclusion

Artificial intelligence (AI) has established itself as one of the most innovative and transformative technologies of the modern era, with applications spanning across all sectors. Its ability to automate complex processes, generate content autonomously and analyze large volumes of data in real time is revolutionizing our world, facilitating new ways of tackling complex problems and optimizing decision-making.

In medicine, AI offers enormous potential for applications such as the accuracy of diagnostics, the personalization of treatments and the optimization of clinical data management. To achieve the successful integration into mainstream medical practice, it is essential to ensure transparency in the operation of the models and to promote their use in an equitable manner, ensuring that technological advances translate into concrete improvements in the quality of healthcare and patient health outcomes.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT and Google Gemini in order to translate, structure the contents and generate ideas. 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 publication.

Funding

None.

Ethical disclosures

None.

Conflicts of interest

None.

References

1. Silver D, Huang A, Maddison C, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. *Nature*. 2016;529:484–9. Available from: <https://doi.org/10.1038/nature16961>.
2. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Proces Syst*. 2017;30:5998–6008. Available from: <https://doi.org/10.48550/arXiv.1706.03762>.
3. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. *Adv Neural Inf Proces Syst*. 2014;3. <https://doi.org/10.1145/3422622>.
4. European Commission. Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts COM/2021/206 final. 2021. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>.