ABHYASMITRA.IN

COMPREHENSIVE UNIVERSITY STUDY NOTES

ARTIFICIAL INTELLIGENCE

Unit I — Introduction to AI & Intelligent Agents

Topics Covered in this Unit

Introduction to AI: Definitions (Thinking/Acting Humanly & Rationally), Turing Test & Chinese Room Argument

Chronological History of AI: Dartmouth 1956, Early Optimism, AI Winters, Expert Systems & Deep Learning Boom

Spectrum of AI Systems: Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI) & Super AI (ASI)

Intelligent Agents & Concept of Rationality: Omniscience vs. Rationality, Performance Measures & Rational Agents

PEAS Framework (Performance, Environment, Actuators, Sensors) & Comprehensive 7-Property Environment Taxonomy

Internal Structures of Intelligent Agents: Simple Reflex, Model-Based, Goal-Based, Utility-Based & Learning Agents

AI in Problem Solving: Formal State Space Representation (Initial State, Actions, Transition Model, Goal Test, Path Cost)

AI Ethics, Safety & Societal Challenges: Algorithmic Bias, Autonomous Weapons, Transparency & Job Displacement

High-Yield University Exam Question Bank with Detailed Model Answers & Unit Summary

Compiled in a structured, easy-to-understand, textbook style format
Specially curated for B.Tech / B.E. / BCA / MCA / B.Sc Computer Science, AI & Data Science Students

COURSE SYLLABUS & MAPPING

[UNIT I SYLLABUS — PRESCRIBED UNIVERSITY CURRICULUM]

This section contains the official syllabus for Unit I. Verify with your university curriculum mapping.

Unit I Syllabus Breakdown:

Unit I - Introduction to AI: Introduce to the fundamentals of AI, history, applications, and challenges. Definition and History of AI, Applications and Impact of AI in Various Domains, Types of AI: Narrow AI, General AI, and Super AI, Intelligent Agents, Agents and Environments, Concept of Rationality, Nature of Environments, Structure of Agents. AI in Problem Solving: State Space Representation, AI Ethics and Challenges

• Course Code: ___________________ • Semester / Year: ___________________

TABLE OF CONTENTS

1.1 Introduction to Artificial Intelligence: Fundamentals & History

Formal Definition: Artificial Intelligence (Russell & Norvig)

Artificial Intelligence (AI) is the study of agents that receive percepts from the environment and perform actions to maximize their chance of successfully achieving their goals (Rational Action). Historically, AI definitions span across a 2×2 matrix: Thinking Humanly (Cognitive Science), Thinking Rationally (Laws of Thought / Logic), Acting Humanly (Turing Test), and Acting Rationally (Rational Agents).

The Four Approaches to AI

Approach Dimension	Human-Centric Focus (Empirical / Psychological)	Rational-Centric Focus (Mathematical / Ideal)
Thinking (Cognitive / Internal)	Thinking Humanly: Cognitive Modeling. Requires experimental validation of biological neural thought processes (fMRI scans, psychology experiments).	Thinking Rationally: Laws of Thought. Formalized deductive logic (Aristotle's Syllogisms) ensuring mathematically sound conclusions from premises.
Acting (Behavioral / External)	Acting Humanly: The Turing Test Approach (Alan Turing, 1950). A machine passes if a human interrogator cannot distinguish it from a human through text conversation.	Acting Rationally: The Rational Agent Approach. Designing computational agents that act to achieve the best expected outcome given their percept history.

Chronological History and Evolution of AI

- **1. Inception & Foundations (1943–1955):** McCulloch & Pitts (1943) artificial binary neuron model; Hebbian learning (1949); Alan Turing's seminal paper 'Computing Machinery and Intelligence' (1950).
- **2. The Birth of AI (Dartmouth Workshop, 1956):** John McCarthy, Marvin Minsky, Claude Shannon, and Nathaniel Rochester coined the term 'Artificial Intelligence'. Newell & Simon demonstrated the Logic Theorist (first AI program solving mathematical theorems).
- **3. Early Enthusiasm & Great Expectations (1952–1969):** Development of General Problem Solver (GPS), Lisp programming language (McCarthy, 1958), Rosenblatt's Perceptron (1958), and ELIZA chatbot (Weizenbaum, 1966).
- **4. The First AI Winter (1974–1980):** Minsky & Papert proved single-layer Perceptrons cannot solve non-linear XOR functions (1969). Combinatorial explosion and failure of machine translation led to severe government funding cuts (Lighthill Report).

- **5. The Expert Systems Boom (1980–1987):** Rule-based expert systems (MYCIN, DENDRAL, XCON) saved enterprises millions of dollars by encoding specialized domain knowledge into IF-THEN production rules.
- **6. The Second AI Winter & Statistical Revolution (1987–2000s):** Expert systems proved brittle and expensive to maintain. AI shifted toward rigorous probabilistic reasoning (Judea Pearl's Bayesian Networks), Support Vector Machines, and Hidden Markov Models.
- **7. The Deep Learning & Generative AI Revolution (2012–Present):** AlexNet (2012) revolutionized computer vision on GPUs. DeepMind's AlphaGo (2016) conquered Go. Vaswani et al. (2017) introduced the Transformer architecture, sparking modern Large Language Models (GPT-4, Gemini).

1.2 Types of AI: Narrow AI, General AI, and Super AI

Classification Level	Cognitive Scope & Autonomy	Core Characteristics	Real-World Examples & Status
Artificial Narrow Intelligence (ANI / Weak AI)	Specialized in a single dedicated task; lacks generalized cognitive transfer.	Operates strictly within programmed parameters; cannot transfer expertise from one domain to another.	Spam filters, Siri/Alexa, AlphaFold, Tesla Autopilot vision, Chess engines. (CURRENT STATUS OF ALL MODERN AI).
Artificial General Intelligence (AGI / Strong AI)	Human-level broad cognitive intellect across all intellectual and creative domains.	Possesses common-sense reasoning, abstract conceptualization, transfer learning, and autonomous adaptability.	Theoretical target of leading research labs (OpenAI, DeepMind, Anthropic). Estimated: Near future (2030s).
Artificial Superintelligence (ASI)	Intellectual capacity drastically exceeding the combined collective intellect of all human brains.	Self-improving recursive algorithms solving physics, interstellar travel, biological immortality, and quantum gravity.	Theoretical future horizon / Technological Singularity (Nick Bostrom's superintelligence thesis).

1.3 Intelligent Agents & The Concept of Rationality

Definition: Agent, Percept, Action & Rationality

An Agent is anything that perceives its environment through Sensors and acts upon that environment through Actuators.

An Agent Function $f: P^* \rightarrow A$ is an abstract mathematical mapping from the complete percept sequence P^* to an action A.

An Agent Program is the concrete physical implementation of the agent function running on an architecture machine.

A Rational Agent is one that selects an action expected to maximize its Performance Measure, given the evidence provided by the percept sequence and whatever built-in knowledge the agent possesses.

Omniscience vs. Rationality

- • **Omniscient Agent:** Knows the actual physical outcome of its actions and acts with 100% infallible success in advance (impossible in the real world).
- • **Rational Agent:** Maximizes expected success based on current knowledge and percept history. Rationality does not mean perfection; a rational agent can make a mistake due to incomplete information, but its decision was optimal given available data.

1.4 The PEAS Framework & Taxonomy of Task Environments

The PEAS Specification

To design a rational agent, the Task Environment must be formally specified via PEAS (Performance Measure, Environment, Actuators, Sensors):

Agent System	Performance Measure (P)	Environment (E)	Actuators (A)	Sensors (S)
Autonomous Taxi Driver	Safe trip, speed, legal compliance, comfort, maximized profit.	Roads, other vehicles, pedestrians, weather conditions.	Steering wheel, accelerator, brakes, horn, turn signals.	Cameras, LiDAR, Radar, GPS, speedometer, odometer, sonar.
Medical Diagnosis Agent	Healthy patient, minimized healthcare cost, zero misdiagnoses.	Patient, hospital clinic staff, laboratory test systems.	Displaying screen diagnoses, medical prescription orders.	Keyboard symptoms input, lab test blood analyzers, vital monitors.
Automated Vacuum Cleaner	Cleanliness of floor, low power consumption, minimized time.	Floors, rooms, walls, dirt, obstacles, carpets.	Drive wheels, vacuum suction motor, rotating brush.	Bump sensors, optical dirt sensors, infrared cliff sensors.

The 7 Core Dimensions of Task Environments

Environment Property	Fully Observable vs. Partially Observable	Deterministic vs. Stochastic	Episodic vs. Sequential
Definitions	Fully: Sensors detect complete state of the environment at all times. Partially: Noisy/missing sensor data.	Deterministic: Next state is completely determined by current state and agent action. Stochastic: Randomness/uncertainty.	Episodic: Agent's experience divided into atomic episodes; past actions don't affect future. Sequential: Current action affects future.
Examples	Chess = Fully; Poker / Driving = Partially.	Chess = Deterministic; Taxi driving = Stochastic.	Defect inspection = Episodic; Chess / Driving = Sequential.

Environment Property	Static vs. Dynamic	Discrete vs. Continuous	Single-Agent vs. Multi-Agent	Known vs. Unknown
Definitions	Static: Environment does not change while agent deliberates. Dynamic: Changes continuously.	Discrete: Finite, distinct states, time steps, and actions. Continuous: Infinite continuous real values.	Single: Agent acts alone. Multi: Multiple competitive or cooperative agents interact.	Known: Agent knows all physics/rules. Unknown: Must explore to learn laws.
Examples	Crossword = Static; Driving = Dynamic.	Chess = Discrete; Taxi = Continuous.	Crossword = Single; Chess = Competitive Multi.	Video game rules = Known; New planet = Unknown.

1.5 Structure of Intelligent Agents

Intelligent agents are constructed using five fundamental architectural archetypes of increasing sophistication:

1. Simple Reflex Agents

- • **Mechanism:** Selects actions based only on the current percept, ignoring complete percept history. Governed by condition-action rules: IF condition THEN action.
- • **Limitation:** Only works in Fully Observable environments; falls into infinite loops in partially observable domains.

2. Model-Based Reflex Agents

- • **Mechanism:** Maintains an internal state that tracks unobserved aspects of the current world. Incorporates two types of knowledge: (1) How the world evolves independently, (2) How the agent's own actions affect the world.

- • **Advantage:** Successfully handles Partially Observable environments.

3. Goal-Based Agents

- • **Mechanism:** Combines internal state with explicit Goal descriptions to evaluate which actions will achieve desired future states. Uses Search and Planning algorithms.

4. Utility-Based Agents

- • **Mechanism:** Maps state history to a real number via a Utility Function U(s), measuring how desirable a state is. Resolves trade-offs between conflicting goals and calculates expected utility under uncertainty.

5. Learning Agents

- • **Four Functional Components:**
  - **1. Learning Element:** Responsible for making improvements by learning from experience.
  - **2. Performance Element:** The operational agent responsible for selecting external actions.
  - **3. Critic:** Evaluates agent behavior against a fixed external performance standard.
  - **4. Problem Generator:** Suggests exploratory actions that lead to new, informative experiences.

1.6 AI in Problem Solving: State Space Representation

Formal 5-Tuple Formulation of a Problem

A well-defined problem in AI is formally specified by 5 components:

1. Initial State: The starting configuration s_0 ∈S.

2. Actions Available: ACTIONS(s) returning valid moves in state s.

3. Transition Model: RESULT(s, a) returning the successor state after executing action a in state s.

4. Goal Test: A predicate GOAL-TEST(s) determining whether state s is a goal state.

5. Path Cost Function: c(s, a, s') assigning a numeric step cost to transitioning from s to s' via a. Total path cost g(n) is the sum of step costs.

State Space Graphs vs. Search Trees

- • **State Space Graph:** An abstract mathematical graph where nodes represent unique physical states and directed edges represent valid actions. May contain cycles and loops.
- • **Search Tree:** An explicit tree overlay constructed dynamically during search. Nodes represent search paths from the root initial state. If an algorithm visits the same state via different paths, multiple distinct tree nodes contain the same underlying state.

1.7 AI Ethics, Safety & Societal Challenges

- **1. Algorithmic Bias & Discrimination:** AI models trained on historical human datasets replicate and amplify historical racial, gender, and socioeconomic biases (e.g., facial recognition disparities, hiring screening algorithms).
- **2. Transparency & Explainability (Black Box AI):** Deep neural networks lack interpretable reasoning, posing severe risks in high-stakes clinical medical diagnostics and judicial bail decisions.
- **3. Safety, Robustness & Value Alignment:** Ensuring super-intelligent AI systems align with human moral values and do not pursue unintended destructive sub-goals (The Alignment Problem).
- **4. Autonomous Weapons & Warfare:** Deploying lethal autonomous weapons systems (LAWS) without human-in-the-loop ethical oversight.
- **5. Economic & Labor Disruption:** Mass automation of cognitive, administrative, and creative tasks altering global workforce stability.

1.8 High-Yield University Exam Question Bank & Model Answers

Part A: Short Answer Questions (2–3 Marks)

- **Q1. State the PEAS description for an Automated Taxi Driver.**

Answer: Performance Measure: Safety, speed, comfort, legal compliance, profit. Environment: Roads, other traffic, pedestrians, weather. Actuators: Steering, accelerator, brakes, signals, horn. Sensors: Cameras, LiDAR, Radar, GPS, speedometer.

- **Q2. Differentiate between an Omniscient Agent and a Rational Agent.**

Answer: An omniscient agent knows the actual future outcome of its actions and acts with infallible perfection. A rational agent acts to maximize its expected performance given its percept history and available knowledge, making the best possible decision under uncertainty without possessing clairvoyance.

- **Q3. What are the 5 components of a formal Problem Formulation in AI?**

Answer: (1) Initial State, (2) Actions Function ACTIONS(s), (3) Transition Model RESULT(s, a), (4) Goal Test predicate GOAL-TEST(s), and (5) Path Cost function c(s, a, s').

- **Q4. Describe a Simple Reflex Agent and state its key limitation.**

Answer: A Simple Reflex Agent selects actions based solely on the current percept via condition-action rules (IF-THEN). Its key limitation is that it can only succeed in fully observable environments; in partially observable environments, it easily enters unrecoverable infinite loops.

Part B: Long Answer Questions (8–10 Marks)

- **Q5. Explain the architecture of all 5 types of Intelligent Agents (Simple Reflex, Model-Based, Goal-Based, Utility-Based, Learning) with detailed block diagrams.**

Model Answer Outline:

- 1. Introduction: Definition of Agent Function f: $P^* \rightarrow A$ and Agent Program.
- 2. Simple Reflex Agent: Condition-Action rules, direct mapping, lack of memory.
- 3. Model-Based Reflex Agent: Internal state, transition model, sensor model for partial observability.
- 4. Goal-Based Agent: Search, planning, evaluating future consequences to satisfy goals.
- 5. Utility-Based Agent: Utility functions, resolving conflicting goals, decision-theoretic rationality under uncertainty.
- 6. Learning Agent: Detailed breakdown of Learning Element, Performance Element, Critic, and Problem Generator.
- 7. Complete architectural block diagrams illustrating data flows for each archetype.

- **Q6. Discuss the Nature of Task Environments in AI. Detail all 7 environmental dimensions with real-world agent examples.**

Model Answer Outline:

- 1. Definition of Task Environment and the PEAS framework.
- 2. Detailed explanation and contrast of all 7 environmental dimensions:
- - Fully Observable vs Partially Observable.
- - Deterministic vs Stochastic vs Strategic.
- - Episodic vs Sequential.
- - Static vs Dynamic vs Semidynamic.
- - Discrete vs Continuous.
- - Single-Agent vs Multi-Agent (Cooperative vs Competitive).
- - Known vs Unknown.
- 3. Comprehensive multi-column classification table evaluating Chess, Medical Diagnosis, Taxi Driving, and Part-Picking Robot across all dimensions.

1.9 Unit I Summary & Quick Revision Sheet

Core Concept	Key Theoretical Takeaways & Exam Points
AI Approaches	Thinking/Acting Humanly (Cognitive/Turing Test) vs Thinking/Acting Rationally (Logic/Rational Agents).
AI Hierarchy	Narrow AI (ANI - Task specific, modern status) → General AI (AGI - Human-level) → Super AI (ASI - Beyond human).
PEAS Framework	Performance, Environment, Actuators, Sensors. 7 Dimensions: Observability, Determinism, Episodicity, Dynamism, Discreteness, Multi-agent, Knowledge.
Agent Types	Simple Reflex (Condition-action), Model-Based (Internal state), Goal-Based (Goals), Utility-Based (Utility score), Learning (Critic + Problem Generator).
State Space Formulation	5-Tuple: Initial State, Actions, Transition Model, Goal Test, Path Cost. Graph vs Tree search.