

Artificial Intelligence

All information sources are shown in red

Definition:

" It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable." (IBM - Artificial Intelligence)

Artificial intelligence makes it possible for machines to learn from experience, adjust to new inputs and perform human-like tasks. Most AI examples that you hear about today – from chess-playing computers to self-driving cars – rely heavily on deep learning and natural language processing. Using these technologies, computers can be trained to accomplish specific tasks by processing large amounts of data and recognising patterns in the data. (SAS - Artificial Intelligence)

In summary, the goal is to create systems that think and act rationally like humans.

Applications:

AI and machine learning technology is used in most of the essential applications of the 2020s, including: search engines (such as Google Search), targeting online advertisements, recommendation systems (offered by Netflix, YouTube or Amazon), targeted advertising (Tik Tok, Instagram, Facebook), virtual assistants (such as Siri or Alexa), autonomous vehicles (self-driving cars like Tesla), automatic language translation (Google Translate), facial recognition (Apple's Face ID) and automatic image annotation (used by Facebook, Apple's iPhoto and TikTok). (Wikipedia - Artificial Intelligence)

Ethics:

Artificial intelligence ethics refers to the moral and philosophical principles that guide the development and use of artificial intelligence technologies. It addresses the ethical considerations and challenges associated with artificial intelligence systems, including their design, deployment, and impact on society.

Transparency and Accountability: AI systems should be transparent in their decision-making processes, and there should be mechanisms for accountability when they make mistakes or cause harm.

Privacy: Respecting individuals' privacy and ensuring that AI systems handle personal data in a secure and responsible manner.

Security: Ensuring the security of AI systems to prevent them from being exploited for malicious purposes, such as cyberattacks or spreading disinformation.

Bias and discrimination: Bias in artificial intelligence refers to the presence of unfair and unjust prejudices in artificial intelligence systems. This bias can result from the data used to train artificial intelligence models, the algorithms themselves, or the human decisions made during system development. Bias can manifest in various ways:

Data Bias: If the training data used to develop an artificial intelligence system is not representative of the real-world population, it can lead to biased outcomes. For example, if a facial recognition system is primarily trained on data from one racial group, it may perform poorly for other racial groups. ([Harvard University - Racial Discrimination in Face Recognition Technology](#))

Algorithmic Bias: Even if the training data is unbiased, the algorithms used in artificial intelligence can introduce bias. This can happen if the algorithms inadvertently learn and amplify existing biases present in the data.

Bias in Decision-Making: Biased decision-making can occur when humans in the development process knowingly or unknowingly introduce their own prejudices into the artificial intelligence system.

Future:

Economic Impact: AI is expected to have a profound impact on the global economy. It has the potential to drive productivity, create new industries, and lead to significant economic growth. However, it may also disrupt traditional job markets, leading to concerns about job displacement. ([The Centre for Economic Policy Research \(CEPR\) - The impact of artificial intelligence on growth and employment](#))

Healthcare: AI has the potential to revolutionize healthcare by improving diagnosis, treatment, and patient care. The stakes include the equitable access to AI-driven healthcare solutions and ensuring patient data privacy. ([National Center for Biotechnology Information - The potential for artificial intelligence in healthcare](#))

Autonomous Weapons: there are risks of autonomous weapons systems, where AI can be used for warfare, making decisions without human intervention.

Misinformation: something that is becoming increasingly common is the use of deepfake technology, which can manipulate audio and video to create convincing but false content. This has implications for the spread of misinformation and disinformation. ([Builtin - 12 Risks and Dangers of Artificial Intelligence](#))

AI Ethics: we need more ethical considerations in AI development in the future about privacy, accountability, and transparency. ([University of Oxford - No need to wait for the future, the danger of AI is already here](#))