

Introduction to Artificial Intelligence

Foundations, Concepts, and Intelligent Agents

Kiran Bagale

IOE, Thapathali Campus

November 2025

Outline

- 1 Definition, Foundation, and History of AI
- 2 AI Tree: Branches and Interdisciplinary Nature
- 3 Knowledge and Learning
- 4 Human Intelligence and Machine Intelligence
- 5 Intelligent Agents and Types
- 6 Formal, Mundane, and Expert Tasks in AI
- 7 Conclusion

What is Artificial Intelligence?

Definition

Artificial Intelligence is the study of agents that receive percepts from the environment and perform actions [1].

- **Thinking Humanly:** Cognitive modeling approach
- **Thinking Rationally:** Laws of thought approach
- **Acting Humanly:** Turing test approach
- **Acting Rationally:** Rational agent approach

AI aims to create systems that can perform tasks requiring human-like intelligence.

AI draws from multiple disciplines:

Technical Foundations:

- Philosophy
- Mathematics
- Economics
- Neuroscience

Engineering Foundations:

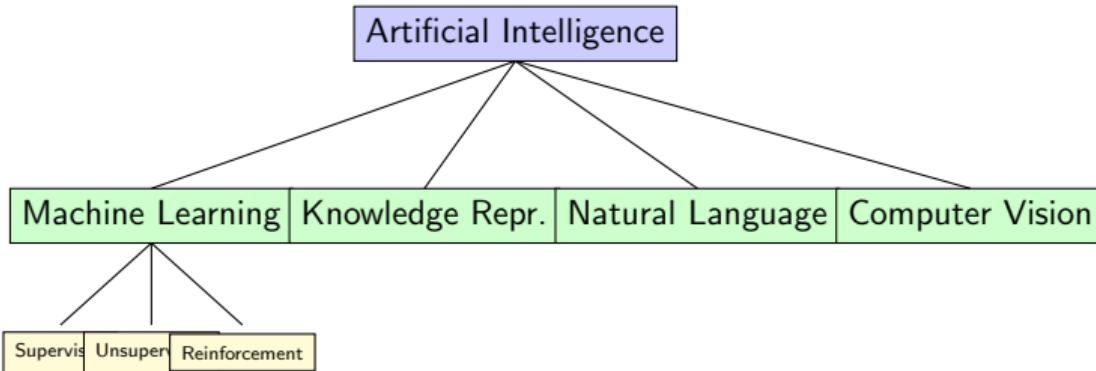
- Computer Science
- Control Theory
- Linguistics
- Psychology

[1] emphasizes the interdisciplinary nature of AI research.

Brief History of AI

- **1943-1956:** Gestation period
 - McCulloch-Pitts neural model (1943)
 - Turing's "Computing Machinery and Intelligence" (1950)
- **1956:** Birth of AI at Dartmouth Conference
- **1956-1974:** Early enthusiasm and success
 - Logic Theorist, General Problem Solver
 - Symbolic AI approaches
- **1974-1980:** First AI Winter
- **1980-1987:** Expert systems boom
- **1987-1993:** Second AI Winter
- **1993-Present:** Modern AI renaissance
 - Machine learning, deep learning
 - Big data and computational power

The AI Tree: Major Branches



[2] describes AI as a collection of interrelated subfields.

Major AI Branches

① Machine Learning

- Learning from data and experience
- Pattern recognition [3]

② Knowledge Representation and Reasoning

- Logical systems, ontologies
- Inference and deduction

③ Natural Language Processing

- Understanding and generating human language
- Machine translation, sentiment analysis

④ Computer Vision

- Image recognition and understanding
- Object detection, scene analysis

⑤ Robotics

- Physical interaction with environment
- Autonomous systems

Interdisciplinary Nature of AI

AI integrates concepts from:

Core Disciplines:

- **Mathematics:** Logic, probability, optimization [4]
- **Computer Science:** Algorithms, data structures
- **Statistics:** Inference, hypothesis testing

Supporting Fields:

- **Cognitive Science:** Human reasoning models
- **Neuroscience:** Brain-inspired computing
- **Linguistics:** Language structure

Key Insight

AI's strength comes from synthesizing insights across multiple domains.

What is Knowledge?

Definition

Knowledge is information that has been organized, structured, and integrated with existing understanding to enable reasoning and decision-making.

Types of Knowledge in AI:

- **Declarative Knowledge:** Facts and relationships ("Paris is the capital of France")
- **Procedural Knowledge:** How to perform tasks ("How to solve an equation")
- **Heuristic Knowledge:** Rules of thumb and best practices
- **Structural Knowledge:** Relationships and hierarchies

[2] emphasizes knowledge representation as fundamental to AI systems.

Importance of Knowledge in AI

Why Knowledge Matters:

- ① **Problem Solving:** Domain knowledge guides search and reasoning
- ② **Inference:** Enables drawing conclusions from observations
- ③ **Communication:** Allows understanding and generation of natural language
- ④ **Planning:** Supports goal-directed behavior
- ⑤ **Learning:** Provides context for interpreting new information

Knowledge Engineering

The process of building knowledge-based systems involves:

- Knowledge acquisition
- Knowledge representation
- Knowledge validation

What is Learning?

Machine Learning Definition

Learning is the process by which an agent improves its performance on future tasks after making observations about the world [1].

Key Components of Learning:

- **Task:** The problem to be solved
- **Experience:** Training data or interactions
- **Performance Measure:** Metric for success
- **Improvement:** Better performance over time

[3] provides mathematical foundations for learning algorithms.

Types of Learning

By Supervision:

- **Supervised Learning**

- Labeled training data
- Classification, regression

- **Unsupervised Learning**

- Unlabeled data
- Clustering, dimensionality reduction

- **Reinforcement Learning**

- Reward signals
- Sequential decision-making

By Approach:

- **Inductive Learning**

- Learning from examples

- **Deductive Learning**

- Logical inference

- **Transfer Learning**

- Applying knowledge across domains

- **Active Learning**

- Query-based learning

Importance of Learning in AI

Why Learning is Central to AI:

- ① **Adaptability:** Systems can handle new situations
- ② **Scalability:** Reduces need for manual programming
- ③ **Performance:** Improves accuracy over time
- ④ **Discovery:** Finds patterns humans might miss
- ⑤ **Automation:** Enables autonomous systems

Real-World Applications

- Medical diagnosis systems
- Autonomous vehicles
- Recommendation systems
- Speech recognition

Characteristics of Human Intelligence:

- **General Intelligence:** Ability to learn and adapt across diverse domains
- **Common Sense Reasoning:** Understanding everyday situations
- **Creativity:** Generating novel ideas and solutions
- **Emotional Intelligence:** Understanding and managing emotions
- **Social Intelligence:** Navigating complex social interactions
- **Transfer Learning:** Applying knowledge across contexts
- **Few-shot Learning:** Learning from limited examples

Humans excel at tasks requiring flexibility, intuition, and contextual understanding.

Characteristics of Machine Intelligence:

- **Narrow Intelligence:** Excellence in specific domains
- **Speed:** Processing vast amounts of data quickly
- **Consistency:** Uniform performance without fatigue
- **Scalability:** Handling massive computational tasks
- **Precision:** Exact calculations and pattern matching
- **Memory:** Perfect recall of training data

Current State

Most AI systems exhibit **narrow intelligence**, excelling in specific tasks but lacking general-purpose reasoning [1].

Comparison: Human vs Machine Intelligence

| Aspect | Human | Machine |
|-------------------|-----------------------|---------------------------|
| Learning Speed | Slow initial learning | Fast with sufficient data |
| Data Requirements | Few examples needed | Requires large datasets |
| Generalization | Excellent | Limited |
| Energy Efficiency | High (~20W) | Low (kW-MW) |
| Common Sense | Natural | Challenging |
| Calculation Speed | Slow | Very fast |
| Consistency | Variable | Highly consistent |
| Creativity | High | Limited |
| Explainability | Good | Often poor |

[4] discusses mathematical approaches bridging these differences.

Bridging the Gap

Approaches to Human-like AI:

① Cognitive Architectures

- Modeling human cognitive processes
- ACT-R, SOAR frameworks

② Neuromorphic Computing

- Brain-inspired hardware
- Spiking neural networks

③ Hybrid Systems

- Combining symbolic and sub-symbolic AI
- Neuro-symbolic integration

④ Meta-Learning

- Learning to learn
- Few-shot learning capabilities

What is an Intelligent Agent?

Definition

An **intelligent agent** is anything that can perceive its environment through sensors and act upon that environment through actuators [1].

Key Components:

- **Percepts:** Inputs from sensors at any given instant
- **Percept Sequence:** Complete history of percepts
- **Agent Function:** Maps percept sequences to actions
- **Agent Program:** Concrete implementation of agent function
- **Actuators:** Mechanisms for taking actions

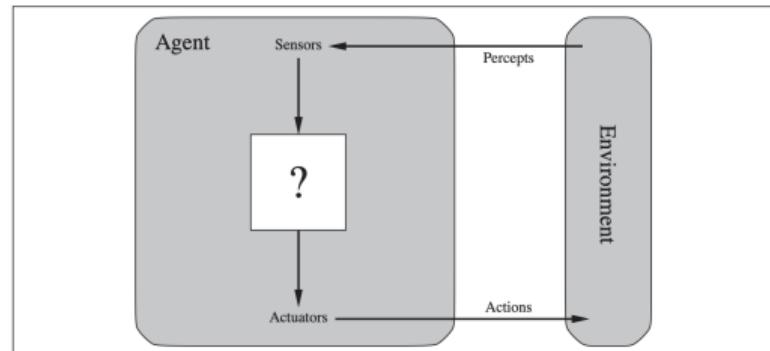


Figure: Figure 2.1 Agents interact with environments through sensors and actuators.

Agent Examples

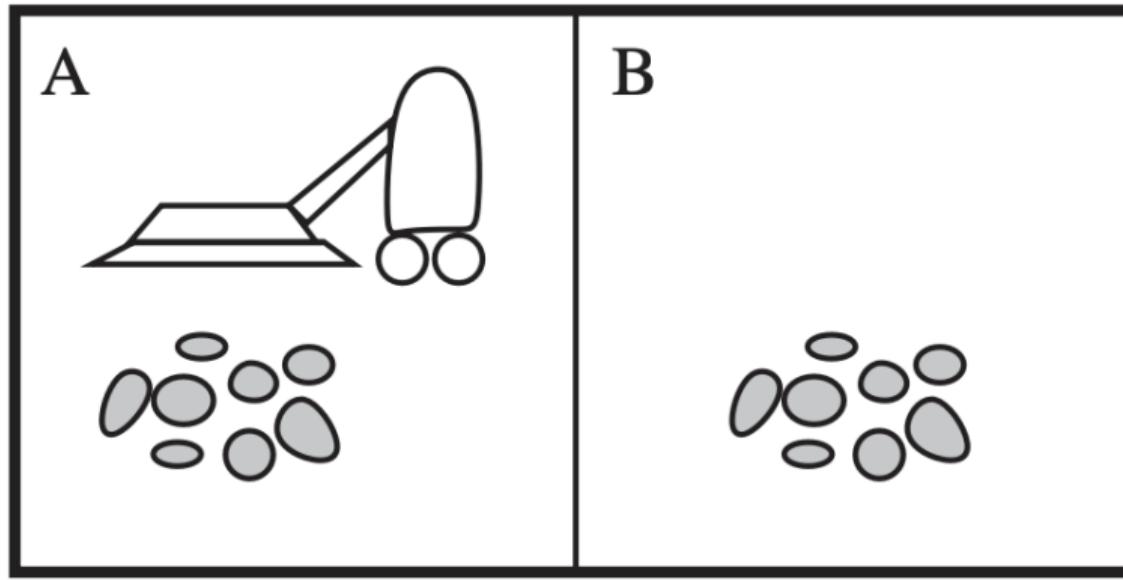


Figure: Figure 2.2 A vacuum-cleaner world with just two locations.

Agent Examples

| Percept sequence | Action |
|------------------------------------|--------|
| [A, Clean] | Right |
| [A, Dirty] | Suck |
| [B, Clean] | Left |
| [B, Dirty] | Suck |
| [A, Clean], [A, Clean] | Right |
| [A, Clean], [A, Dirty] | Suck |
| : | : |
| [A, Clean], [A, Clean], [A, Clean] | Right |
| [A, Clean], [A, Clean], [A, Dirty] | Suck |
| : | : |

Figure: Figure 2.3 Partial tabulation of a simple agent function for the vacuum-cleaner world shown in Figure 2.2.

Rationality

What is rational at any given time depends on four things:

- The performance measure that defines the criterion of success.
- The agent's prior knowledge of the environment.
- The actions that the agent can perform.
- The agent's percept sequence to date.

A **rational agent** selects actions that maximize its expected performance measure, given its percept sequence and built-in knowledge.

PEAS Description:

- **Performance Measure:** Criteria for success
- **Environment:** Task context
- **Actuators:** Available actions
- **Sensors:** Information sources

Example

Example: Autonomous Taxi

- **P:** Safe, fast, legal, comfortable, maximize profit
- **E:** Roads, traffic, pedestrians, weather
- **A:** Steering, accelerator, brake, signal
- **S:** Cameras, GPS, speedometer, sensors

| Agent Type | Performance Measure | Environment | Actuators | Sensors |
|---------------------------------|-------------------------------------|----------------------------------|---|---|
| Medical diagnosis system | Healthy patient, reduced costs | Patient, hospital, staff | Display of questions, tests, diagnoses, treatments, referrals | Keyboard entry of symptoms, findings, patient's answers |
| Satellite image analysis system | Correct image categorization | Downlink from orbiting satellite | Display of scene categorization | Color pixel arrays |
| Part-picking robot | Percentage of parts in correct bins | Conveyor belt with parts; bins | Jointed arm and hand | Camera, joint angle sensors |

Properties of Task Environments

- **Fully vs Partially Observable**
 - Complete vs incomplete information
- **Single vs Multi-agent**
 - Competitive, cooperative, mixed
- **Deterministic vs Stochastic**
 - Predictable vs uncertain outcomes
- **Episodic vs Sequential**
 - Independent vs dependent actions
- **Static vs Dynamic**
 - Environment changes during deliberation
- **Discrete vs Continuous**
 - Finite vs infinite states/actions
- **Known vs Unknown**
 - Rules known or must be learned

[1] emphasizes matching agent design to environment properties.

Formal Tasks in AI

Definition:

Tasks with well-defined rules, clear objectives, and measurable outputs.

Examples:

- Search problems (A*, BFS, DFS)
- Constraint satisfaction (Sudoku, graph coloring)
- Formal logic and reasoning
- Game playing (Chess, Go)
- Automated planning (STRIPS, PDDL)
- Machine learning tasks with clear metrics
- Program verification

STRIPS (STanford Research Institute Problem Solver) and its successor, PDDL (Planning Domain Definition Language)

Characteristics:

- Clear inputs and outputs
- Deterministic or probabilistic
- Easy to evaluate correctness

Mundane Tasks in AI

Definition:

Everyday human tasks involving perception, common sense, and intuition. Hard for machines.

Examples:

- Visual perception (face/object recognition)
- Speech recognition
- Natural language understanding
- Motor tasks (walking, grasping)
- Common-sense reasoning
- Contextual understanding (emotion, sarcasm)

Characteristics:

- Unstructured, ambiguous data
- Requires robust generalization
- Uses deep learning and perception models

Expert Tasks in AI

Definition:

Tasks requiring specialized domain knowledge and expert-level reasoning.

Examples:

- Medical diagnosis and radiology
- Legal document analysis
- Financial forecasting and trading
- Scientific discovery (e.g., protein folding)
- Engineering design and optimization
- Industrial automation and robotics

Characteristics:

- Domain-specific knowledge
- High accuracy and reliability
- Often uses hybrid symbolic + ML systems

Summary

| Category | Meaning | Examples | | Difficulty |
|----------|------------------------|-------------------------|--|-----------------------|
| Formal | Rule-based, structured | Search, logic, planning | | Moderate |
| Mundane | Everyday tasks | Vision, NLP, speech | | Very Hard |
| Expert | Professional tasks | Medicine, law | | Very Hard + High risk |

| Task Environment | Observable | Agents | Deterministic | Episodic | Static | Discrete |
|---------------------------|------------|--------|---------------|------------|---------|------------|
| Crossword puzzle | Fully | Single | Deterministic | Sequential | Static | Discrete |
| Chess with a clock | Fully | Multi | Deterministic | Sequential | Semi | Discrete |
| Poker | Partially | Multi | Stochastic | Sequential | Static | Discrete |
| Backgammon | Fully | Multi | Stochastic | Sequential | Static | Discrete |
| Taxi driving | Partially | Multi | Stochastic | Sequential | Dynamic | Continuous |
| Medical diagnosis | Partially | Single | Stochastic | Sequential | Dynamic | Continuous |
| Image analysis | Fully | Single | Deterministic | Episodic | Semi | Continuous |
| Part-picking robot | Partially | Single | Stochastic | Episodic | Dynamic | Continuous |
| Refinery controller | Partially | Single | Stochastic | Sequential | Dynamic | Continuous |
| Interactive English tutor | Partially | Multi | Stochastic | Sequential | Dynamic | Discrete |

Figure: Examples of task environments and their characteristics.

Types of Agents

1. Simple Reflex Agents

- Select actions based on current percept only
- Condition-action rules: "If dirty then vacuum"
- Fast but limited; no memory of past

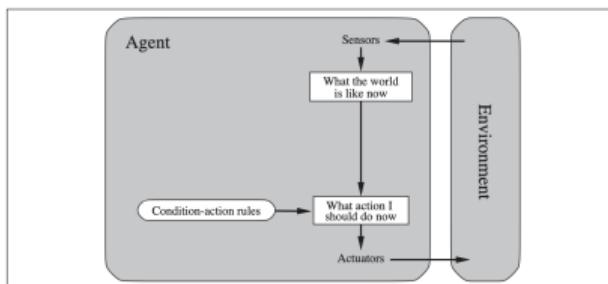


Figure: Schematic diagram of a simple reflex agent.

- **function** SIMPLE-REFLEX-AGENT(percept)
returns an action **persistent**: rules, a set of condition-action rules
 $\text{state} \leftarrow \text{INTERPRET-INPUT}(\text{percept})$
 $\text{rule} \leftarrow \text{RULE-MATCH}(\text{state}, \text{rules})$
 $\text{action} \leftarrow \text{rule.ACTION}$
return action

Types of Agents (2/5)

2. Model-Based Reflex Agents

- Maintain internal state/model of world
- Track aspects not directly observable
- Update state based on percepts and actions

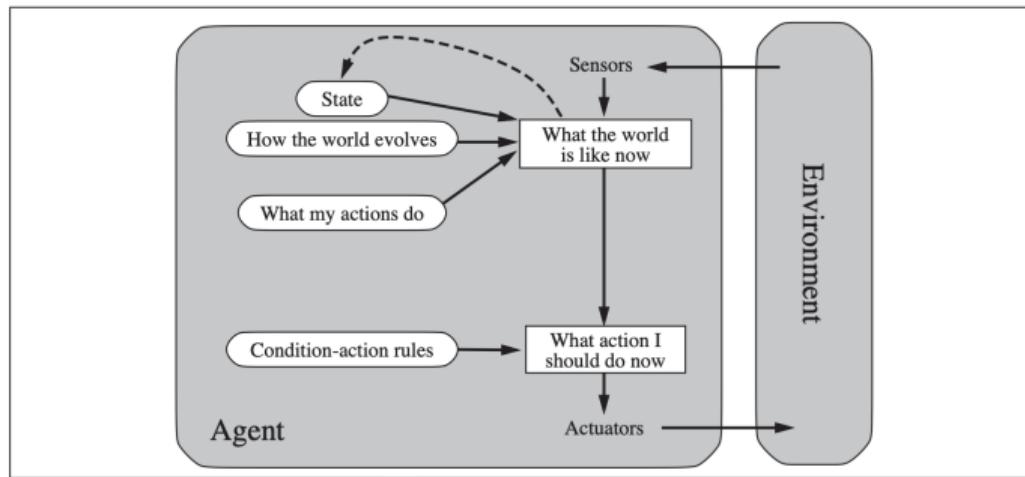


Figure: A model-based reflex agent.

Types of Agents (3/5)

3. Goal-Based Agents

- Act to achieve specified goals
- Consider future consequences of actions
- Search and planning for goal achievement

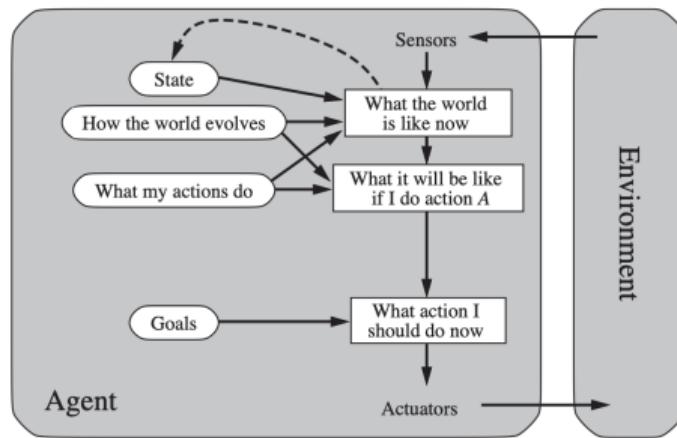


Figure: A model-based, goal-based agent. It keeps track of the world state as well as a set of goals it is trying to achieve, and chooses an action that will (eventually) lead to the achievement of its goals.

Types of Agents (4/5)

4. Utility-Based Agents

- Use utility function to evaluate states
- Handle conflicting goals and uncertainty
- Maximize expected utility

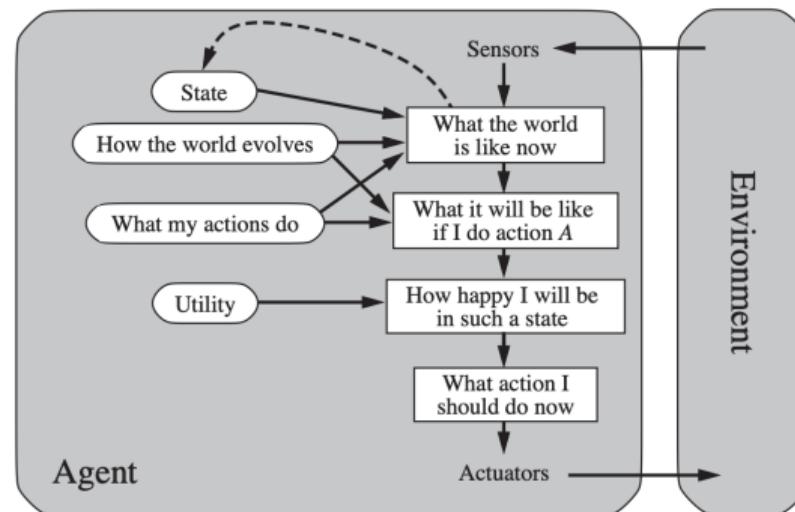


Figure: A model-based, utility-based agent.

Types of Agents (5/5)

5. Learning Agents

- Improve performance through experience
- Components: learning element, performance element, critic, problem generator
- Most general and capable agent type

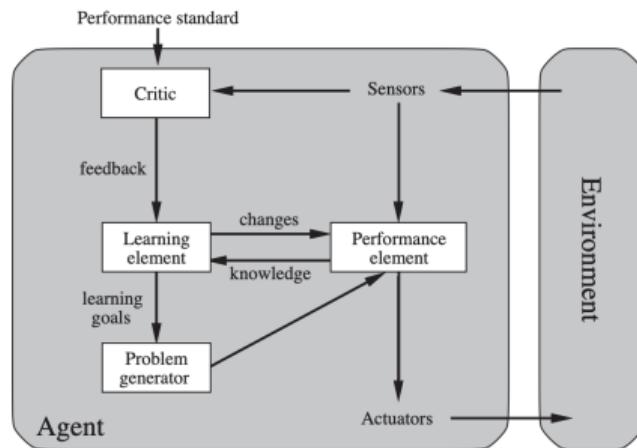


Figure: A general learning agent.

Hierarchy

Each agent type builds upon and extends the capabilities of simpler types.

Implementation Approaches:

- **Reactive Architectures**

- Direct perception-action mapping
- Subsumption architecture (Brooks)

- **Deliberative Architectures**

- Symbolic reasoning and planning
- Explicit world models

- **Hybrid Architectures**

- Combine reactive and deliberative layers
- Three-layer architecture (Sensory, Planning, Behavior)

- **Layered Architectures**

- Multiple levels of abstraction
- Hierarchical control

Subsumption architecture is a reactive robotic architecture heavily associated with behavior-based robotics which was very popular in the 1980s and 90s.

Summary

Key Takeaways:

- AI seeks to create intelligent behavior in machines through multiple approaches
- AI is inherently interdisciplinary, drawing from mathematics, computer science, neuroscience, and more
- Knowledge representation and learning are fundamental to AI systems
- Human and machine intelligence have complementary strengths and weaknesses
- Intelligent agents provide a unifying framework for understanding AI systems
- Different agent types suit different task environments

Looking Forward

Modern AI combines these foundational concepts with powerful computational resources and big data to create increasingly capable systems.

References I

-  Russell, S., Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
-  Rich, E., Knight, K., Nair, S. B. (2009). *Artificial Intelligence* (3rd ed.). McGraw-Hill.
-  Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
-  Deisenroth, M. P., Faisal, A. A., Ong, C. S. (2020). *Mathematics for Machine Learning*. Cambridge University Press.

Thank You!

Questions?