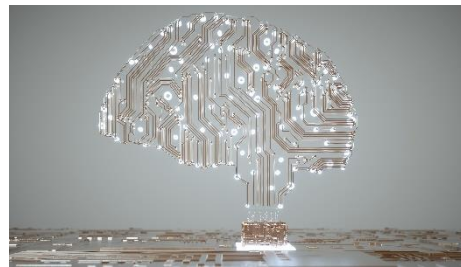


Rapid Technology Assessment: Artificial Intelligence

Artificial intelligence (AI) refers to machine-driven capability to perform tasks and solve problems, imitating human intelligence. AI has huge potential to transform almost every aspect of our lives: how we produce goods and deliver services, protect our national security, develop science and technology, and interact with machines and each other.

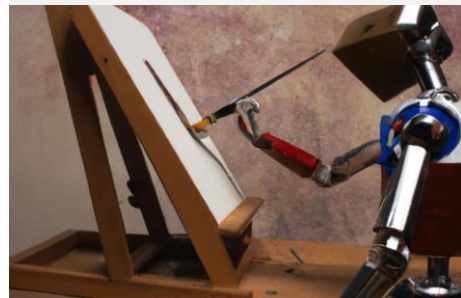
Introduction

- **AI is a rapidly expanding and changing field with many emerging trends, technologies, and capabilities.** Research output has grown exponentially in the last 30 years with the number of academic publications doubling roughly every two years. Alongside this, the demand for computing power has increased as more and ever larger AI systems are developed.
- **Two broad approaches underpin modern AI systems: symbolic AI and machine learning (ML).** In symbolic AI, knowledge and reasoning are explicitly written as rules in the system, whereas ML models learn patterns from data through training. ML has become the dominant approach and underpins many recent breakthroughs and novel applications of AI.
- **Important AI capabilities include** natural language processing (interpreting text/speech), computer vision (interpreting images/videos), systems able to control multiple agents (e.g. drones or satellites), and content generation (e.g. images, videos, text).



UK Position

- **The UK has a strong research base in AI** ranking 4th globally for research publications (2017-2022) and producing impactful research, particularly in areas of AI fairness, transparency, robustness and reliability.
- However, the UK files fewer patents than leading nations (7th). Investment activity is highest in the USA, with the UK ranking 3rd-6th across different AI sub-technologies.



DALL-E Prompt: Robot painting a picture

Recent Developments

- **Increasingly large and complex ML models pre-trained on vast unlabelled datasets have been released** that are more readily adapted to new tasks and can demonstrate new capabilities. These foundation models can be fine-tuned for a new task using much less data than starting from scratch (known as transfer learning). Examples include GPT-3 and DALL-E.
- **The ability of AI systems to generate new and unique content increased dramatically in 2022**, including high quality images, audio, video and naturally phrased text. Generative AI models such as DALL-E, ChatGPT, CoPilot and Stable Diffusion have received widespread attention.
- **There is growing interest and activity in assurance of AI systems** (considering reliability, robustness, fairness) and improving our ability to interpret and explain AI decision making, often challenging when there are billions of complex parameters involved. This could increase developer and user confidence in deploying AI systems in high impact areas.

Opportunities

- **The revolutionary potential of AI is already impacting a range of sectors, services, and fields of research.** From health, life sciences, sustainability and the environment, defence, to space and other parts of the economy there are a range of opportunities for AI systems to improve our lives and tackle key challenges.
- **Advances in generative AI offer a range of opportunities to different sectors.** Potential uses are vast, from research, product design, software development, to customer service and marketing. However, current models have limitations and there are numerous challenges and risks to consider.
- **Increased explainability of AI**, understanding biases and uncertainties, and privacy-preserving methods such as federated learning could enable wider use of AI in high impact areas or where sensitive data is involved.
- **Experts highlighted UK strengths in key AI capabilities** natural language processing, computer vision and multi-agent systems.

Challenges & Risks

- **Ensuring AI systems are understandable tools that can be trusted.** Future governance, standards and regulation will be important alongside technical measures to assure the trustworthiness of AI systems, particularly for high impact areas such as healthcare or autonomous vehicles. The UK has research strengths in this field and is actively developing this sector through the UK National AI Strategy.
- **Improving AI sustainability.** Training large ML models is energy intensive and there is increasing interest in more sustainable approaches that use less energy and computing power. Neuro-symbolic AI combines coded logic with machine learning which could reduce energy need as well as improve model transparency.
- **Foundation models** could underpin a significant proportion of the future AI ecosystem, with any defects or biases in the foundation model being inherited. Due to the high cost, developing foundation models could become a capability limited to a small number of organisations with control over access.