Lecture(4):

Introduction to AI:

Definition of AI

History of AI

AI techniques and approaches

Challenges and ethical considerations

1. Introduction to AI

Artificial intelligence (AI) is a branch of computer science that aims to design systems capable of simulating human intelligence, such as learning, decision-making, problem-solving, and perception. Russell and Norvig define AI as:

"The study of intelligent agents": entities that perceive their environment and take actions to maximize the chances of achieving their goals.

Main types:

Narrow AI: Specializes in specific tasks (e.g., Netflix recommendations).

Artificial General Intelligence (AGI): Thinks like humans in all domains (not yet realized).

Source: Chapter 1, Section 1.1 (4th edition, 2020).

2. History of Artificial Intelligence (AI)

Beginnings (1943–1956):

The artificial neuron model (McCulloch and Pitts, 1943).

The Turing test (1950) to measure machine intelligence.

Optimistic period (1956–1974):

Dartmouth conference (1956) where the term "artificial intelligence" was coined.

The emergence of systems such as ELIZA (language processing).

The first AI winter (1974–1980): Funding declined due to the difficulty of achieving exaggerated expectations.

Resurgence of interest (1980–2000):

Expert systems such as MYCIN for medical diagnosis.

The development of machine learning algorithms.

Modern era (2000–present):

The big data revolution and deep learning.

Achievements such as AlphaGo and ChatGPT.

Source: Chapter 1, Section 1.3 and Chapter 28 (4th Edition).

3. AI Techniques and Approaches

A. Classic Approaches:

Symbolic AI:

Relies on logical rules and knowledge representation (e.g., Prolog).

Limitations: Limitations in handling unstructured data.

Neural Networks:

Simulate the structure of the human brain through layers of connected nodes.

Evolution: From simple models (Perceptron) to deep networks (CNN, RNN).

B. Modern Approaches:

Machine Learning:

Supervised Learning: Using labeled data (e.g., image classification).

Unsupervised Learning: Discovering patterns in unlabeled data (e.g., clustering).

Reinforcement Learning: Learning with rewards (e.g., AlphaGo).

Natural Language Processing (NLP):

Models like GPT-4 for text understanding and machine translation.

Source: Chapters 2 (Intelligent Agents), 7 (Machine Learning), and 21 (Neural Networks).

4. Challenges and Ethical Considerations

A. Technical Challenges:

The Explainability Problem: Difficulty understanding the decisions of deep learning systems (“black boxes”).

Algorithmic Bias: Models being biased due to imbalanced training data (e.g., discrimination in hiring).

The Need for Big Data: The complexity of training models without violating privacy.

B. Ethical Considerations:

Privacy: The collection of personal data without consent (e.g., facial recognition in surveillance).

Technological Unemployment: The replacement of jobs by intelligent systems (e.g., self-driving cars).

Autonomous Weapons: The use of AI in warfare without human control.

Human Presence: Long-term risks such as the loss of control over AGI (superintelligence).