Chapter 1
Introduction to Artificial Intelligence

Intelligence

Intelligence is:

– The ability to reason

– The ability to understand

– The ability to create

– The ability to Learn from experience

– The ability to plan and execute complex

tasks The intelligent behavior may include

– Everyday tasks: recognize a friend, recognize who is calling, translate from one language to another, interpret a photograph, talk, and cook a dinner

– Formal tasks: prove a logic theorem, geometry, calculus, play chess, checkers, or Go

– Expert tasks: engineering design, medical designers, financial analysis

Artificial Intelligence

AI is the branch of computer science concerned with making computers behave like humans. In other words, AI is the science and engineering of making intelligent machines, especially intelligent computer programs. The process may include

- Learning (Gaining of information and rules for using the information)
- Reasoning (Using the rules to reach approximate or definite conclusions)
- Self-Correction

According to Barr and Feigenbaum:

"Artificial Intelligence is the part of computer science concerned with designing intelligence computer systems, that is, systems that exhibit the characteristics we associate with intelligence in human behavior."

According to Elaine Rich:

"AI is the study of how to make computers do things at which, at the moment, people are better" An AI system should have

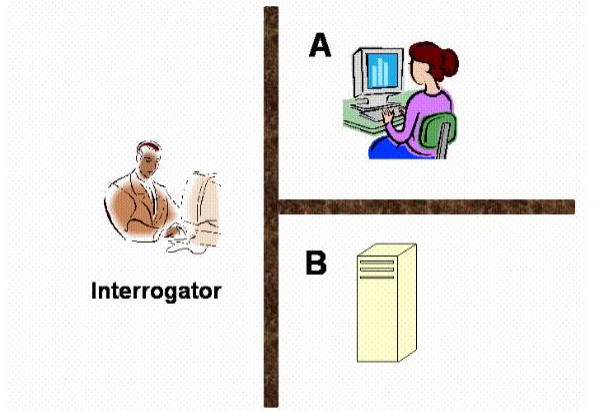
- Capability to provide reason about something
- Capability of natural language processing
- Capability of learning past experience
- Capability of self-correction

Views of AI fall into four categories

Thinking humanly	Thinking rationally
Acting humanly	Acting rationally

- **Acting humanly:** *The Turing Test approach*

  The Turing Test is a method for determining whether or not a computer is capable of thinking like a human. The test is named after Alan Turing, an English mathematician who pioneered artificial intelligence during the 1940s and 1950s, and who is credited with devising the original version of the test. According to this kind of test, a computer is deemed to have artificial intelligence if it can mimic human responses under specific conditions.

Consider the following setting. There are two rooms, A and B. One of the rooms contains a computer. The other contains a human. The interrogator is outside and does not know which one is a computer. He can ask questions through a teletype and receives answers from both A and B. The interrogator needs to identify whether A or B are humans. To pass the Turing test, the machine has to fool the interrogator into believing that it is human.	

To pass a Turing test, a computer must have following capabilities:

Natural Language Processing: Must be able to communicate in English successfully

Knowledge representation: To store what it knows and hears.

Automated reasoning: Answer the Questions based on the stored information.

Machine learning: Must be able to adapt in new circumstances.

- **Thinking humanly:** *The cognitive modeling approach*

  Make the machines having mind like natural mind.

  Cognition: The action or process of acquiring knowledge and understanding through thought, experience and senses.

  How do humans think?

  Requires scientific theories of internal brain activities (cognitive model). Once we

have precise theory of mind, it is possible to express the theory as a computer program.

Two ways of doing this is:

Predicting and testing human behavior (cognitive science)

Identification from neurological data (Cognitive neuro science)

- **Thinking rationally:** *The "laws of thought approach"*

  Aristotle was one of the first who attempt to codify the right thinking that is irrefutable reasoning process. He gave Syllogisms that always yielded correct conclusion when correct premises are given.

  For example:

  Ram is a man

  Man is mortal i.e.

  Ram is mortal

  These laws of thought were supposed to govern the operation of the mind; their study initiated a field called **logic.** The logistic tradition in AI hopes to create intelligent systems using logic programming.

- **Acting rationally :** *The rational agent approach*

  An **agent** is something that acts.

  Computer agent is expected to have following attributes:

  - Autonomous control
  - Perceiving their environment
  - Persisting over a prolonged period of time
  - Adapting to change
  - And capable of taking on another's goal

  Rational behavior: doing the right thing

  The right thing: that which is expected to maximize goal achievement, given the available information

  Rational Agent is one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome.

  In this approach the emphasis is given to correct inferences.

AI and related fields

Different fields have contributed to AI in the form of ideas, viewpoints and techniques.

***Philosophy*:**

Logic, reasoning, mind as a physical system, foundations of learning, language and rationality.

***Mathematics*:**

Formal representation and proof algorithms, computation, undesirability, intractability, probability.

***Psychology*:**

Adaptation, phenomena of perception and motor control.

***Economics*:**

Formal theory of rational decisions, game theory.

***Linguistics*:**

Knowledge representation, grammar

***Neuro science*:**

Physical substrate for mental activities

***Control theory*:**

Homeostatic systems, stability, optimal agent design

Brief History of AI

The term “Artificial Intelligence” was used for the first time in 1956 by an American scientist John McCarthy who is referred to as the Father of AI. McCarthy also come up with a programming language called LISP (i.e. List-Processing), which is still used to program computer in AI that allow the computer to learn. Further, the major achievements can be listed as below:

1943	First electronic computer “Colossus” was developed.
1949	First commercial stored program computer was developed.
1950	- Alan Turing proposes the Turing test as a measure of machine intelligence. - Claude Shannon published a detail analysis of chess playing as search. - Isaac Asimov published his three laws of Robotics.
1951	The first working AI programs were written to run on the Ferranti Mark machine of the University of Manchester; a checkers-playing program written by Christopher Stavechey and a chess-playing program is written by Dietrich Prinz
1955	The first Dartmouth college summer AI conference is organized by John McCarthy, Marvin Minsky, Nathan Rochester of IBM and Claude Shannon.
1956	- The name artificial intelligence is used for the 1st time as the topic of the second Dartmouth Conference, organized by John McCarthy. - The first demonstration of the Logic Theorist (LT) written by Allen Newell, J.C. Shaw and

	Merbart Simon pus is called the first AI program
1957	The general problem Solver (GPS) demonstrated by Newell, Shaw and Simon
1958	John McCarthy at MIT invented the Lisp Programming Language.
1959	- John McCarthy and Marvin Minsky founded the MIT AI Lab. - First industrial robot company, animation was established.
1972	Prolog programming language was developed by Alain Colmerauer
1980	First National Conference of the American Association for Artificial Intelligence (AAAI) was held at Stratford.
Mid 1980's	Neural networks become widely used with the Back propagation algorithm.
1994	AI system exist in real environments with real sensory inputs (i.e. Intelligent Agents)
1997	First time AI system controlled a spacecraft named "Deep Space II"
2007	Checkers is solved by a team of researchers of the University of Alberta.
Present	Programmers are still trying to develop a computer which can successfully pass the "Turing Test".

Application of AI

Artificial intelligence has been used in a wide range of fields including medical diagnosis, stock trading, robot control, law, remote sensing, scientific discovery and toys. Many thousands of AI applications are deeply embedded in the infrastructure of every industry. In the late 90s and early 21st century, AI technology became widely used as elements of larger systems, but the field is rarely credited for these successes.

Game Playing

Machines can play master level chess. There is some AI in them, but they well against people mainly through brute force method, looking at hundreds of thousands of positions.

Speech Recognition

It is possible to instruct some computers using speech. In 1990s, computer speech recognition reached a practical level for limited purposes.

Understanding Natural Language

To perform many natural language processing tasks such as machine translation, summarization, information extraction, word sense disambiguation need the AI in machine.

Computer Vision

Computer vision is concerned with the theory behind artificial system that extract information from images. The image data can take many forms such as videos sequences views from multiple cameras and data from a medical scanner. Application range from simple tasks such as industrial machine, vision system which count bottles speeding by on a production line to research into artificial intelligence and computers or robots that can comprehended the world around them.

Expert System

Expert system needs the AI to perform its task. One of the first expert system was MYCIN in 1974 which diagnosis bacterial infections of the blood and suggests treatments. It did better that makes medical students practicing doctors provided to limitations were observed.

Finance

Financial institutions have long used artificial neural network systems to detect charges or claims outside of the norm, flagging these for human investigation. Use of AI in banking can be traced back to 1987 when Security Pacific National Bank in USA set-up a Fraud Prevention Task force to counter the unauthorized use of debit cards.

Hospitals and medicine

Artificial neural networks are used as clinical decision support systems for medical diagnosis, such as in Concept Processing technology in EMR software.

Other tasks in medicine that can potentially be performed by artificial intelligence include:

Computer-aided interpretation of medical images. Such systems help scan digital images, *e.g.* from computed tomography, for typical appearances and to highlight conspicuous sections, such as possible diseases. A typical application is the detection of a tumor.

Heart sound analysis

Companion robots for the care of the elderly

Heavy industry

Robots have become common in many industries. They are often given jobs that are considered dangerous to humans. Robots have proven effective in jobs that are very repetitive which may lead to mistakes or accidents due to a lapse in concentration and other jobs which humans may find degrading. Japan is the leader in using and producing robots in the world. In 1999, 1,700,000 robots were in use worldwide.

Online and telephone customer service

Artificial intelligence is implemented in automated online assistants that can be seen as avatars on web pages. It can avail for enterprises to reduce their operation and training cost. A major underlying technology to such systems is natural language processing.

Toys and games

The 1990s saw some of the first attempts to mass-produce domestically aimed types of basic Artificial Intelligence for education, or leisure. This prospered greatly with the Digital Revolution, and helped introduce people, especially children, to a life of dealing with various types of Artificial Intelligence. AI has also been applied to video games, for example video game bots, which are designed to stand in as opponents where humans aren't available or desired

Music

The evolution of music has always been affected by technology. With AI, scientists are trying to make the computer emulate the activities of the skillful musician. Composition, performance, music theory, sound processing are some of the major areas on which research in Music and Artificial Intelligence are focusing.

Aviation

The Air Operations Division (AOD) uses AI for the rule based expert systems. The AOD has use for artificial intelligence for replacement operators for fighting and training simulators, mission management aids, support systems for tactical decision making, and post processing of the simulator data into symbolic summaries.

Knowledge and Learning

Knowledge is the information about a domain that can be used to solve problems in that domain. To solve many problems requires much knowledge, and this knowledge must be represented in the computer. As part of designing a program to solve problems, we must define how the knowledge will be represented.

A representation of some piece of knowledge is the internal representation of the knowledge. A representation scheme specifies the form of the knowledge. A knowledge base is the representation of all of the knowledge that is stored by an agent.

A good representation should be

Rich enough to express the knowledge needed to solve the problem. Willing for efficient computation

Able to be acquired from people, data and past experiences.

Knowledge is the body of facts and principles. Knowledge can be language, concepts, procedures, rules, ideas, abstractions, places, customs, and so on. (Study of knowledge is called Epistemology)

Types of knowledge

The types of knowledge include procedural knowledge, declarative knowledge and heuristic knowledge.

- **Meta Knowledge**

  It is knowledge about knowledge and how to gain them.

- **Procedural knowledge**

  Procedural knowledge is related to the performance of some task. For example, sequence of steps to solve a problem is procedural knowledge.

- **Declarative knowledge**

  Declarative knowledge is passive knowledge in the form of statements of facts about the world. For example, mark statement of a student is declarative knowledge.

- **Heuristic knowledge**

  Heuristic knowledge is used to make judgments and also to simplify solution of problems. It is acquired through experience. An expert uses his knowledge that he has gathered due to his experience and learning.
- **Structural Knowledge**

  Describes what relationship exists between objects.

Learning:

Learning is acquiring new or modifying existing knowledge, behaviors, skills, values and may involve synthesizing different types of information.

Machine learning, a branch of AI, is a scientific discipline concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data such as from sensor data or database.

Benefits of AI

1. Efficiency and Productivity
   - AI can automate repetitive tasks, streamline workflows, and handle large amounts of data, boosting productivity across industries. This enables quicker and more cost-effective solutions in areas like manufacturing, logistics, and customer service.
2. Enhanced Decision-Making
   - AI can analyze complex data patterns far beyond human capacity, supporting informed decision-making in fields such as finance, healthcare, and business strategy. AI-assisted decision-making can lead to more accurate diagnoses, better risk assessments, and optimized business operations.
3. Personalization
   - In fields like e-commerce, entertainment, and education, AI can personalize experiences to individual preferences. For instance, AI-driven recommendations on streaming platforms or personalized learning plans in education allow users to have tailored experiences.
4. Medical Advancements
   - AI is revolutionizing healthcare by helping with early disease detection, diagnosis, and personalized treatment plans. AI can analyze medical data to identify

patterns in symptoms and outcomes, contributing to preventative care and more effective treatments.

5. Scientific Discovery and Innovation
    - AI accelerates scientific research by simulating complex phenomena, processing vast amounts of research data, and identifying new scientific insights. In fields like chemistry, biology, and climate science, AI aids in drug discovery, genetic research, and understanding environmental impacts.
6. Improved Accessibility
    - AI tools, such as speech-to-text, image recognition, and language translation, improve accessibility for individuals with disabilities. This helps in creating a more inclusive society where services and information are available to everyone.
7. Environmental Solutions
    - AI contributes to environmental monitoring, climate modeling, and energy efficiency. It helps track deforestation, optimize energy use, and predict weather patterns, aiding in sustainable practices and conservation efforts.

Risks of AI

1. Bias and Discrimination
    - AI systems can perpetuate or even amplify biases present in their training data. This can lead to unfair treatment in areas like hiring, lending, and criminal justice, where biased algorithms may discriminate against certain groups.
2. Privacy Invasion
    - AI often relies on large datasets that include personal information, raising significant privacy concerns. AI-powered surveillance tools, facial recognition, and predictive analytics can intrude on individuals' privacy if misused.
3. Job Displacement and Economic Inequality
    - AI-driven automation could displace jobs in sectors such as manufacturing, transportation, and customer service, potentially leading to unemployment and increased economic inequality. While new jobs may emerge, there is concern about the pace of this transition and its impact on workers.
4. Security Threats and Cybersecurity Risks

- AI can be exploited by malicious actors to develop sophisticated cyberattacks, deep fakes, and misinformation campaigns. These technologies can compromise national security, spread disinformation, and erode public trust in digital media.

5. Loss of Autonomy and Human Control
    - Autonomous systems, such as self-driving cars or drones, operate independently, raising concerns about their behavior in unexpected situations. Ensuring that these systems act safely and under human control, particularly in high-stakes applications, is a major challenge.
6. Transparency and Accountability
    - Many AI systems, especially complex models like deep neural networks, operate as "black boxes," making it difficult to understand their decision-making processes. This lack of transparency can create accountability issues, especially if AI systems make critical mistakes in areas like healthcare or law enforcement.
7. Existential Risks from Advanced AI
    - There are concerns about the potential development of superintelligent AI that could surpass human control. If not carefully designed and managed, highly autonomous AI systems could operate in ways that threaten humanity. This is a long-term risk that has prompted research into AI safety and ethical alignment.

Ethics and Societal Implications

Artificial intelligence, in its essence, refers to the simulation of human intelligence processes by machines, particularly computer systems. The ethical implications of artificial intelligence pertain to the ethical challenges, dilemmas, and consequences that arise from the deployment and application of AI technologies. This encompasses ethical considerations regarding the use of AI in decision-making, potential biases within AI systems, accountability for machine-generated outcomes, and the overall impact of AI on individuals and society.

Ethical implications of AI

1. The Control Problem and Value Alignment

- Control Problem: As AI systems become more capable, ensuring they act in alignment with human values becomes critical. The "control problem" addresses the challenge of designing AI that can be safely controlled, especially as systems approach superintelligence.

- Value Alignment: The authors emphasize aligning AI goals with human ethical values, suggesting that poorly defined goals in AI could lead to harmful unintended consequences. AI systems need to be designed to understand and prioritize human intentions.

2. Autonomous Systems and Decision-Making

- Ethical Dilemmas in Autonomy: AI-driven autonomous systems (e.g., self-driving cars) must often make complex ethical decisions, particularly in life-or-death scenarios.
- Transparency in Decision-Making: For autonomous systems to be trusted, their decision-making processes must be transparent. The authors call for AI designs that allow stakeholders to understand and evaluate how decisions are made, particularly in critical or safety-sensitive areas.

3. Bias and Fairness in AI Systems

- Bias in Training Data: AI systems learn from data that often contains human biases. If unchecked, these biases can lead to discriminatory or unfair outcomes in applications like hiring, criminal justice, and lending.
- Fairness and Inclusivity: The need for fairness in AI by designing systems that actively mitigate biases is needed. Ongoing monitoring and adjustments are essential to ensure ethical AI applications.

4. Privacy and Surveillance

- Data Privacy Risks: AI often relies on large datasets, some of which contain personal information. The authors highlight the importance of protecting individual privacy, particularly as AI is used in fields like healthcare and law enforcement.
- Surveillance Concerns: AI-powered surveillance technologies, such as facial recognition, raise ethical concerns about privacy, consent, and misuse by governments or organizations. Safeguards and regulations are essential to prevent the abuse of such technologies.

5. Accountability and Transparency

- Black Box Models: Many AI models, especially those based on deep learning, operate as “black boxes” that are difficult to interpret. This lack of transparency raises issues of accountability, especially in applications where AI impacts lives.
- Explainable AI: The authors advocate for explainable AI that allows users to understand AI decisions. This transparency is necessary to establish trust and ensure that AI developers can be held accountable for outcomes.

6. Economic Impact and Employment Displacement

- Automation and Job Loss: The book addresses the economic implications of AI, particularly the potential for job displacement due to automation. As AI and robotics automate routine tasks, many jobs could be replaced, disproportionately affecting certain industries and populations.
- Preparing for Economic Shifts: To mitigate these impacts, it is suggested that proactive approaches like workforce retraining, education, and policy measures to support displaced workers.

7**. Long-Term Existential Risks**

- Superintelligence Risks: The authors acknowledge the long-term risks associated with AI systems that could surpass human intelligence. They discuss the need for safety measures to prevent super intelligent AI from acting counter to human interests.
- AI Safety Research: Ongoing research in AI safety and ethics is encouraged to address potential existential risks, ensuring that AI systems remain beneficial and under human control.

Governance and regulation

Artificial intelligence (AI) governance refers to the processes, standards and guardrails that help ensure AI systems and tools are safe and ethical. AI governance frameworks direct AI research, development and application to help ensure safety, fairness and respect for human rights.

Effective AI governance includes oversight mechanisms that address risks such as bias, privacy infringement and misuse while fostering innovation and building trust. An ethical AI-centered approach to AI governance requires the involvement of a wide range of stakeholders, including AI developers, users, policymakers and ethicists, ensuring that AI-related systems are developed and used to align with society's values.

AI governance is essential for reaching a state of compliance, trust and efficiency in developing and applying AI technologies. With AI's increasing integration into organizational and governmental operations, its potential for negative impact has become more visible.