

# Agentic AI for Information Retrieval

CHIIR 2026 Tutorial

From Ranked Lists to Goal-Oriented Information Seeking

Tanya Roosta and Preetam Dammu

# Tutorial Overview

In this tutorial we will cover:

1. Information seeking and the agentic shift
2. Foundations of agentic AI
3. Agentic workflows for IR
4. Prompting and planning
5. Evaluation methods
6. Case studies and open research questions

# Module 1: Information Seeking and the Agentic Shift

The background of the slide features abstract, overlapping geometric shapes in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are primarily located on the right side of the slide, creating a modern, dynamic aesthetic.

# Introduction to Module 1

- ▶ Information retrieval systems have traditionally focused on:
  - Query understanding
  - Document ranking
  - Relevance estimation
- ▶ However, many real-world information needs require:
  - Multi-step reasoning
  - Comparison of sources
  - Synthesis of evidence

# Traditional IR Workflow

- ▶ Classic IR interaction loop:
  1. User formulates query
  2. System returns ranked list
  3. User inspects results
  4. User clicks documents
  5. User reformulates query
- ▶ This process may repeat many times before the user finds an answer.

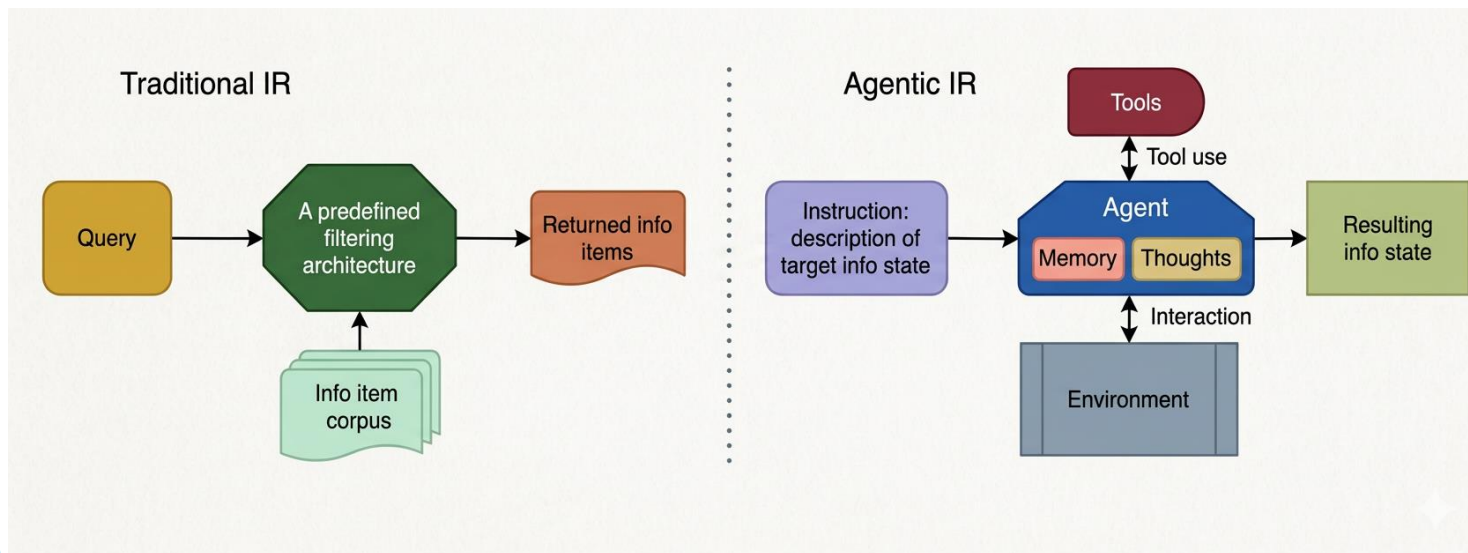


# Beyond Ranked Lists

- ▶ **Ranked lists:**
  - Users must compare multiple sources.
  - Users must synthesize evidence.
  - This places high cognitive load on users.
- ▶ **Agentic shift: transitioning from passive retrieval to active goal pursuit:**
  - Goal-driven workflows.
  - Agents break down objectives and execute actions across multiple steps.
  - Evidence is synthesized automatically.

# The Monolithic to Agentic Shift

- ▶ In the era of LLM-driven AI agents, the definition of “Information” in IR is evolving from the information items from the corpus to the information states from the wild\*.
- ▶ Agentic systems shift the paradigm from Document retrieval To Goal-directed problem solving.
- ▶ The unit of interaction becomes **task completion**, not just retrieval.



# Module 2: Agentic AI Foundations

The background features abstract, overlapping geometric shapes in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are primarily located on the right side of the frame, creating a modern, tech-oriented aesthetic.

# Definition of Agentic AI

AI systems that can **perceive**, **reason**, **make decisions**, and **act autonomously** in pursuit of goals.

## Beyond Reactive

Agentic AI goes beyond reactive AI, which simply responds to situations. It actively pursues goals and objectives, adapting its actions based on feedback and learning from experience.

## Proactive and Adaptive

Instead of passively waiting for instructions, agentic AI systems are proactive and adaptive. They can anticipate needs, generate solutions, and autonomously make decisions within defined parameters.

# Characteristics

## 1 Goal-Oriented

Agentic AI systems are designed with specific goals in mind, driving their actions and decisions.

## 3 Self-Learning

Agentic AI continuously learns and improves its performance based on data and feedback.

## 2 Autonomous

They can operate independently, making decisions and taking actions without constant human intervention.

## 4 Contextual Awareness

They can understand and adapt to changing contexts, making decisions based on real-time information.

# Benefits of AI Agents

## Increased Efficiency

Agentic AI can automate tasks, optimize processes, and free up human resources for more strategic work.

## Enhanced Productivity

By making intelligent decisions and taking proactive actions, agentic AI can boost productivity across industries.

## Improved Decision-Making

Leveraging data analysis and predictive modeling, agentic AI can support better informed decision-making.

## Personalized Experiences

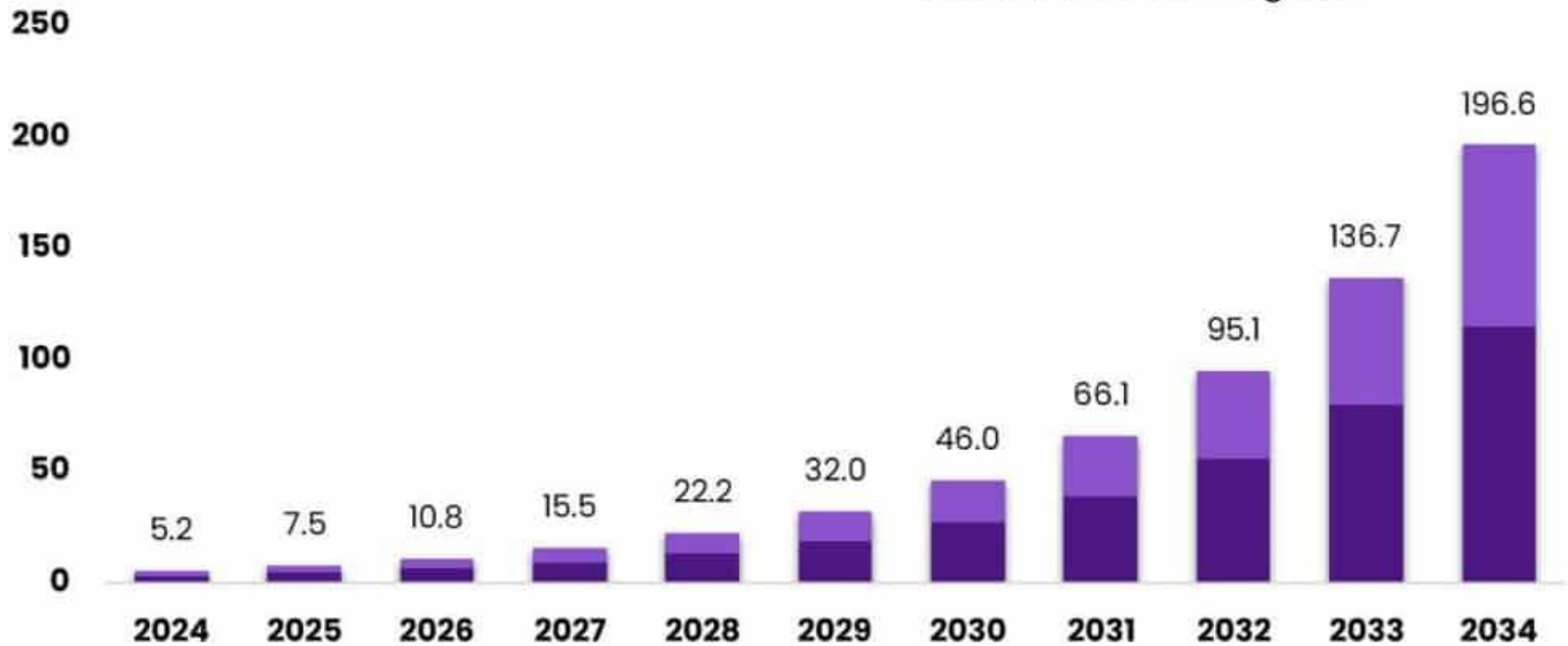
Agentic AI can tailor experiences to individual needs and preferences, leading to greater customer satisfaction.

# Importance of AI Agents

- ▶ The global AI market has a projected value of \$1.81 trillion by 2030, exhibiting a compound annual growth rate (CAGR) of 35.9% from 2025 to 2030
- ▶ AI agents enhance operational efficiency, which can lead to productivity increases of up to 30% in industries such as manufacturing and logistics.
  - Example: AI-driven predictive maintenance can reduce equipment downtime by 40%, significantly lowering operational costs.

# Global Agentic AI Market

Size, by Product Type, 2025-2034 (USD Billion)



The Market will Grow  
At the CAGR of:

**43.8%**

The Forecasted Market  
Size for 2034 in USD:

**\$196.6B**

**market.us**  
ONE STOP SHOP FOR THE REPORTS

# Historical Context

The background of the slide is white with abstract, overlapping geometric shapes in various shades of blue (light blue, medium blue, and dark blue) on the right side, creating a modern, dynamic look.

# Early Foundations of Agentic AI (1950s-1980s)

- ▶ **Turing (1950):** "Computing Machinery and Intelligence" introduced the idea of machines exhibiting intelligent behavior.
- ▶ **Early Rule-Based Systems (1950s-1970s):** These systems operated on predefined rules to perform specific tasks, often in well-defined domains like game playing (e.g., checkers).
- ▶ **Emergence of Agent Concepts (1950s-1980s):** The concept of intelligent agents began to take shape, with researchers exploring systems that could sense, reason, and act in an environment. John McCarthy and Oliver Selfridge at MIT were among the first to describe the concept of intelligent agents.
- ▶ **Limitations:**
  - Brittle systems; lacked robustness and adaptability in open-ended environments.
  - No unified notion of autonomy, goal generation, or learning from experience.

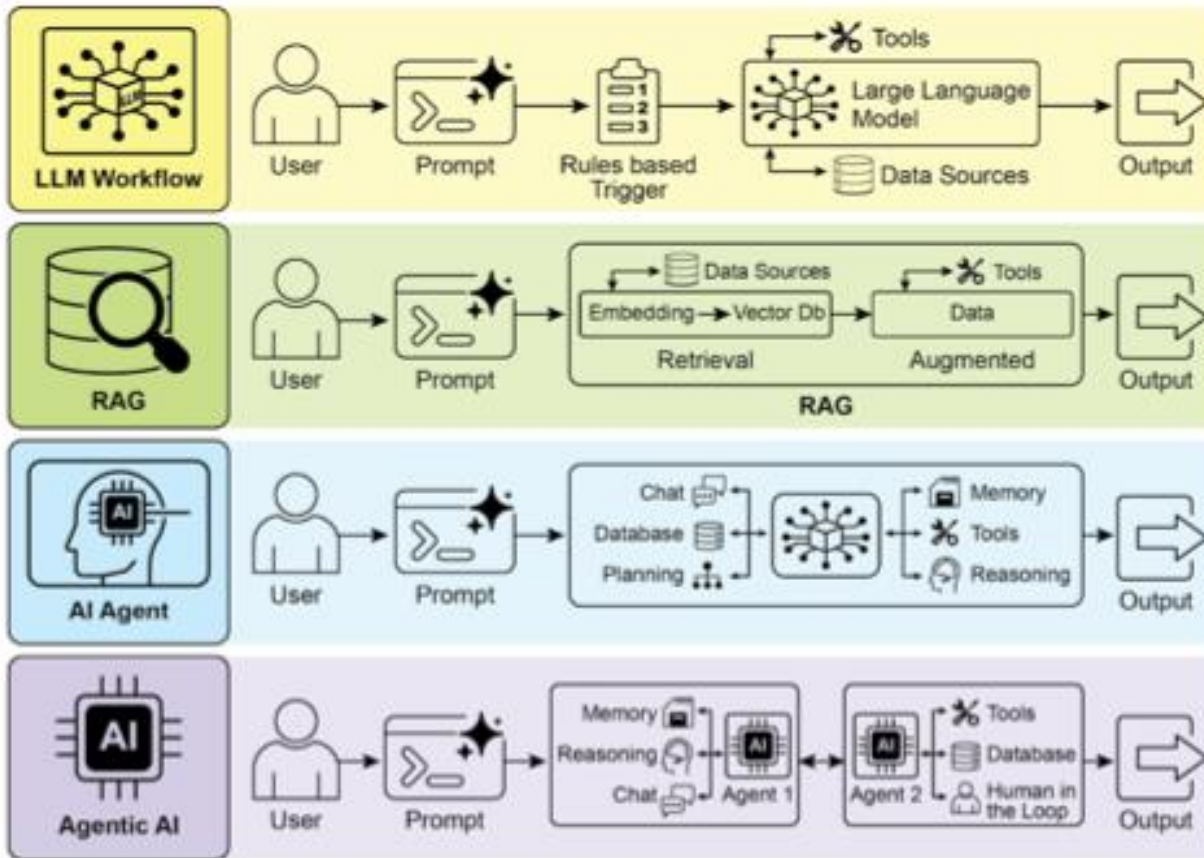
# Rise of Rational and Reactive Agents (1980s-2000s)

- ▶ **Rational Agent Model (Russell & Norvig):** Agents perceive, reason, and act to maximize utility.
- ▶ **BDI Architecture (Belief-Desire-Intention):** Popular framework for modeling human-like decision-making.
- ▶ **Reactive Agents (Brooks, 1986):** Emphasis on situated action, behavior-based robotics.
- ▶ **Key Developments:**
  - Autonomous robots and game-playing agents (e.g., RoboCup, Deep Blue).
  - Introduction of **multi-agent systems** (negotiation, cooperation, competition).

# Modern Agentic AI (2010s-Present)

- ▶ Deep Reinforcement learning and Imitation Learning: Agents learn policies from trial and error (e.g., AlphaGo, OpenAI Five).
- ▶ Cognitive Architectures: Integration of memory, planning, and self-reflection (e.g., ACT-R, SOAR\*).
- ▶ LLM-powered Agents (2020s): Autonomous task agents (e.g., AutoGPT, BabyAGI) with reasoning chains and tool use.
- ▶ Emerging Trends:
  - Agents as co-workers (e.g., AI teammates, assistants).
  - Growing focus on agency, alignment, self-improvement, and autonomy under uncertainty.

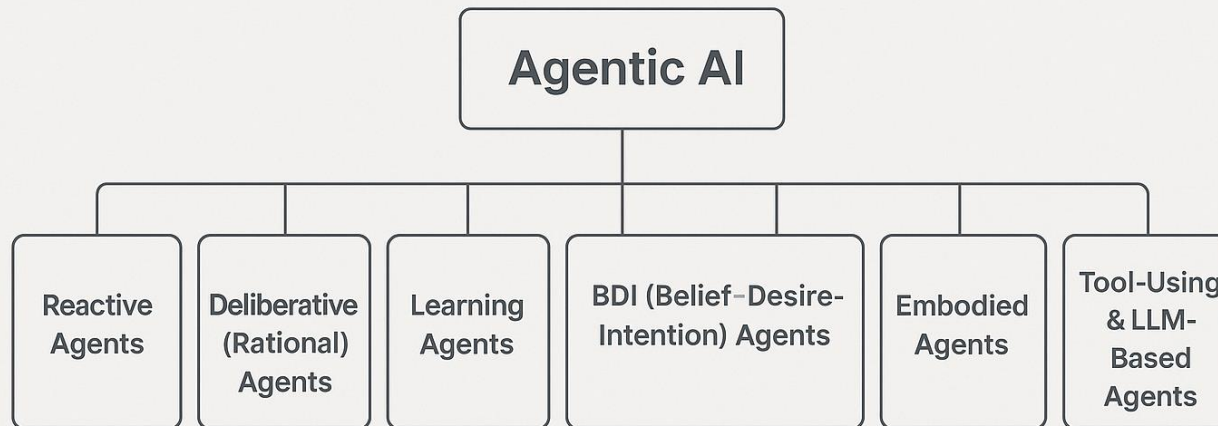
# And in the last 2-3 years...



# Types of Agentic AI

The background of the slide is white with abstract, overlapping geometric shapes in various shades of blue (light blue, medium blue, and dark blue) on the right side, creating a modern, tech-oriented aesthetic.

# Types of Agentic AI

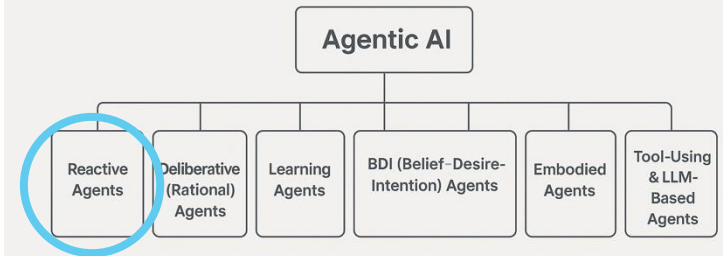


# Reactive Agents

## ▶ Reactive Agents

- **Behavior:** Respond to current stimuli; no internal state or memory.
- **Examples:** Braitenberg vehicles, simple robotic agents.
- **Strengths:** Fast, robust in predictable environments.
- **Limitations:** Can't learn, plan, or reflect on experience.

## Types of Agentic AI

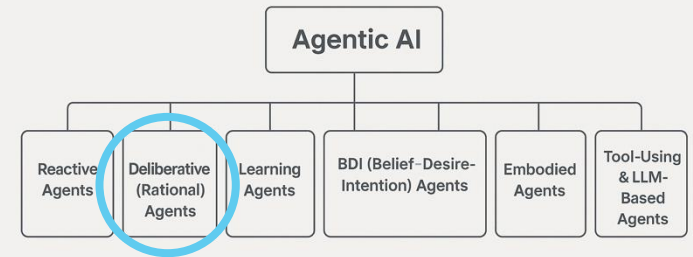


# Rational Agents

## ▶ Deliberative (Rational) Agents

- **Behavior:** Build and use internal models to plan actions.
- **Core Components:** Perception → Belief update → Planning → Action.
- **Examples:** Chess-playing AIs (e.g., Deep Blue), GOFAI planning systems.
- **Limitations:** Struggle in dynamic or uncertain environments.

## Types of Agentic AI

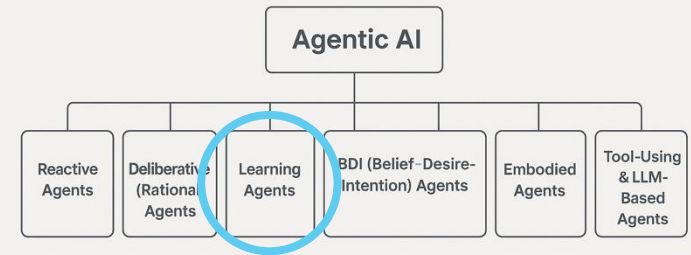


# Learning Agents

## ▶ Learning Agents

- **Behavior:** Improve performance through experience (exploration + exploitation).
- **Techniques:** Reinforcement Learning, imitation Learning.
- **Examples:** AlphaGo, OpenAI Five, autonomous driving systems.
- **Strengths:** Adaptable, scalable in complex tasks.
- **Risks:** Sample inefficiency, safety in real-world deployment.

## Types of Agentic AI

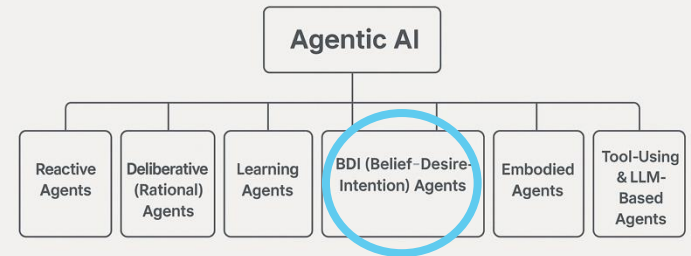


# BDI Agents

## ▶ BDI (Belief-Desire-Intention) Agents

- **Behavior:** Modeled after human reasoning; maintain beliefs, desires (goals), and intentions (committed plans).
- **Examples:** simulation-based agents in games or training.
- **Uses:** Multi-agent systems, decision-support systems.

## Types of Agentic AI

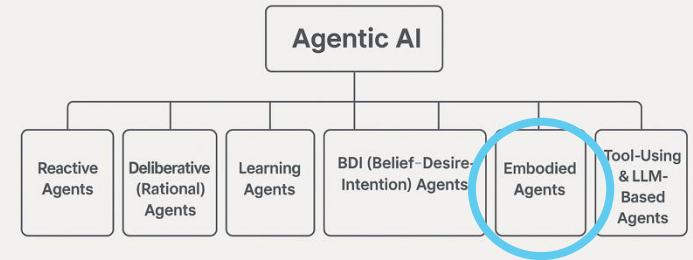


# Embodied Agents

## ▶ Embodied Agents

- **Definition:** Agents with a physical or virtual body interacting in a physical/simulated world.
- **Examples:** Robots, virtual avatars, embodied conversational agents.
- **Focus:** Sensory grounding, human-agent interaction.

## Types of Agentic AI

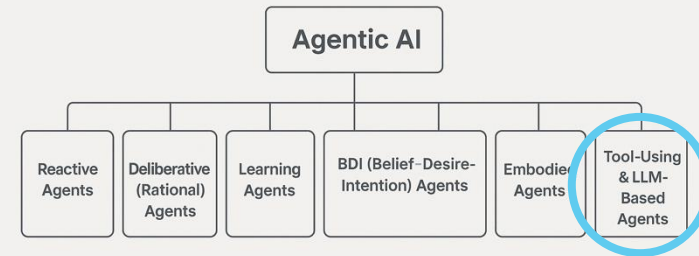


# LLM-Based Agents

## ▶ Tool-Using and LLM-Based Agents

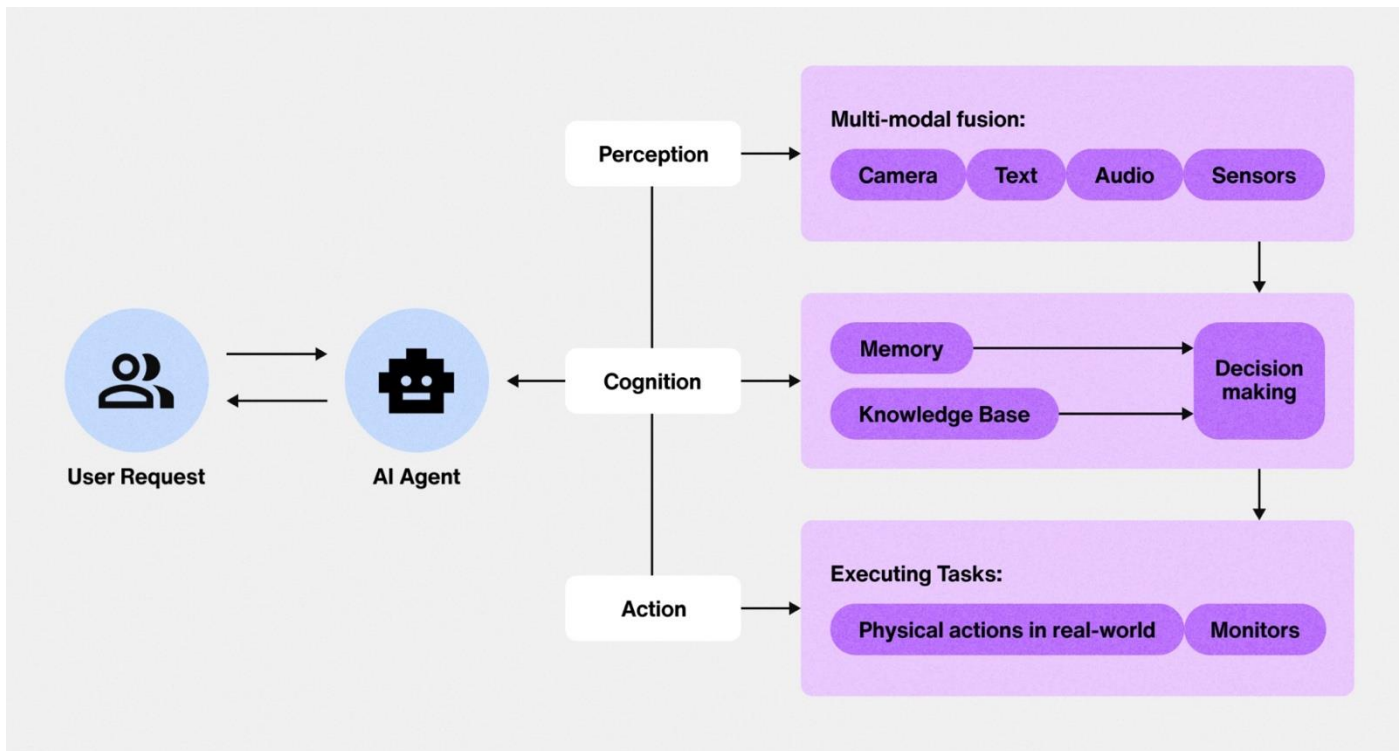
- LLMs orchestrating actions using APIs/tools (e.g., web search, code execution).
- **Examples:** AutoGPT, BabyAGI, ChatGPT agent.
- **Features:** Planning, memory, reflection, autonomy in open-ended tasks.
- **Challenges:** Alignment, hallucination, evaluation of outcomes.

## Types of Agentic AI



# Agentic AI Platform Reference Architecture

The background features abstract, overlapping geometric shapes in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are primarily located on the right side of the slide, creating a modern, tech-oriented aesthetic.



# Profiling

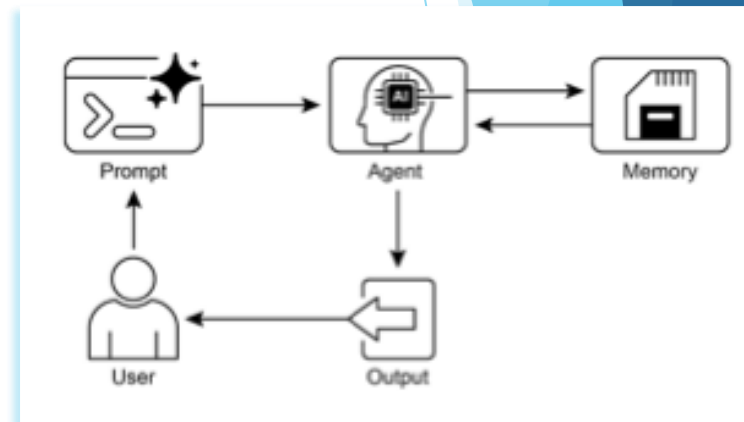
- ▶ Profiling is where an agent is assigned roles that shape its behavior. It defines the agent's identity, objectives, and boundaries, setting guidelines for interactions with users and systems.
- ▶ Examples:
  - "Coder Agent" can handle repetitive programming tasks, debug code, or create new scripts, e.g. GitHub Copilot.
  - "Content Creator Agent" might draft articles or design graphics.
  - "Project Manager Agent" could prioritize tasks and manage workflows.

# Knowledge

- ▶ The agent's knowledge provides domain-specific expertise, helping it understand tasks and make decisions based on factual data.
- ▶ It typically consists of semantic memory (general facts, concepts, and rules about the world), and instructions for handling queries.
- ▶ This is often achieved using pre-trained AI models, structured knowledge bases, and mechanisms for continuous learning.

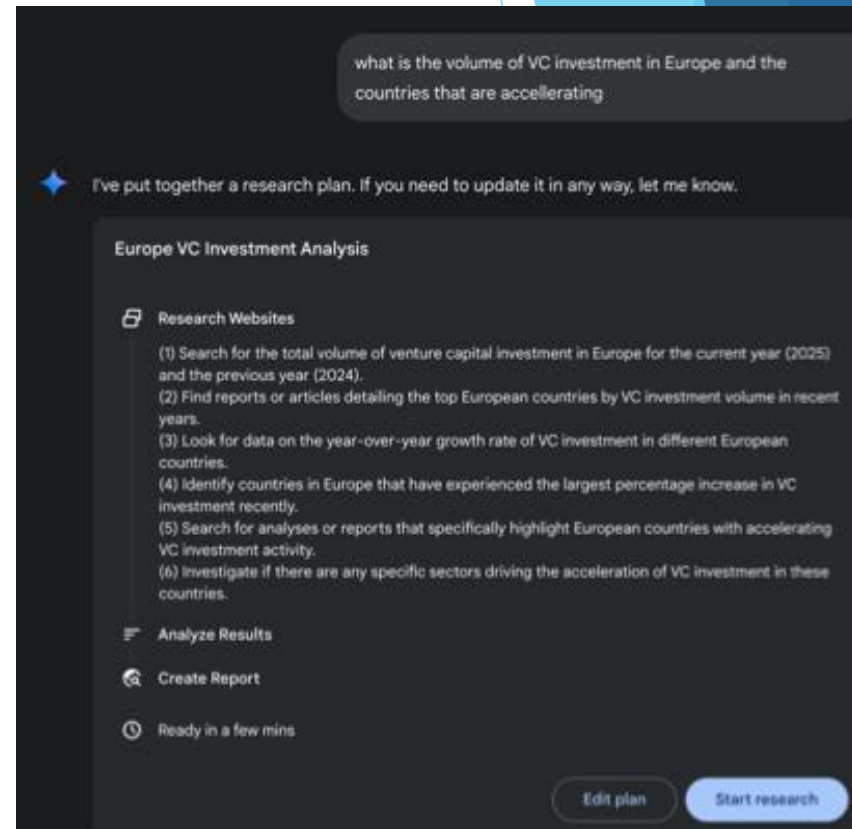
# Memory

- ▶ Memory includes the ability to store and retrieve interaction-specific data, e.g., user input from previous queries, past experiences, and their outcomes, to make the agent adaptable.
- ▶ Memory types: short-term, long-term.
- ▶ Example use cases: Chatbots and Conversational AI, task-Oriented Agents, personalized experience agents.



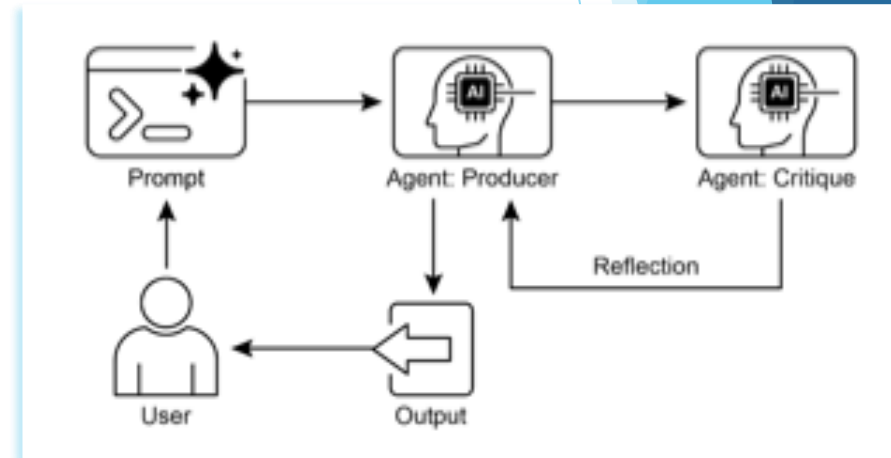
# Reasoning and Planning

- ▶ By applying logical reasoning, the agent uses AI algorithms and heuristics to tackle complex situations, facilitating problem-solving, task decomposition, and strategic planning.
- ▶ Before responding to a query, the agent generates a sequence of actions to ensure a reliable outcome. Planning techniques such as [chain-of-thought reasoning](#), and [subgoal decomposition](#) are often employed.



# Reflection

- Reflection enables agents to process feedback and learn from their experiences. Instead of depending solely on user feedback, the model evaluates its responses critically, identifies improvements, and revises accordingly.
- Beyond self-reflection, external tools like unit tests can further validate outputs, helping models identify and resolve errors. Multi-agent frameworks take this concept even further, one agent generates content, while another critiques it, enabling collaborative improvement.



# Actions

- ▶ The agent's ability to take actions bridges its internal reasoning with the external world, enabling it to achieve its goals through precise execution. This involves **function calling** as a core mechanism, where the agent interacts with APIs, software, or hardware to perform tasks, seamlessly integrating with external tools or services.
- ▶ Connecting these components requires a unified framework that supports all aspects of the agent's functionality. Seamless data exchange between components ensures that, for example, memory feeds into reasoning, and reflection informs future planning.

# Agentic AI Software Stack

The background features abstract, overlapping geometric shapes in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are primarily located on the right side of the frame, creating a modern, tech-oriented aesthetic.

**Actuators**  
Augment native agentic capabilities

UI Automation: Browserbase Browser Use OpenAI Operator  
Sandboxes: E2B Integrations: composio  
App-specific: blender CLOUDFLARE stripe snowflake todoist

**Context**  
Critical for accuracy and relevance

Knowledge Memory  
 Pinecone  zep cognee  
 neo4j mongoDB  mem0

Context Interface

Model Context Protocol

**Agent Frameworks**  
Design patterns & orchestration

LangChain crewai LlamaIndex agno Letta

**Durable Execution**  
Reliably execute multi-step workflows

Temporal AKKA DBOS ingest

**AI Inference Stack**  
Full stack around foundation models

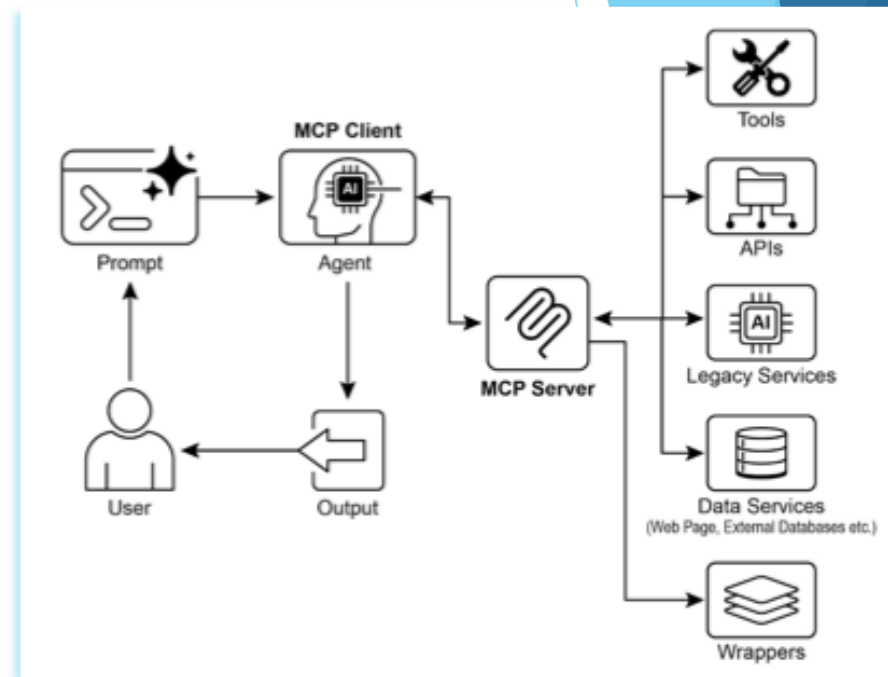
AI LLaMA by Meta deepseek

Observability / Security  
Inter-agent Comms (Google A2A)

wing

# AI Agent Protocols




- ▶ Agent protocols are **standardized frameworks** that define the rules, formats, and procedures for structured communication among agents and between agents and external systems.
- ▶ A unified protocol for agent systems not only addresses current interoperability issues, but it also creates a connected network of intelligence.





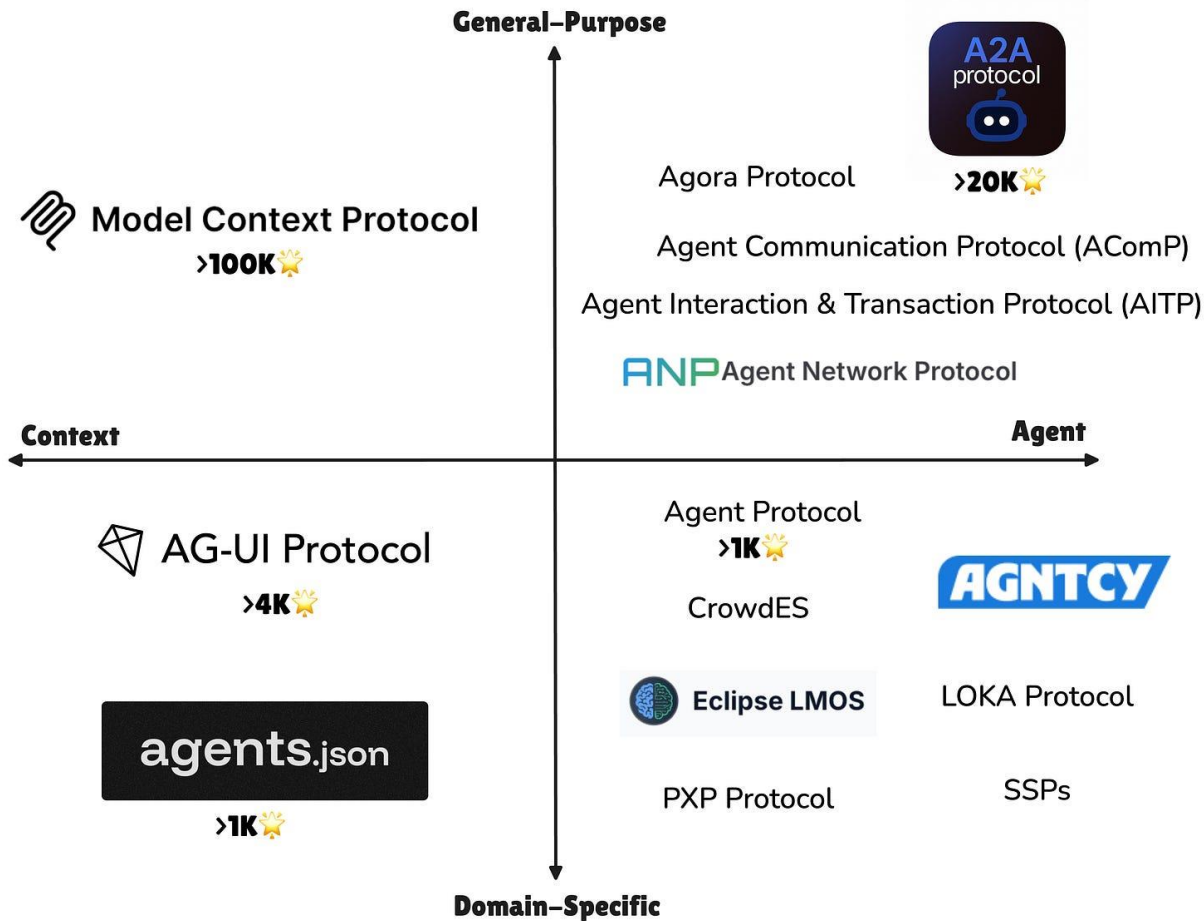
Rakesh Gohel   
@rakeshgohe101

# Comparing most popular AI Agent Protocols

Aspects	<b>MCP</b> <small>(Model Context Protocol)</small>	<b>A2A</b> <small>(Agent to Agent Protocol)</small>	<b>ANP</b> <small>(Agent Network Protocol)</small>	<b>ACP</b> <small>(Agent Communication Protocol)</small>
Developed by	<b>ANTHROPIC</b>			
Architecture	Client-Server	Centralized Peer-to-peer	Decentralized Peer-to-Peer	Brokered client-server
Agent Discovery	Manual registration	Agent Card retrieval via HTTP	Search Engine Discovery	Registry-based
Session Support	Stateless	Session-aware or stateless	Stateless; DID-authenticated	Session-aware with run state tracking
Transport Layer	HTTP, Stdio,SSE	HTTP, optional SSE	HTTP with JSON-LD	HTTP with incremental streams
Strengths	Best for tool calling	Inter-agent negotiation	AI-native protocol negotiation	Tool modularity
Limitations	Best for tool calling	Assumes agent catalog	High negotiation overhead	Registry required

# Agent Communication Protocols Landscape

GenerativeProgrammer.com Updated: 06/2025



# Evaluation of Agentic AI Systems

The background of the slide is white with abstract, overlapping geometric shapes in various shades of blue (light blue, medium blue, and dark blue) on the right side, creating a modern, tech-oriented aesthetic.



# Evaluation Dimensions

- ▶ **Autonomy**: Does the agent operate independently with minimal human intervention? Can it initiate, plan, and execute actions toward goals?
- ▶ **Goal Alignment**: Are the agent's objectives aligned with user intent and organizational goals? Is it steerable and responsive to new constraints?
- ▶ **Adaptability and Learning**: Can the agent adapt to changing environments or user feedback? Does it improve over time through experience or data?
- ▶ **Transparency and Explainability**: Can the agent explain its decisions and reasoning? Is its internal decision process interpretable?
- ▶ **Safety and Robustness**: Does it avoid unsafe, unethical, or undesired behavior? How well does it perform under adversarial or uncertain conditions?

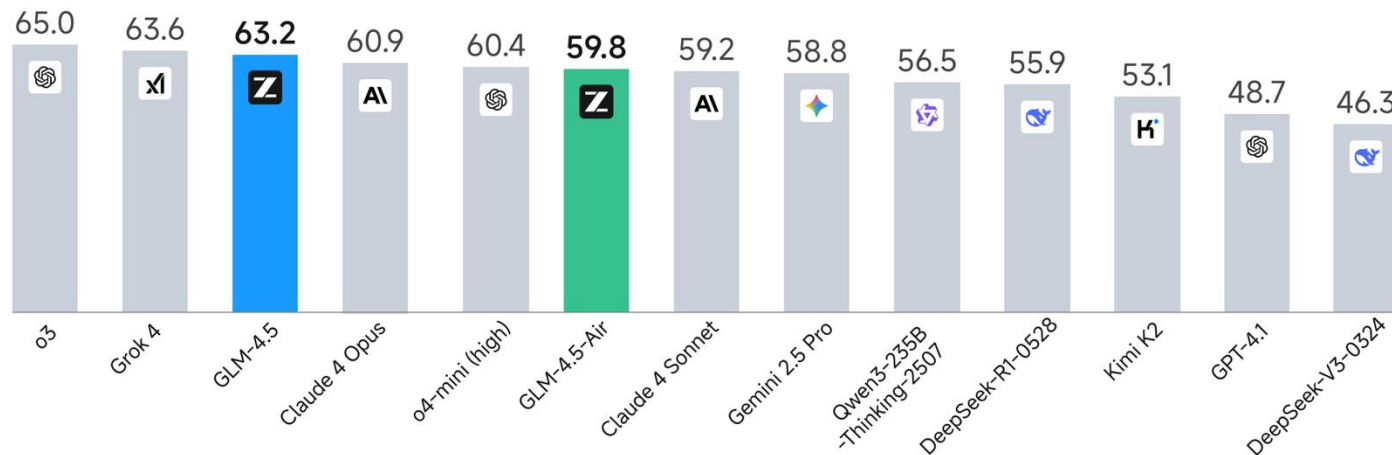
# Evaluation Methodologies

- ▶ **Behavioral Benchmarks**
  - Task success rate (completion, efficiency, correctness)
  - Goal pursuit under uncertainty or obstacles
- ▶ **Simulation Environments**
  - Test agents in sandboxed environments with varying complexity
  - Evaluate response to edge cases and open-ended scenarios
- ▶ **Human-in-the-Loop Evaluation**
  - Expert assessments of decision quality and usability
  - User satisfaction, trust, and cognitive load
- ▶ **Agentic Metrics**
  - Initiative Index: frequency of self-directed actions
  - Delegation Score: appropriateness of seeking help or escalating tasks
  - Planning Depth: complexity of multi-step plans

# LLM Performance Evaluation: Agentic, Reasoning, And Coding Benchmarks

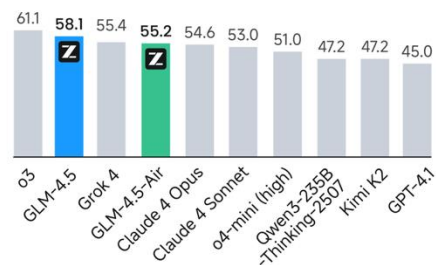


12 benchmarks: MMLU-Pro, AIME 24, MATH-500, SciCode, GPQA, HLE, LCB (2407-2501), SWE-Bench Verified, Terminal-Bench, TAU-Bench, BFCL V3, BrowseComp



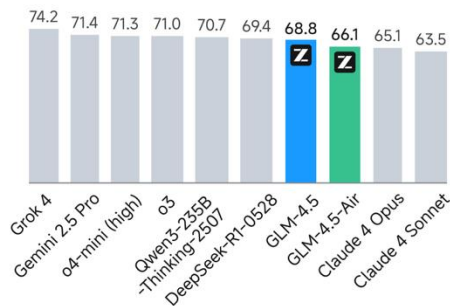
## Agentic

Agentic Benchmarks: TAU-Bench, BFCL V3 (Full), BrowseComp



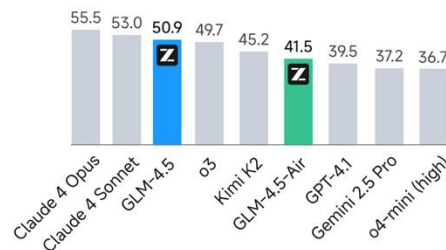
## Reasoning

Reasoning Benchmarks: MMLU-Pro, AIME 24, MATH 500, SciCode, GPQA, HLE, LCB (2407-2501)



## Coding

Coding Benchmarks: SWE-Bench Verified, Terminal-Bench



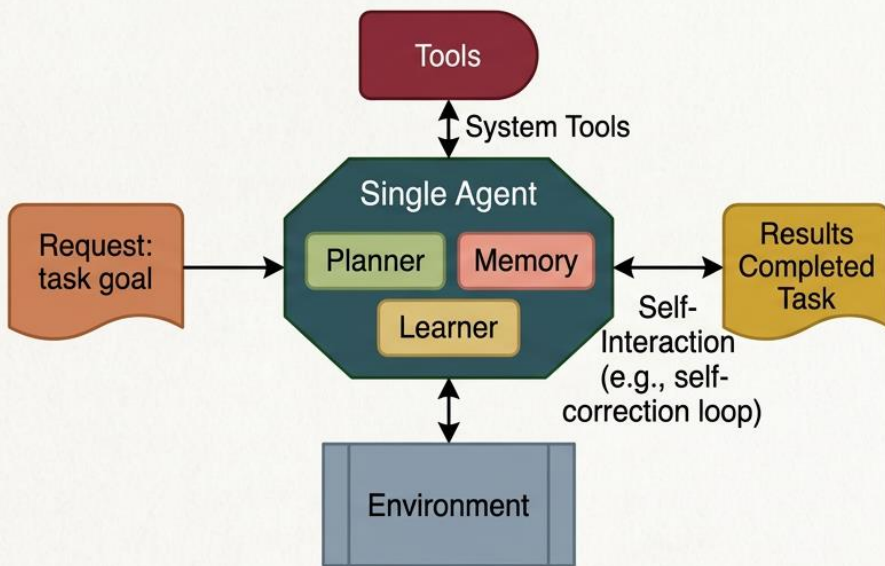
# Module 3: Agentic AI in IR and Generative IR

# Agentic AI in IR

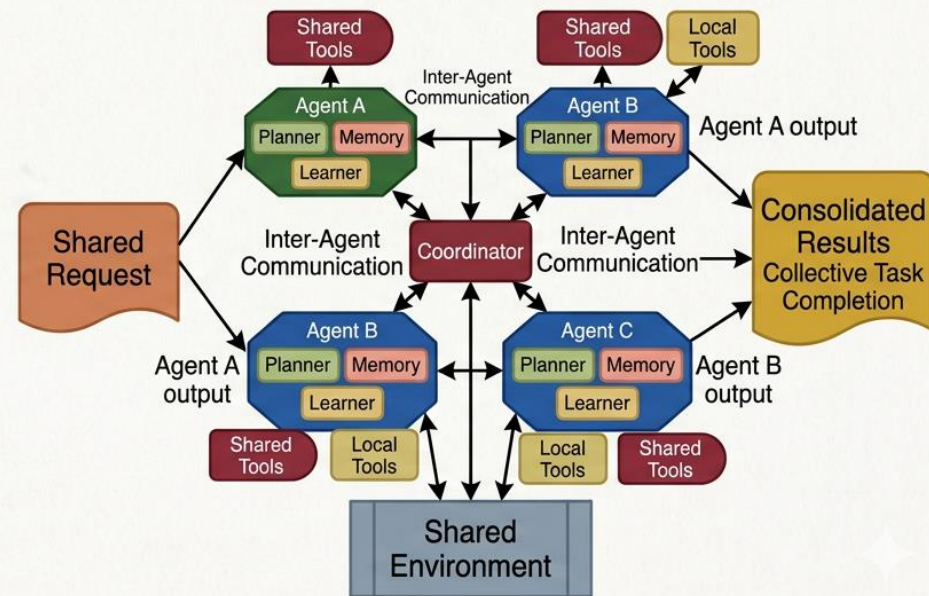
- ▶ In IR systems, agents can orchestrate workflows such as:
  - Query decomposition
  - Multi-source retrieval
  - Evidence aggregation
  - Answer synthesis

# System Design Choices

## Single Agent System

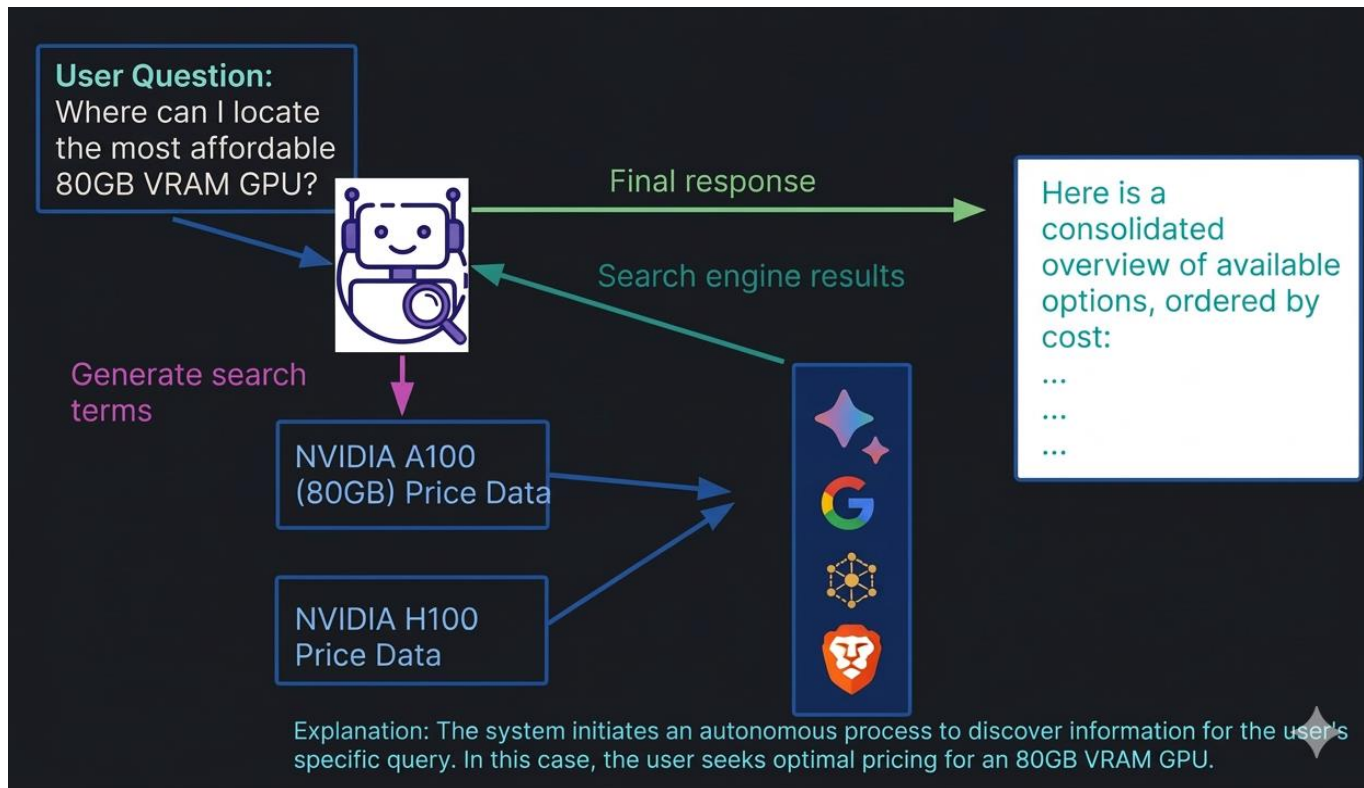


## Multi Agent System



# Task Decomposition

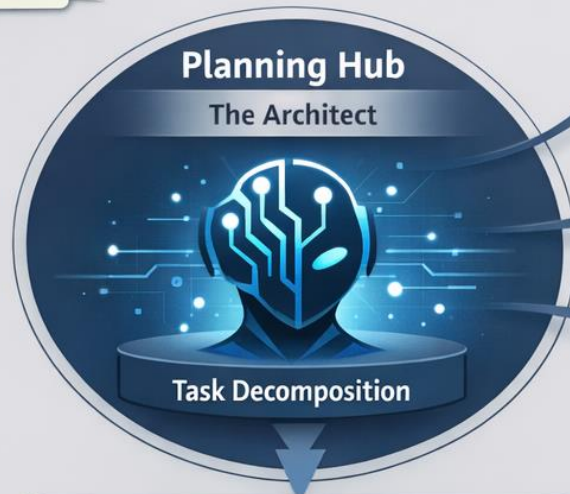
- ▶ Many complex information needs require decomposition, i.e. splitting multifaceted user requests into a hierarchy of simple tasks.



# Planning and Subgoal Decomposition

## User Intent

"Find the cheapest laptop with a dedicated GPU."



### Subgoal 1:

Identify Budget-Friendly Laptop GPUs



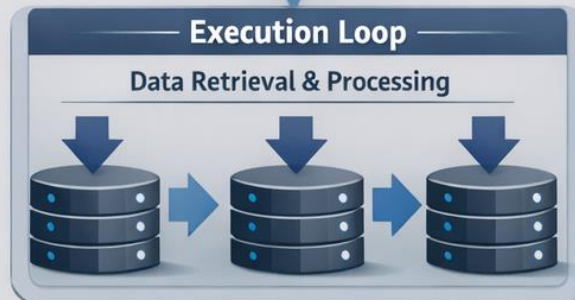
### Subgoal 2:

Scrape & Compare Prices from Retail Sites



### Subgoal 3:

Check Student Discounts & Deals



## Final Output

### Best Option:

Affordable Laptop with  
Dedicated GPU:  
\$599 with Student Discount

# Tool Choice & Orchestration

- ▶ Agents need to orchestrate across diverse digital resources like metasearch and internal APIs, and choose appropriate tools.
- ▶ Examples include:
  - Search engines
  - Domain-specific APIs
  - Browsing tools
  - Retrieval-augmented generation pipelines.

# Retrieval Patterns

- ▶ Agents need to actively initiate lookups and corroborate data.
- ▶ Common retrieval patterns include:
  - Multi-query retrieval
  - Iterative retrieval
  - Retrieval-augmented generation
- ▶ These patterns help ensure coverage of relevant information.

# Temporal Drift

- ▶ Changes in underlying information over time (facts, policies, prices, scientific knowledge) that make previously correct answers outdated.
- ▶ Multi-step pipelines amplify drift:
  - Errors can propagate across planning → retrieval → reasoning
- ▶ Agents rely on mixed sources:
  - Pretraining + retrieval + memory → inconsistent time horizons
- ▶ Stored memory becomes stale:
  - Long-term memory may conflict with fresh retrieval results
- ▶ Failure modes:
  - Stale answers
  - Temporal inconsistencies
  - False confidence
  - Evaluation leakage

# Dataset Contamination

- ▶ Evaluation datasets may contain information already seen during model training. This can artificially inflate performance.
  - Pretraining + Retrieval overlap: The model may have already “seen” the answer—even before retrieval.
  - Multi-step pipelines hide leakage: Planning, reasoning, and synthesis can mask memorization as intelligence.
  - Evaluation becomes misleading: High task success may reflect recall from training, not true reasoning.
- ▶ Strategies include:
  - Timestamp filtering
  - Live retrieval evaluation
  - Dataset auditing

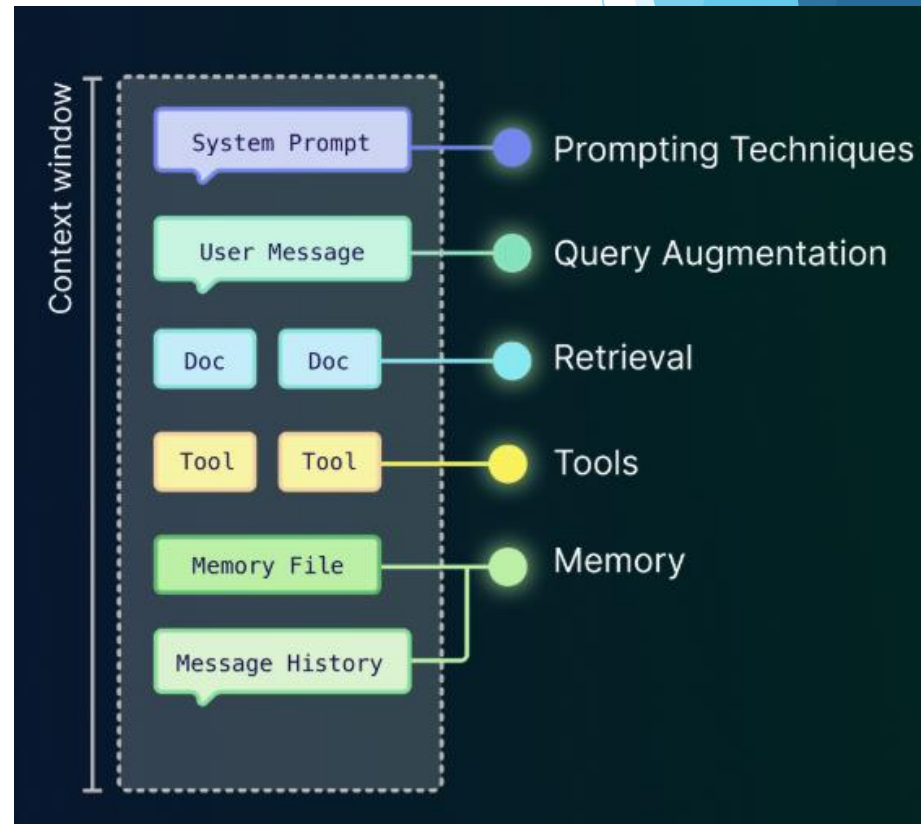
# Prompting for Agentic IR

- ▶ Prompting in agentic AI for information retrieval is about shaping the agent's behavior as a planner and decision-maker, not just a text generator.
  - Prompt Chaining: Decomposing complex tasks into focused, sequential prompts where output from one step acts as input for the next
  - Routing Pattern: Implementing conditional logic to direct the flow of control to different specialized agents or tools based on user intent
- ▶ Effective prompts explicitly define the goal, constraints, available tools, and required evidence, guiding the agent to decompose complex queries, select appropriate retrieval strategies, and ground its answers in verifiable sources.
- ▶ Structured prompting improves transparency, controllability, and auditability.



# Context Management

- ▶ Agents must manage context effectively.
- ▶ Challenges include:
  - Limited context windows
  - Information overload
  - Irrelevant retrieval results
- ▶ Strategies include:
  - Memory summarization
  - Retrieval filtering
  - Context pruning





# Memory Management

- ▶ Memory serves as the **central intelligence hub** for an agentic IR system. It enables the agent to record, recall, and learn from previous experiences to enhance its future accuracy.
- ▶ By integrating historical context, specific user needs, and external data, the system undergoes constant evolution. This creates a loop of continuous improvement, resulting in responses that are increasingly reliable and contextually aware.
- ▶ Memory management design dimensions:
  - Short term vs. long term memory
  - Explicit history vs. compressed embedding based

# Inspectability and Auditability

- ▶ Query Understanding & Reformulation Audit
- ▶ Retrieval Provenance & Coverage
- ▶ Evidence Attribution (Grounding)
- ▶ Ranking & Selection Transparency
- ▶ Multi-Hop Reasoning Trace
- ▶ Context Construction & Prompt Assembly
- ▶ Answer Faithfulness vs Helpfulness
- ▶ Source Quality & Trustworthiness
- ▶ Retrieval-Time Security Risks
- ▶ Temporal Drift & Freshness
- ▶ User Personalization & Filter Bubbles



# Module 4: Measuring What Matters

The background features abstract, overlapping geometric shapes in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are primarily located on the right side of the frame, creating a modern, layered effect against the white background.

# Beyond n-grams

- ▶ Traditional metrics such as BLEU or ROUGE are insufficient.
- ▶ Agentic IR evaluation should measure:
  - Correctness
  - Evidence support
  - Adequacy
  - Efficiency
- ▶ New for agentic AI compared to traditional ML:
  - Not just **model weights** → entire **decision process**
  - Not just **predictions** → **actions with consequences**
  - Not just **static evaluation** → **dynamic trajectories**

# Retrieval Quality

These are your classic IR metrics, but now applied per step and across hops.

## ▶ Key Metrics

- Recall@k (per hop + cumulative)
- nDCG@k / MRR
- Coverage (did we retrieve all necessary evidence?)
- Diversity (source/domain diversity)
- Redundancy rate

## ▶ New Agentic Twist

- Multi-hop Recall: Did all required pieces get retrieved across steps?
- Retrieval Efficiency: Relevant docs / total retrieved

# Multi-Step Reasoning & Planning Metrics

- Task Success Rate (TSR): Did the agent fully answer the query correctly?
- Hop Efficiency: useful hops/total hops
- Plan Optimality: Compare agent plan vs ideal plan
- Error Propagation Rate: % of failures caused by earlier steps
- Recovery Rate: Can the agent self-correct after bad retrieval?

# Grounding & Faithfulness Metrics

- Citation Coverage: % of claims backed by sources
- Faithfulness Score: Are claims supported by retrieved evidence?
- Hallucination Rate: Unsupported or fabricated content
- Attribution Precision / Recall: Correct mapping of claims to passages
- Evidence Sufficiency: Is retrieved evidence enough to justify answer?

# Source Quality & Fairness Metrics

- ▶ Especially critical in open-domain IR (news, health)
- ▶ Key Metrics:
  - Source Authority Score
  - Freshness Score
  - Viewpoint Diversity
  - Bias Amplification
  - Source Concentration Index

# Temporal & Freshness Metrics

- ▶ IR is time-sensitive
- ▶ Key Metrics:
  - Freshness Match Rate
  - Staleness Error Rate
  - Temporal Consistency



# Important Composite Metrics

- ▶ End-to-End Task Success:
  - Combines retrieval + reasoning + synthesis
- ▶ Grounded Answer Rate:
  - Correct AND supported by evidence
- ▶ Multi-Hop Retrieval Coverage:
  - Did the agent find all necessary info?
- ▶ Hallucination Rate
  - Must be minimized
- ▶ Efficiency (Cost vs Quality):
  - Tokens, latency, hops

$$\text{Efficiency} = \frac{\text{Task Success Score}}{\text{Tokens} + \alpha \cdot \text{Latency} + \beta \cdot \text{Hops}}$$

# Module 5: Case Studies and Open Problems

# Google Deep Research

- ▶ Deep Research exemplifies agentic IR by treating retrieval as an **ongoing, goal-directed process** rather than a one-shot query.
- ▶ Iterative, multi-step retrieval
- ▶ Planning-driven
- ▶ Evidence aggregation & synthesis
- ▶ Self-evaluation and refinement
- ▶ Focus on completeness over speed



# Q&A for Dynamic Information

- ▶ “What is the safest and most cost-effective cholesterol medication for a 55-year-old with diabetes and mild kidney disease?”
- ▶ A standard search engine would:
  - Return documents (papers, guidelines, blogs)
  - Require the user to:
    - Compare medications
    - Interpret clinical trade-offs
    - Cross-check safety
- ▶ This query is:
  - Multi-hop
  - High-stakes
  - Requires synthesis + reasoning

# How an Agentic IR System Solves This?

- ▶ **Step 1: Goal Interpretation (Intent → Structured Task)**
  - The agent transforms the query into a structured goal:
    - Not just keyword matching
    - Convert query to decision problem
- ▶ **Step 2: Task Decomposition (Planning)**
  - Identify candidate medications
  - Retrieve clinical guidelines
  - Compare
  - Retrieve cost
  - Synthesize recommendation

# How an Agentic IR System Solves This?

- ▶ Step 3: Multi-Source Retrieval (Tool Use)
  - ▶ Tools used:
    - PubMed / medical papers
    - Clinical guidelines (e.g., AHA, ADA)
    - Drug databases
    - Pricing APIs
  - ▶ Retrieval strategy:
    - ▶ Multi-query retrieval
      - “statins kidney disease safety”
      - “cholesterol drugs diabetes guidelines”
    - ▶ Iterative retrieval
      - Refine queries based on gaps

# How an Agentic IR System Solves This?

- ▶ Step 4: Evidence Aggregation & Filtering
- ▶ Step 5: Reasoning & Trade-off Analysis
  - Agent performs multi-constraint reasoning
- ▶ Step 6: Answer Synthesis (Grounded Output)
  - Citation
  - Justification
  - Tradeoffs
- ▶ Step 7: Verification & Reflection
  - Agent performs:
    - Self-checking: Did I consider contraindications
    - Ensures all claims are grounded: ? Is every claim supported by retrieved evidence?

# Open Research Questions

Agentic IR forces us to rethink retrieval as a sequential decision-making problem—where the core challenges are no longer ranking documents, but planning search, grounding reasoning, and governing autonomous information access.

- User modeling: Challenges in fine-tuning agents to specific user personas and enterprise contexts
- How should agents plan search vs. retrieve greedily
- safety and misinformation
- Grounding and faithfulness at scale
- Memory design for long horizon IR agents
- Cost vs. quality optimization
- Governance of agentic systems

These topics present major opportunities for CHIIR research.

# Conclusion

- ▶ Agentic systems transform information retrieval by shifting from:
  - queries → goals
  - documents → answers with evidence
  - click logs → interaction traces
- ▶ For IR researchers, this introduces new opportunities in:
  - Evaluation
  - User interaction studies
  - System design

Thank You

Questions?