
AI1001 - Assignment 4

Shivram S
ai24btech11031@iith.ac.in

1 History of Artificial Intelligence

The first work generally recognized as AI was a model of neurons made by McCulloch and Pitts in 1943. Another influential model was Hebbian Learning - a simple rule that could be used to 'teach' networks. The first neural computer was built in 1950 at Harvard by Minsky and Edmonds.

The 1950s were a period when AI researchers worked on tasks such as games, puzzles, mathematics and IQ tests. In 1955, McCarthy, along with Minsky, Shannon and Mathaniel Rochester, carried out a 2-month study workshop on intelligence at Dartmouth. Some of the exploratory works of this period include Logic Theorist, General Problem Solver, Geometry Theorem Prover, Arthur Samuel's checkers playing programs, and McCarthy's Advice Taker.

In the 1960s, Minsky developed the idea of microworlds - limited domains that required intelligence to solve. The **blocks** microworld developed the ideas of computer vision and constraint propagation. There was also some early interest in building neural networks, such as an enhancement of Hebbian Learning by Widrow, and the creation of **perceptrons** by Frank Rosenblatt.

Eventually, AI systems failed to meet predictions made by AI researchers. There were several reasons for this failure, such as AI systems merely imitating humans rather than carefully analyzing the task, microworld algorithms being unable to scale up to larger problems, and a lack of computing power. Structures such as the perceptron were also fundamentally limited in what they could represent.

To approach the problem of scalability, **weak methods** were replaced by powerful, domain-specific knowledge that allowed larger reasoning steps. Programs such as **Dendron**, **Mycin**, **R1** and **SHRDLU**, developed using this method, were called **expert systems**. The concepts of **certainty factors** and **frames** were incorporated into these programs.

Neural networks regained interest in the mid-1980s. Several groups applied back-propagation to problems in computer science. This was followed by interest in probabilistic approaches such as Hidden Markov Models for speech recognition and Bayesian networks for representing uncertain knowledge. Markov decision processes gained popularity in robotics and process control.

Advances in computing power and the creation of the World Wide Web facilitated the creation of very large data sets, known as **big data**. This led to the development of learning algorithms designed to take advantage of these. **Deep learning**, which uses multiple layers of simple adjustable neurons, gained popularity and was used in several programs such as **AlexNet** and **AlphaGo**.

2 Risks and Benefits of AI

AI has both risks and benefits. Even though it has led to dramatic acceleration in manufacturing and scientific research, its misuse can lead to risks such as:

1. Usage of AI in **lethal autonomous weapons**. Absence of human supervision allows a small group to deploy an arbitrarily large number of weapons against human targets.

2. AI may be used for **mass surveillance** of individuals to detect activities of interest. Using AI to tailor information flows can allow for **controlling behaviour** to some extent, which is of great concern in politics.
3. Improper use of AI and biases in training data can lead to **biased decision making**, putting certain sections of society at a disadvantage.
4. AI might have an **impact on employment**, rendering some activities economically inviable.
5. Use of AI in **safety-critical applications** is risky. Technical and ethical standards have to be developed when human lives are at stake.
6. **Cybersecurity**: AI may be misused to create highly effective tools for automated blackmail and phishing attacks.

Turing compares the development of artificial intelligence to the evolution of primates into gorillas and humans. Just like gorillas have lost control of their futures, humans might lose control of their futures to AI. In such a case, we would be incentivized to stop work on AI and give up on its benefits.

3 AI Index Report 2024

Generative AI investment is skyrocketing. The United States is the leading source of top AI. Frontier models are getting more expensive, and industry continues to dominate frontier research. Despite increase in research in open-source AI, closed models still outperform open ones.

The number of AI parents have skyrocketed, with 61.1% originating from China. Interest in agentic systems have increased. More students are opting for CS and AI degrees, but AI PhDs are migrating to industry at an alarming rate.

Researchers are trying to develop multimodal AI which is flexible and capable of multiple types of tasks. Google's Gemini and OpenAI's GPT-4 achieve this to some extent. LLMs have also increased the flexibility of robots. AI has begun to help accelerate progress in science and medicine.

AI beats humans on some tasks like image classification and visual reasoning but trails behind on tasks like commonsense reasoning and planning. But standardized evaluation of models is still lacking. Harder benchmarks are emerging, and human evaluation is gaining popularity.

People have become more aware and more nervous about AI's potential impact. AI has captured policymakers' attention, and AI regulations in the US have increased sharply. Concerns such as impersonation, bias, intellectual property rights, privacy, data security and reliability have been raised. Several vulnerabilities have been discovered in LLMs and extreme AI risks have become difficult to analyze.