AKTU · 3RD YEAR · SEMESTER 5 · BCAI501

Artificial Intelligence

Unit 4 — Software Agents

- Architecture for Intelligent Agents
- Agent Communication
- Negotiation and Bargaining
- Argumentation among Agents
- Trust and Reputation in Multi-agent Systems



DECODED BY

Chirag Jain

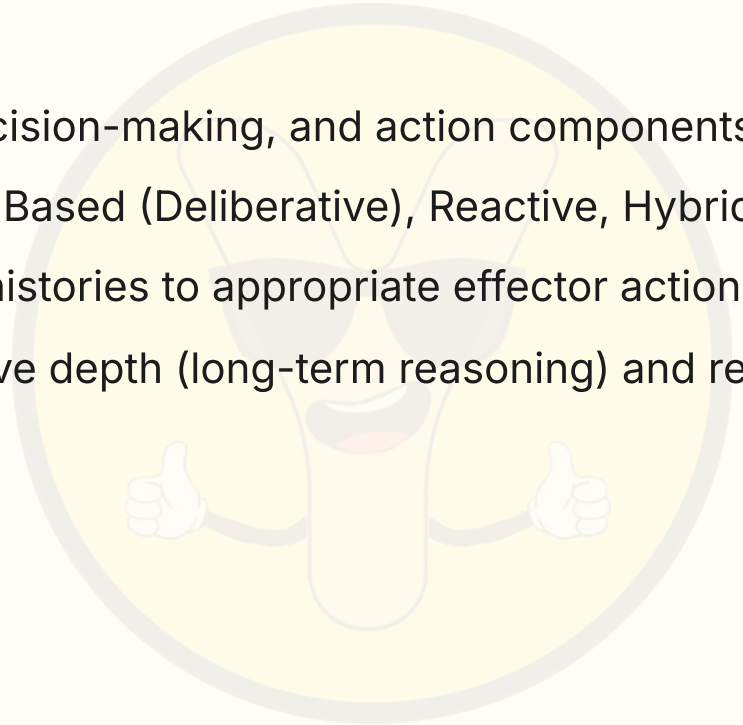
Founder, Ycotes · Gold Medalist in Artificial Intelligence, IIT Madras · taught 10,000+ students

theycotes.com

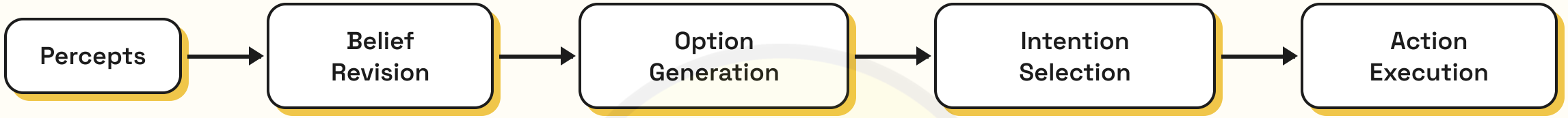
Architecture for Intelligent Agents

Agent architectures define the software structure, data flows, and reasoning mechanisms powering autonomous agents.

- Architectures organize sensing, decision-making, and action components into concrete software models.
- Key architectural paradigms: Logic-Based (Deliberative), Reactive, Hybrid, and BDI architectures.
- Primary goal: Map sensor percept histories to appropriate effector actions effectively.
- Trade-offs exist between deliberative depth (long-term reasoning) and reactive speed (real-time response).



Belief-Desire-Intention (BDI) Architecture Flow



Sequential data processing flow inside a standard BDI agent execution cycle.

Classification of Agent Architectures

Intelligent agent software architectures are broadly categorized into four primary paradigms.

- Logic-Based (Deductive)
  - Uses formal logic for symbolic reasoning
  - Proves theorems to determine optimal actions
- Reactive (Subsumption)
  - Direct mapping from percepts to actions
  - No explicit internal world model or complex reasoning
- Belief-Desire-Intention (BDI)
  - Mental state modeling (Beliefs, Desires, Intentions)
  - Balances goal commitment with plan adaptability
- Hybrid Architectures
  - Combines reactive layers (speed) with deliberative layers (planning)
  - Examples: Interrap, TouringMachines



Reactive vs Deliberative Architecture

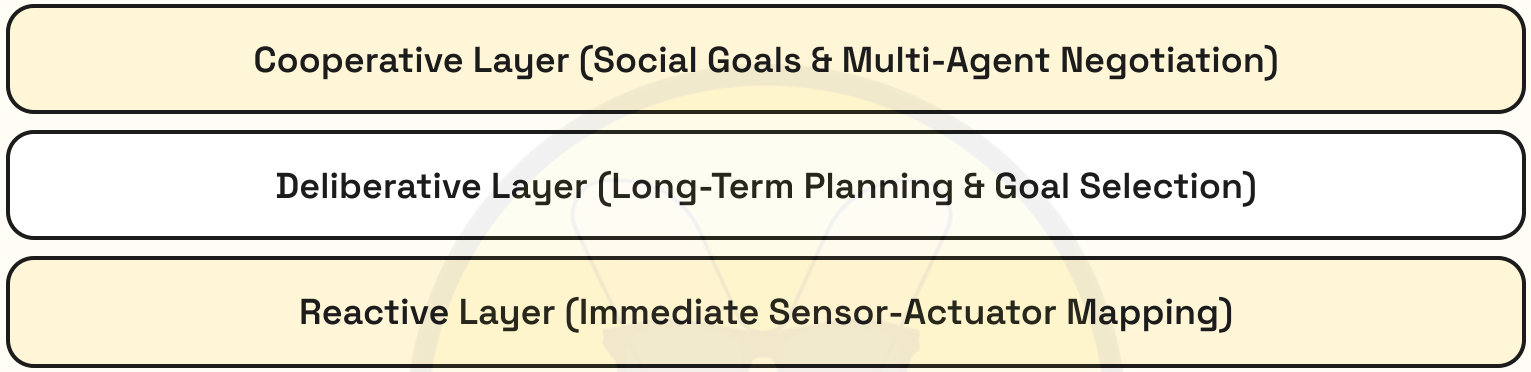
Parameter	Reactive Architecture	Deliberative Architecture
Internal Model	None or minimal state representation	Explicit symbolic world model
Response Time	Very fast (Real-time performance)	Slower (Reasoning bottleneck)
Reasoning Type	Condition-Action rules / Subsumption	Logic deduction and state search
Dynamic Environments	Highly robust and adaptive	Struggles if environment changes fast
Complex Planning	Poor long-term goal planning	Excellent long-term plan generation

BDI Architecture Mechanics

The BDI model mimics human problem-solving by organizing agent state into three mental attitudes.

- Beliefs B: Represents information the agent has about the current state of the environment.
- Desires D: Represents all possible goals or outcomes the agent would like to achieve.
- Intentions I: Represents chosen desires the agent has committed to executing.
- Intention Filter: Prevents re-planning at every step; agent executes plan until failure or completion.
- Plan Library: Pre-compiled procedural plans selected based on current active intentions.

Layered Hybrid Agent Architecture



Three-tiered Interrap hybrid architecture balancing fast reactions with high-level deliberation.

Agent Communication & Speech Act Theory

SPEECH ACT THEORY

A pragmatic theory of language proposed by Austin and Searle where communication is treated as actions (speech acts) that change the mental state of recipient agents.

- Locutionary Act: The physical act of uttering a meaningful string of words.
- Illocutionary Act: The underlying intention or force of the speaker (e.g., requesting, informing, promising).
- Perlocutionary Act: The actual psychological or behavioral effect on the receiver.
- Forms the theoretical foundation for standard Agent Communication Languages (ACL).

Agent Communication Languages: KQML & FIPA-ACL

ACLs provide standardized message formats allowing autonomous agents to exchange knowledge.

- KQML (Knowledge Query and Manipulation Language): Pioneered message-passing primitives called performatives.
- FIPA-ACL: The current international standard developed by FIPA with formal modal-logic semantics.
- Performatives: Communicative verbs indicating intent (e.g., inform, request, propose, agree, refuse).
- Message Components: Communicative act, sender, receiver, content, ontology, protocol, language.
- Ontologies: Shared vocabularies enabling agents to understand content domain semantics.

Comparison: KQML vs FIPA-ACL

Feature	KQML	FIPA-ACL
Standardization Body	ARPA Knowledge Sharing Effort	FIPA / IEEE Standards Association
Formal Semantics	Informal / Ad-hoc definition	Formal Modal Logic (SL Language)
Core Performatives	evaluate, ask-one, tell, reply	inform, request, propose, agree, refuse
Content Representation	Knowledge Interchange Format (KIF)	Semantic Language (SL), RDF, XML
Protocol Support	Basic message routing	Built-in interaction protocols (CNP, Auction)

Contract Net Protocol (CNP)

A high-level agent interaction protocol for distributed task allocation among autonomous agents.

- Manager Agent: Identifies task, broadcasts Request for Proposal (RFP) to potential contractors.
- Contractor Agents: Evaluate task feasibility and return bids specifying cost/quality parameters.
- Bid Evaluation: Manager evaluates bids against internal utility functions.
- Awarding: Manager selects best contractor and sends acceptance; rejects remaining bidders.
- Execution & Result: Selected contractor executes task and returns final output to manager.

Negotiation and Bargaining in MAS

AGENT NEGOTIATION

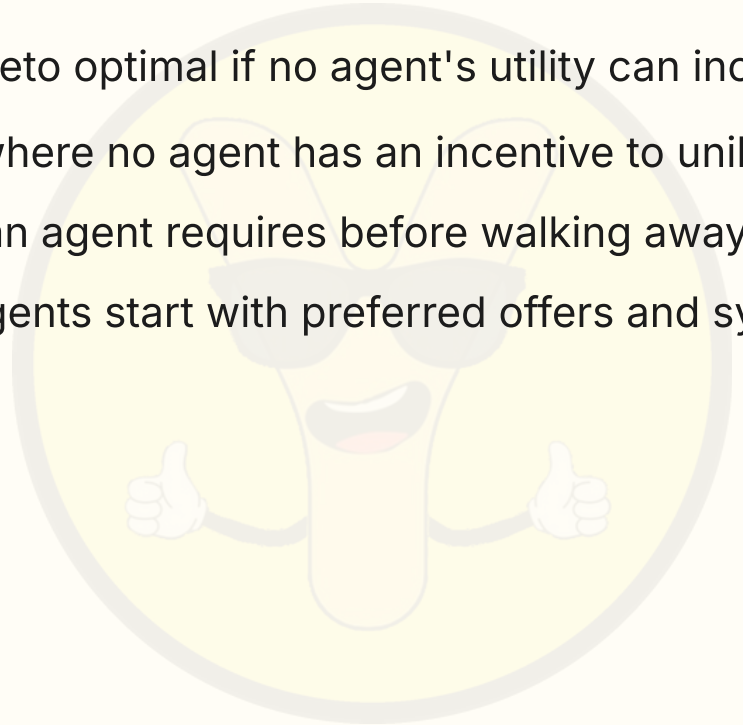
The process by which a group of autonomous agents attempt to reach a mutually acceptable agreement over a set of issues through structured proposal exchange.

- Negotiation Domain: Set of possible agreements or outcomes Ω.
- Utility Function: Function $U_i : \Omega \to \mathbb{R}$ mapping outcomes to agent preferences.
- Negotiation Protocol: Standardized rules governing turn-taking, proposals, and termination.
- Negotiation Strategy: Private algorithm used by an agent to determine bids and concessions.

Game-Theoretic Bargaining Concepts

Mathematical foundation evaluating the optimality and stability of negotiation outcomes.

- Pareto Efficiency: Outcome x is Pareto optimal if no agent's utility can increase without reducing another's.
- Nash Equilibrium: Strategy profile where no agent has an incentive to unilaterally change strategy.
- Reservation Price: Minimum utility an agent requires before walking away from negotiation.
- Monotonic Concession Protocol: Agents start with preferred offers and systematically concede utility over time.



Negotiation Utility Calculation

Two agents negotiate the division of 2 resource bundles. Bundle X gives utilities $U_A(X) = 8, U_B(X) = 2$. Bundle Y gives $U_A(Y) = 5, U_B(Y) = 6$. Bundle Z gives $U_A(Z) = 3, U_B(Z) = 4$. Identify Pareto optimal outcomes.

1. Step 1: Compare Bundle X $(8, 2)$ and Bundle Z $(3, 4)$: Neither strictly dominates the other.
2. Step 2: Compare Bundle Y $(5, 6)$ and Bundle Z $(3, 4)$: $U_A(Y) > U_A(Z)$ and $U_B(Y) > U_B(Z)$. Thus Y strictly dominates Z.
3. Step 3: Compare Bundle X $(8, 2)$ and Bundle Y $(5, 6)$: Increasing A's utility reduces B's utility. Neither dominates the other.
4. Step 4: Remove dominated Bundle Z from consideration.

The Pareto optimal set of outcomes consists of Bundle X $(8, 2)$ and Bundle Y $(5, 6)$.

Argumentation among Agents

ABSTRACT ARGUMENTATION FRAMEWORK

Formulated by Dung (1995), a framework represented as a pair $AF = \langle AR, Att \rangle$, where AR is a set of arguments and $Att \subseteq AR \times AR$ is a binary attack relation.

- Conflict-free Set: A set $S \subseteq AR$ is conflict-free if no argument in S attacks another argument in S.
- Defense / Acceptability: A set S defends argument a if S attacks every argument b that attacks a.
- Admissible Extension: A conflict-free set S that defends all its elements.
- Grounded Extension: The unique minimal admissible extension under set inclusion (least fixed point).

Computing Dung's Argumentation Extensions

Given Argumentation Framework $AF = \langle \{a, b, c\}, \{(a, b), (b, c)\} \rangle$, compute the Conflict-free sets, Admissible sets, and Grounded extension.

1. Step 1: List all Conflict-free sets (no internal attacks): $\emptyset, \{a\}, \{b\}, \{c\}, \{a, c\}$.
2. Step 2: Identify unattacked arguments: Argument a receives no attacks.
3. Step 3: Check defense: Argument a attacks b. Since b attacks c, argument a defends c.
4. Step 4: Form Admissible sets (conflict-free and self-defending): $\emptyset, \{a\}, \{a, c\}$.
5. Step 5: Compute Grounded extension: Start with unattacked $\{a\}$, add defended $c \rightarrow \{a, c\}$.

Admissible sets: $\emptyset, \{a\}, \{a, c\}$. Grounded Extension: $\{a, c\}$.

Argumentation Dialogues and Protocols

Agents engage in structured dialogues to resolve conflicting beliefs, goals, or plans.

- Persuasion Dialogue: One agent seeks to convince another to accept a proposition.
- Inquiry Dialogue: Agents collaborate to establish truth or answer a shared question.
- Deliberation Dialogue: Agents decide on the optimal joint plan or course of action.
- Negotiation Dialogue: Agents bargain over limited resources using logical arguments.
- Dialogue Moves: Assert, Challenge, Defend, Accept, Retract.

Trust and Reputation in MAS

TRUST VS REPUTATION

Trust is an agent's subjective belief in the reliability of a partner based on direct experiences, whereas Reputation is an aggregated public evaluation derived from recommendations across the agent community.

- Direct Trust: Computed directly from past personal interactions with a target agent.
- Indirect Reputation: Aggregated feedback shared by third-party observer agents.
- Purpose: Enables safe interaction in open, decentralized dynamic multi-agent networks.
- Dynamic Update: Scores increment with successful interactions and decay over time or failure.

Beta Reputation System Formulation

Uses the Beta probability distribution function $Beta(\alpha, \beta)$ to calculate trust scores.

- Feedback Counters: Track successful outcomes r and unsuccessful outcomes s.
- Distribution Parameters: $\alpha = r + 1$ and $\beta = s + 1$ (assuming uniform prior).
- Expected Reputation Score: Computed as $E(R) = \frac{\alpha}{\alpha+\beta} = \frac{r+1}{r+s+2}$.
- Uncertainty: Variance decreases as total interactions $(r + s)$ increase.
- Recency Weighting: Applies a decay factor $\gamma \in (0, 1]$ to prioritize recent interactions.

Calculating Beta Reputation Score

An e-commerce agent has completed 18 positive transactions and 2 negative transactions with a client. Prior parameters are $\alpha_0 = 1, \beta_0 = 1$. Compute the updated expected trust score $E(R)$.

1. Step 1: Identify given values: Positive $r = 18$, Negative $s = 2$.
2. Step 2: Calculate parameter $\alpha = r + \alpha_0 = 18 + 1 = 19$.
3. Step 3: Calculate parameter $\beta = s + \beta_0 = 2 + 1 = 3$.
4. Step 4: Substitute into Beta expectation formula: $E(R) = \frac{\alpha}{\alpha+\beta} = \frac{19}{19+3}$.
5. Step 5: Compute final value: $E(R) = \frac{19}{22} \approx 0.8636$.

The updated expected trust score is 0.8636 (or 86.36%).

Reputation Attacks & Countermeasures

Attack Type	Threat Mechanism	Defense Strategy
Sybil Attack	Attacker creates fake virtual identities to inflate ratings	Identity verification, cryptographic certificates, entrance fees
Bad-Mouthing	Malicious agents broadcast false negative ratings	Weighted aggregation based on recommender trust, outlier filtering
Ballot-Stuffing	Colluding agents send false positive ratings to boost score	Transaction proof requirement, statistical anomaly detection
Whitewashing	Agent resets identity to erase accumulated bad history	High entry fee for new agents, low default initial trust assignment
On-Off Attack	Agent behaves well intermittently to maintain average trust	Asymmetric update (heavy penalty for failure, slow reward for success)

Exam Corner: High-Yield AKTU Questions

- **Q1** Explain the Belief-Desire-Intention (BDI) agent architecture with a clean block diagram and execution cycle.
- **Q2** Compare Reactive and Deliberative agent architectures on response speed, internal model, and planning capabilities.
- **Q3** Differentiate between KQML and FIPA-ACL. Explain speech act theory and performatives with suitable examples.
- **Q4** Describe the Contract Net Protocol (CNP) steps for task distribution in a Multi-Agent System.
- **Q5** Define Pareto Efficiency and Nash Equilibrium in the context of multi-agent negotiation and bargaining.
- **Q6** Explain Dung's Abstract Argumentation Framework. Define Conflict-free, Admissible, and Grounded extensions.
- **Q7** Detail the Beta Reputation System model. Calculate trust score given 14 positive and 6 negative interactions.
- **Q8** Write short notes on Sybil Attack, Whitewashing, and Bad-Mouthing attacks in multi-agent trust models.

Unit 4 done. You are exam-ready.

Revise the Exam corner once more, then open Unit 5 on theycotes.com. Every AKTU subject, unit by unit, free. Notes by Chirag Jain, Ycotes.

