Introduction to Artificial Intelligence: Concepts, Capabilities, and Fundamentals

Hertzian

Introduction to AI

Artificial Intelligence (AI) refers to the development of computer systems that can perform tasks that typically require human intelligence. These tasks include learning, reasoning, problem solving, perception, and language understanding. The goal of AI is to create systems that can process information, make decisions, and improve over time without explicit human intervention.

While the term "AI" is often associated with human-like robots, the reality is much broader. AI powers everything from voice assistants (like Siri and Alexa) to recommendation systems on Netflix and self-driving cars. AI exists on a spectrum - from simple rule-based systems to advanced learning models that can analyse vast amounts of data.

The Evolution of AI

AI has gone through cycles of excitement and disappointment, shaping its path to where it is today. These cycles shaped the development of the field:

Early AI (1950s - 1970s): AI began with the creation of the computer. In the 1950s and 60s, researchers created the first early AI programs that could play games like chess or solve basic problems. They used a brain-inspired called system the Perceptron to have machines make decisions. It was an early major milestone. At the time, it felt like the future had arrived.

But the hardware simply wasn't up to the task. These systems worked in tightly controlled settings, but couldn't handle the messiness of the real world. As expectations outpaced what AI could deliver, funding dried up and enthusiasm cooled. By the late 1970s, the first "AI winter" had set in and the field lost momentum and support.

Second Wave (1980s-1990s): In the 1980s, the rise of development in Expert systems reignited excitement. These were computer programs designed to replicate the knowledge and decision-making skills of human specialists.

For a while, expert systems looked like the answer. Businesses and governments began using them to solve real-world problems, from diagnosing medical conditions to managing factory operations. But there was a catch: these systems were expensive to build, hard to update, and often struggled outside of very specific tasks.

Worse, they could not learn from new information or adapt over time—they were more like static encyclopaedias than curious students. Once again, promises failed to match performance. As doubts crept in and funding pulled back, the field slipped into its second AI winter.

Modern Revival (2000s-present): In the 2000s, AI began to find its feet again, this time with much more solid ground to stand on. A combination of faster computers, massive amounts of digital data, and breakthroughs in algorithms brought fresh energy to the field.

The biggest game-changer was deep learning, which allowed machines to learn from experience, rather like the human brain does. Instead of telling a computer every step to follow, you could feed it examples and it would figure out patterns on its own.

This led to huge leaps in things like image recognition, voice assistants, and language translation.

Now, AI has embedded itself within everyday life, and the dreams of those early researchers have finally been realised.

AI Subfields

Artificial Intelligence is a broad field with many specialised areas that help machines perform tasks that usually require human intelligence. In this section we will break down some of the key subfields of AI:

Natural Language Processing (NLP)

Natural Language Processing is the area of study to get computers to understand and communicate using human language. This is what powers chatbots, voice assistants like Siri and Alexa, and even translation tools like Google Translate. It allows AI to read, summarise, evaluate, and even generate text in a way that sounds human.

Computer Vision

Computer Vision enables AI to see and interpret images and videos. It's used in facial recognition, self-driving cars (to recognize road signs and obstacles), and even in medical imaging to detect diseases from X-rays and MRIs.

Robotics

This is where AI meets physical machines. Robots equipped with AI can perform complex tasks, from assembling cars in factories to performing surgeries with high precision. Some robots, like those in warehouses, can move items around on their own using AI-powered navigation.

Speech Recognition

Speech recognition allows AI to listen and understand spoken words. This is what enables voice commands in smartphones, smart speakers, and even dictation software that turns speech into text. It's a key part of making AI interact with humans more naturally.

How AI Learns

Artificial Intelligence learns through a process that mimics human learning but relies on vast amounts of data and computational power. This learning process can be broadly categorised into three main types: supervised learning, unsupervised learning, and reinforcement learning.

Supervised Learning

Supervised learning involves training an AI system using labelled data. This means that the algorithm is provided with the expected output given a certain input, and it learns to map inputs to the correct outputs based on patterns in the data. For example, in image recognition, an AI might be trained with thousands of pictures of cats and dogs, each labelled accordingly so that it can correctly classify new, previously unseen images.

Due to the data needing correct labels, the cost for collecting or creating these datasets can often be extensive as the data needs to be of a high quality and high volume. It is with supervised learning that the phrase "garbage in, garbage out" originates, as a low-quality dataset will lead to a bad model.

Unsupervised Learning

In contrast to supervised learning, unsupervised learning involves data that is not labelled. The AI analyses large sets of data to find hidden patterns, correlations, or groupings without explicit guidance. This approach is often used in market segmentation, anomaly detection, and recommendation systems. For instance, an AI model might group customers with similar purchasing habits without being explicitly told which categories exist.

The costs for unsupervised learning are significantly lowered compared to a supervised learning model, however, its uses are not as extensive as they are restricted to analysing existing data.

Reinforcement Learning

Reinforcement learning is inspired by behavioural psychology and involves an AI agent learning through trial and error. The system interacts with an environment and receives rewards or penalties based on its actions. Over time, it adjusts its strategy to maximise rewards. This method is commonly used in robotics, gaming, and autonomous systems, such as self-driving cars learning to navigate safely, or a mouse bot navigating a maze.

Common AI Models and Algorithms

Linear Regression

Linear regression is one of the simplest AI algorithms. It tries to find a straight-line relationship between two or more variables. Imagine you're plotting points on a graph — house size on one axis and price on the other — and you're trying to draw a line that best fits all those dots. The algorithm works out the slope and position of that line so it can make predictions for new data. It is best when you believe there is a fairly straightforward relationship between inputs and outputs.

Linear Regression is great for predicting continuous values. For example, you might use linear regression to predict house prices based on square footage, number of bedrooms, and distance to the city centre.

Decision Trees

Decision trees mimic the way we often make decisions. They break down a problem by asking yes/no questions at each stage. For instance, "Is the applicant's income over £30,000?" If yes, go down one path; if no, go down another. The tree continues to branch based on further questions until it arrives at a decision. The algorithm figures out the best questions to ask by looking at the data and picking the ones that split the groups most cleanly.

They are used a lot in finance, such as deciding whether to approve a credit card application. Based on factors like income, employment status, and credit score, the tree helps automate the decision. They're easy to understand and explain to non-technical teams, which makes them popular in business settings.

Support Vector Machines (SVM)

SVMs draw boundaries between different categories in your data. The model tries to place dividing lines between groups in such a way that it's as far away from each group as possible. This helps reduce the chance of making a mistake.

Picture it as if you're trying to separate animals with fences. With an SVM you want to place a fence halfway between two groups of animals and ensure the fence is far from any outliers, to best separate them for when you get a new animal to add.

SVMs are often used for things like email spam detection. The algorithm looks at examples of spam and non-spam emails and finds the best way to tell them apart. Once trained, it can classify new emails quickly and accurately, even if they contain words it's never seen before.

K-Means Clustering

An example of unsupervised learning, K-means clustering groups data into a set number of clusters based on similarity. This is widely used in marketing, for customer segmentation. If you're a retailer, you might use k-means to find distinct groups of customers based on their shopping habits — such as frequent buyers, bargain hunters, or seasonal shoppers — and then tailor your offers to each group.

Neural Networks

A key algorithm to be aware of is Neural Networks. One of the most advanced forms of AI today, Neural networks are inspired by how the human brain works, though in a much simpler form. At their core, they consist of layers of individually simple 'neurons'. Each neuron receives one or more inputs, does some basic calculations, turns a switch on or off, and sends the result to the next layer.

There are three main types of layers:

Input layer: where the data enters. This can take many forms, some examples could be pixels of an image, or numbers in a spreadsheet.

Hidden layers: where the actual learning and decisions happen. These layers process and transform the information using weighted connections.

Output layer: where the result comes out - for instance "yes, this is a cat" or "no, it's not."

The network learns by adjusting the strength (or weight) of the connections between neurons. Over many repetitions, neural networks learn to make very accurate predictions.

Neural networks are the engine behind many of the AI tools we see today. They are capable of performing a large number of tasks that are not feasible through other algorithms. One major use is in image recognition. For example, Facebook's photo tagging or Apple's Face ID relies on convolutional neural networks (a specialised type of neural network) to scan your face and identify unique features like the shape of your eyes or nose.

They're also used in:

- Speech recognition (like Siri or Alexa)
- Language translation (like Google Translate)
- Playing games (like DeepMind's AlphaGo)
- Generating content (like ChatGPT, or Deepseek)

Case Studies: Early AI Adoption

When artificial intelligence first began making its way into mainstream business, it was often trialled in very specific, carefully controlled environments. The industries that led the way did so out of necessity, curiosity, or a desire for competitive advantage. Below are some early case studies that highlight how AI was initially implemented, why it worked, and how it developed over time.

Manufacturing: Predictive Maintenance

General Electric (GE) was one of the earliest adopters of AI in manufacturing, particularly through predictive maintenance. Using AI-powered sensors and data analytics, GE could anticipate when a machine part was likely to fail, allowing for repairs before costly breakdowns occurred.

Why it worked: It solved a clear, costly problem and delivered an immediate return on investment by reducing downtime of machines and potential costs.

How it evolved: What began as predictive maintenance grew into a broader strategy called the "Industrial Internet of Things," where connected machines continuously optimise performance with real-time AI analysis.

Retail: Personalised Recommendations at Amazon

Amazon introduced AI to enhance customer experience through personalised recommendations. By analysing browsing history, past purchases, and customer behaviour, AI systems suggested products tailored to individual users.

Why it worked: It directly influenced customer spending by making online shopping feel more personal and convenient.

How it evolved: Recommendation systems became more complex, using deep learning models and cross-platform data to predict not just what customers want, but when they want it.

Healthcare: Diagnosing Eye Disease

In a pioneering partnership, DeepMind (a UK-based AI company) and Moorfields Eye Hospital applied AI to diagnose eye diseases from retinal scans. The AI system could detect conditions like age-related macular degeneration with remarkable accuracy.

Why it worked: It addressed a major bottleneck in diagnostics, speeding up analysis and reducing errors.

How it evolved: The collaboration expanded into broader diagnostics, setting a precedent for AI-assisted medicine that balances innovation with patient safety and ethical oversight.

Finance: Fraud Detection at MasterCard

MasterCard used AI to detect fraudulent transactions by identifying unusual patterns in spending behaviour in real time. Unlike traditional rule-based systems, AI could adapt and learn from new types of fraud.

Why it worked: Financial fraud is a moving target, and AI's ability to learn and adapt made it a powerful solution.

How it evolved: AI-driven fraud detection became the norm across the financial sector, often integrated with biometric security and customer verification tools.

Agriculture: Crop Monitoring with John Deere

John Deere incorporated AI into its farming equipment to monitor crop health and soil conditions. Using sensors and image recognition, machines could detect issues like pest infestations or nutrient deficiencies.

Why it worked: It gave farmers actionable data, allowing for targeted treatments and better yields.

How it evolved: These tools are now part of precision agriculture, where AI coordinates planting, watering, and harvesting to optimise every square metre of farmland.

Ethics of AI

As Artificial Intelligence becomes more capable and far-reaching, so are the ethical dilemmas it raises. While AI systems can enhance lives and unlock innovation, they also present serious questions about fairness, responsibility, and long-term impact. Ethical AI means building and using technology in ways that are respectful, transparent, and accountable.

Algorithmic Bias

AI models learn from data. If that data reflects past discrimination—intentionally or not - the model may replicate or even amplify it. For example, a facial recognition system trained mostly on lighter skin tones may struggle to recognise people with darker skin. The danger here is not just technical failure, but social harm, particularly when used in policing, hiring, or healthcare.

Where does bias come from?: Bias in AI can come from several different places. As we have covered, AI learns from data - and if that data reflects existing inequalities, the AI will learn those too. For example, if a hiring algorithm is trained on a dataset of past employees that's mostly male, it may start to prefer male applicants, even if unintentionally.

The choices made by developers - what data to use, what problem to solve, and how to evaluate success - can all introduce bias. Even well-meaning teams can embed their own assumptions into systems without realising it.

Transparency and Explainability

AI decisions aren't always easy to unpack. Some models, especially deep learning systems, work in ways that are complex and impossible to understand, even to the people who built them. This creates challenges in critical areas where we need to understand the reasoning behind decisions, such as courtrooms, insurance assessments, or medical diagnostics.

Consent and Data Ownership

A growing concern in the AI space is the use of data without proper consent—especially in creative domains like text, images, and music. Generative AI models (like ChatGPT or image generators like Midjourney and Stable Diffusion) are often trained on massive datasets scraped from the internet. These datasets may include copyrighted material, private content, or personal artwork uploaded without the creators' knowledge or permission.

Artists, writers, and musicians have raised concerns about their work being used to train AI systems that then mimic their style, sometimes in direct competition with them. This raises uncomfortable questions: Who owns the output of an AI trained on someone else's work? Should creators be compensated? Should they have the right to opt-out?

The current legal landscape is murky, and many feel the ethical lines are being crossed even if the laws haven't caught up yet.

Environmental Impact

Behind the sleek interface of an AI tool lies a hidden cost: energy consumption. Training large AI models, especially in deep learning, requires huge amounts of computing power. Data centres run around the clock, consuming vast amounts of electricity and water to keep servers cool.

Training a large AI model can use vast amounts of electricity. For example, training a cutting-edge language model like GPT or a large image generator may take weeks on hundreds of specialised processors (called GPUs), running continuously. A 2019 study found that training one such model could produce as much carbon dioxide as five round-trip flights between London and New York.

And that is just the training phase. Once deployed, these models continue to consume a significant amount of energy every time someone uses them. With millions of queries a day, the environmental toll adds up fast.

As AI becomes more powerful and popular, we face a choice: scale up without thinking or grow more thoughtfully. The good news is that more sustainable approaches are beginning to take shape. Researchers are developing smaller and more efficient models that require less energy to train and run. Some companies now power their facilities with renewable energy, such as wind, solar, or hydro.

Additionally, it is important to encourage developers and users to consider when and where AI is truly necessary and to choose lightweight models where possible.

Conclusion

Artificial Intelligence is no longer a distant concept reserved for science fiction—it's here, shaping our daily lives, industries, and futures. From predictive healthcare and smart farming to personalised shopping and self-driving cars, AI has quietly but powerfully embedded itself into how the world works.

This document has explored how AI has evolved, from early rule-based systems to today's complex learning models, and introduced the key subfields, learning types, and algorithms that underpin technology. It's also shown how businesses have adopted AI in real-world settings, with both success and growing pains.

But understanding AI isn't just about what it can do. It's also about recognising its limitations and responsibilities. Ethical concerns like algorithmic bias, non-consensual data use, and environmental cost highlight the importance of designing AI that is fair, transparent, and sustainable. As AI becomes more advanced, we must remain thoughtful about the impact it has, not just on profits and productivity, but on people and the planet.

AI offers remarkable potential, but its future depends on how we shape it. With the right mix of curiosity, caution, and care, we can build intelligent systems that truly serve the public good.

Test your learning with the attached quiz.