AI / LLM-based Agents
&
Security

LLMs

ChatGPT

Claude

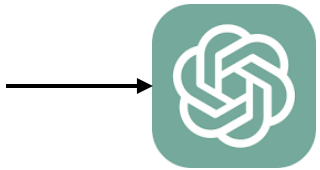
deepseek

LLM & Agent

ChatGPT

LLM & Agent

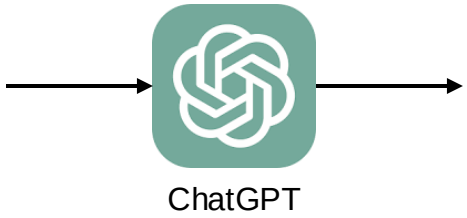
“Help me plan a 3-day trip to New York City.”



ChatGPT

LLM & Agent

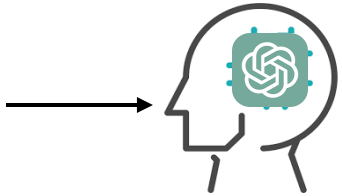
“Help me plan a 3-day trip to New York City.”



“Here’s a suggested itinerary:
Day 1: Visit Times Square, Central Park, and the MET.
Day 2: Explore the Statue of Liberty and the Brooklyn Bridge.
Day 3:…”

LLM & Agent

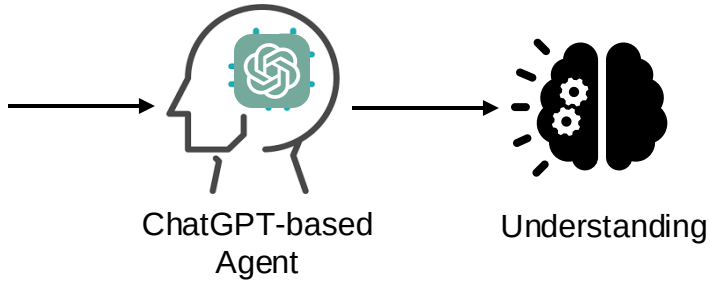
“Help me plan a 3-day trip to New York City.”



ChatGPT-based Agent

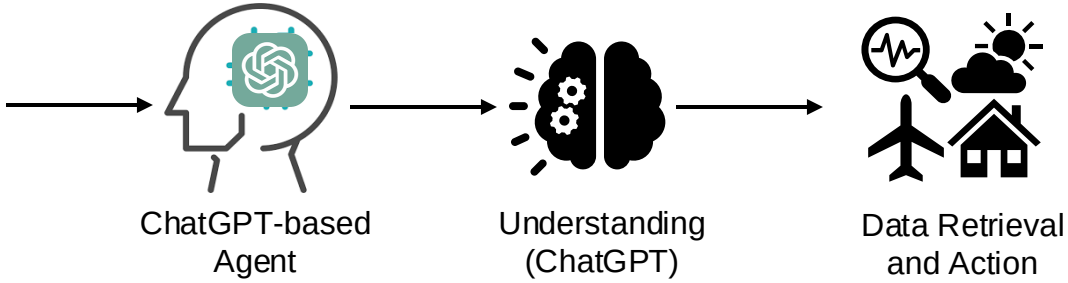
LLM & Agent

“Help me plan a 3-day trip to New York City.”



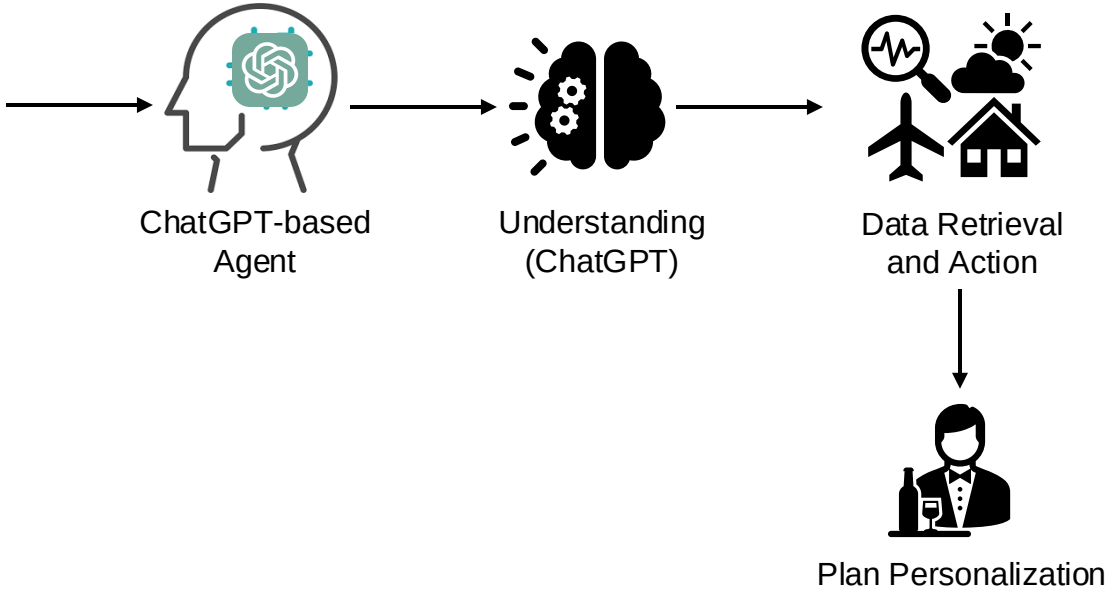
LLM & Agent

“Help me plan a 3-day trip to New York City.”

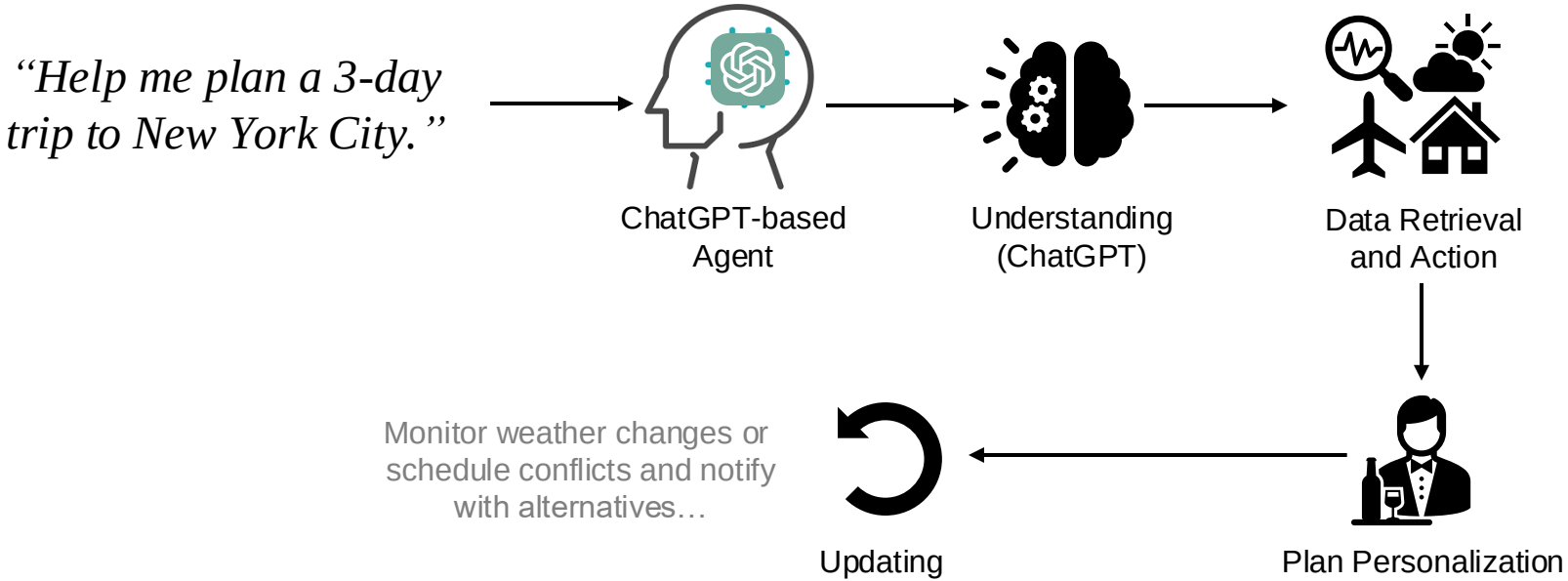


LLM & Agent

“Help me plan a 3-day trip to New York City.”



LLM & Agent


"Help me plan a 3-day trip to New York City."
ChatGPT-based Agent
Understanding (ChatGPT)
Data Retrieval and Action
Plan Personalization
Updating
Monitor weather changes or schedule conflicts and notify with alternatives…

What is an Agent?

An **AI/LLM-based agent** can be defined as an application that attempts to achieve a goal by observing the world and acting upon it using the tools that it has at its disposal

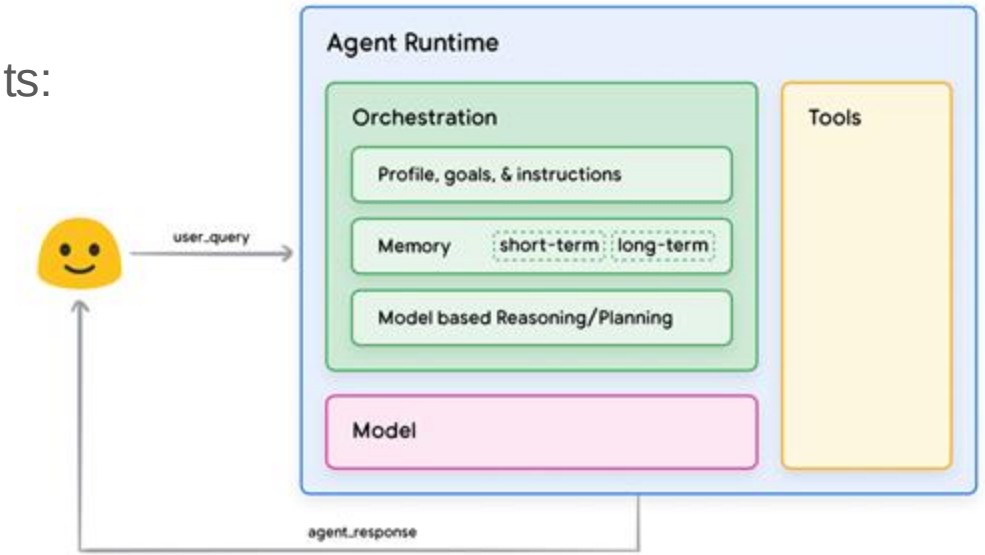
Agents are autonomous and act independently of human intervention; they can reason about what they should do next to achieve the ultimate goal

Foundational Components of an Agent

The foundational components that drive the agent's behavior, actions, and decision making

Three essential components:

- Model
- Tools
- Orchestration Layer

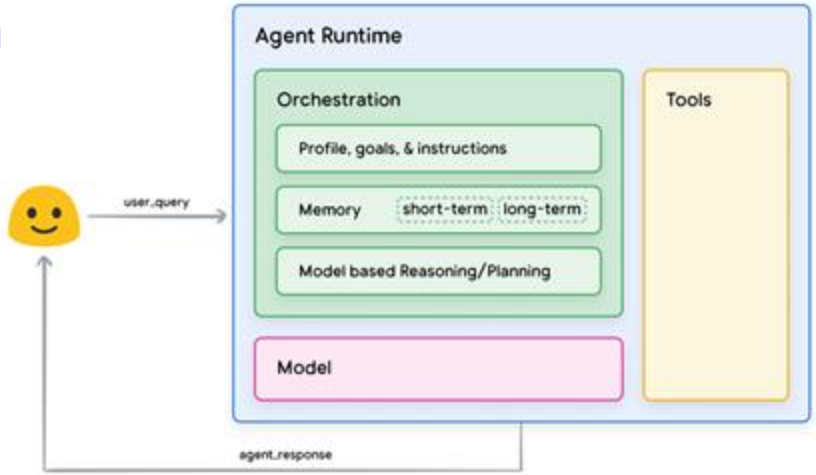


Google. Agents – white paper.

Foundational Components of an Agent

The Model:

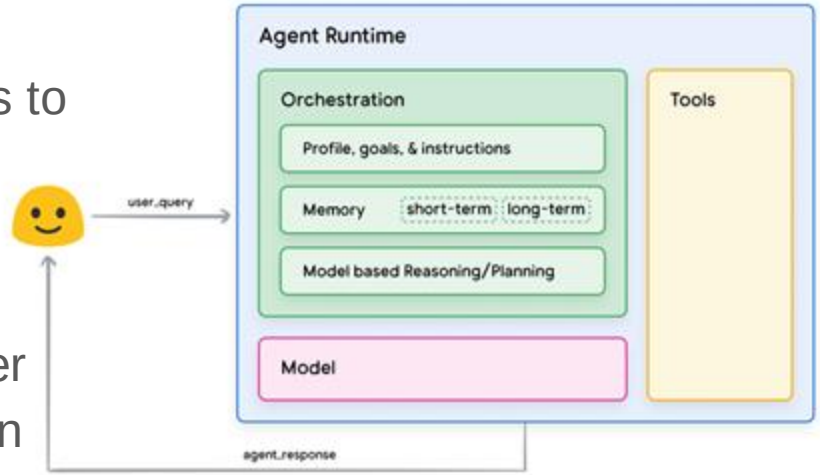
- Refers to the language model (LM) that will be utilized as the centralized decision maker for agent processes
- The model in an agent can be one or multiple LM's of any size that can instruction based reasoning and logic frameworks



Foundational Components of an Agent

The Tools:

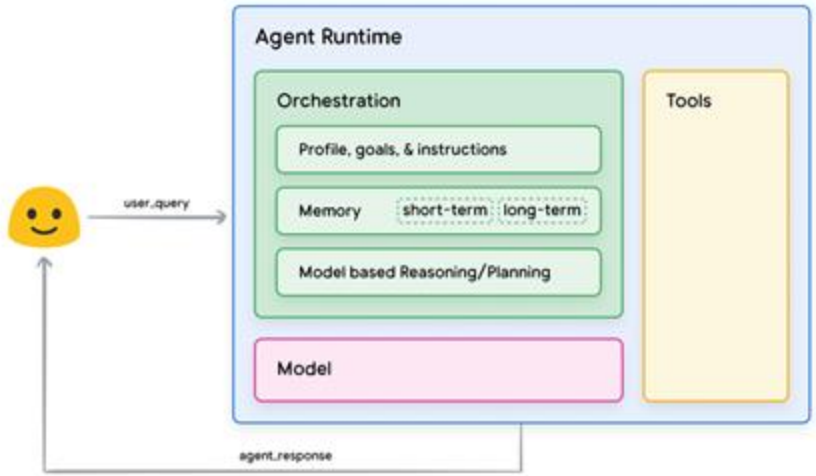
- Foundation models remain constrained by the inability to interact with the outside world
- Tools bridge the gap, empowering agents to interact with external data and services
- Typically align with web API methods
- *E.g.,* a tool can update customer information in a database or fetch weather data to influence a travel recommendation



Foundational Components of an Agent

The Orchestration Layer:

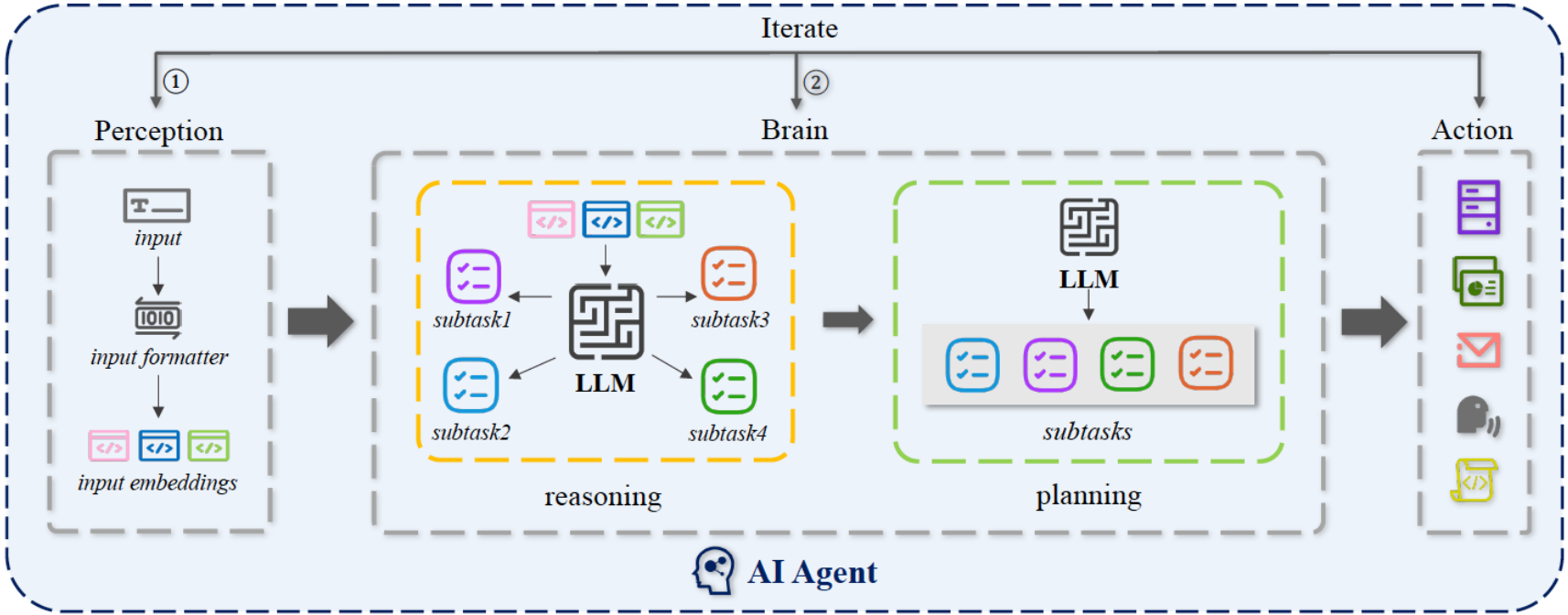
- Describes a cyclical process that governs how the agent takes in information, performs internal reasoning, and then inform its next action or decision
- In general, the loop will continue until the agent reaches its goal or a stop point



Capabilities of Agents

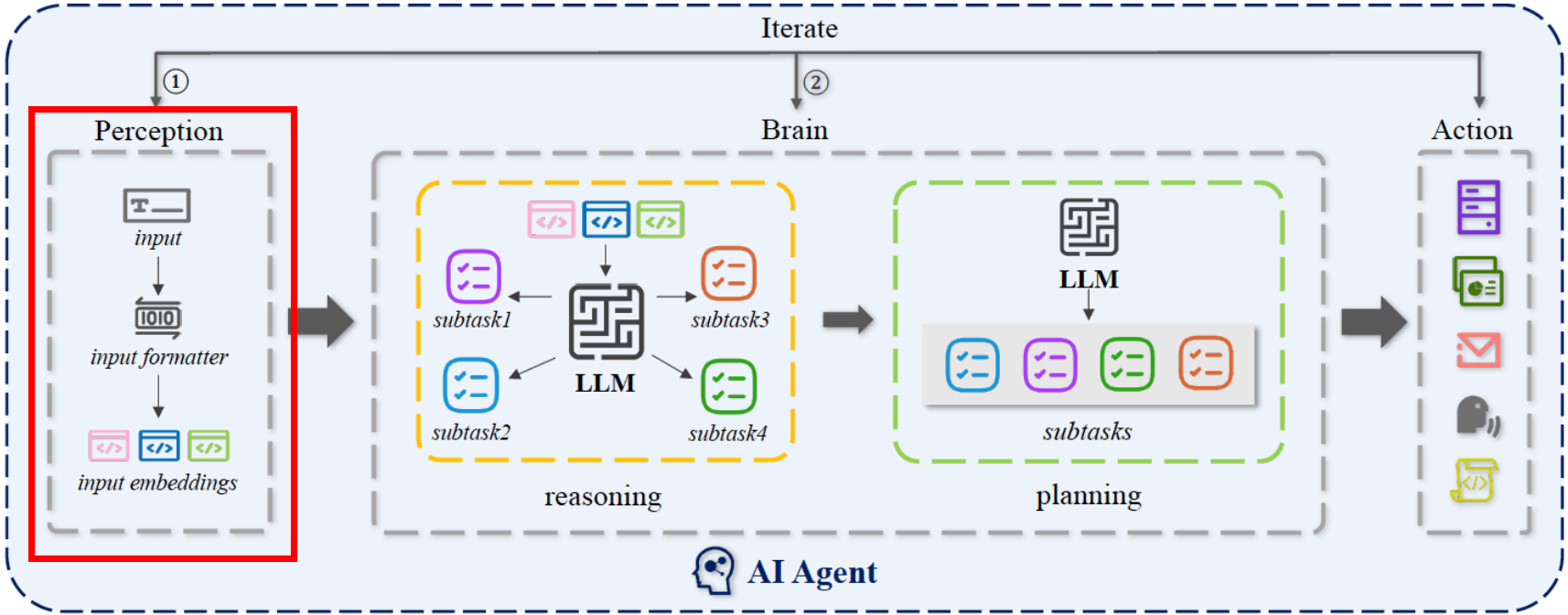
- Tool Utilization
- Advanced Reasoning
- Tailored Generation
- Levels of Autonomy
- Integration with Other AI systems

Conceptual Framework of AI Agent

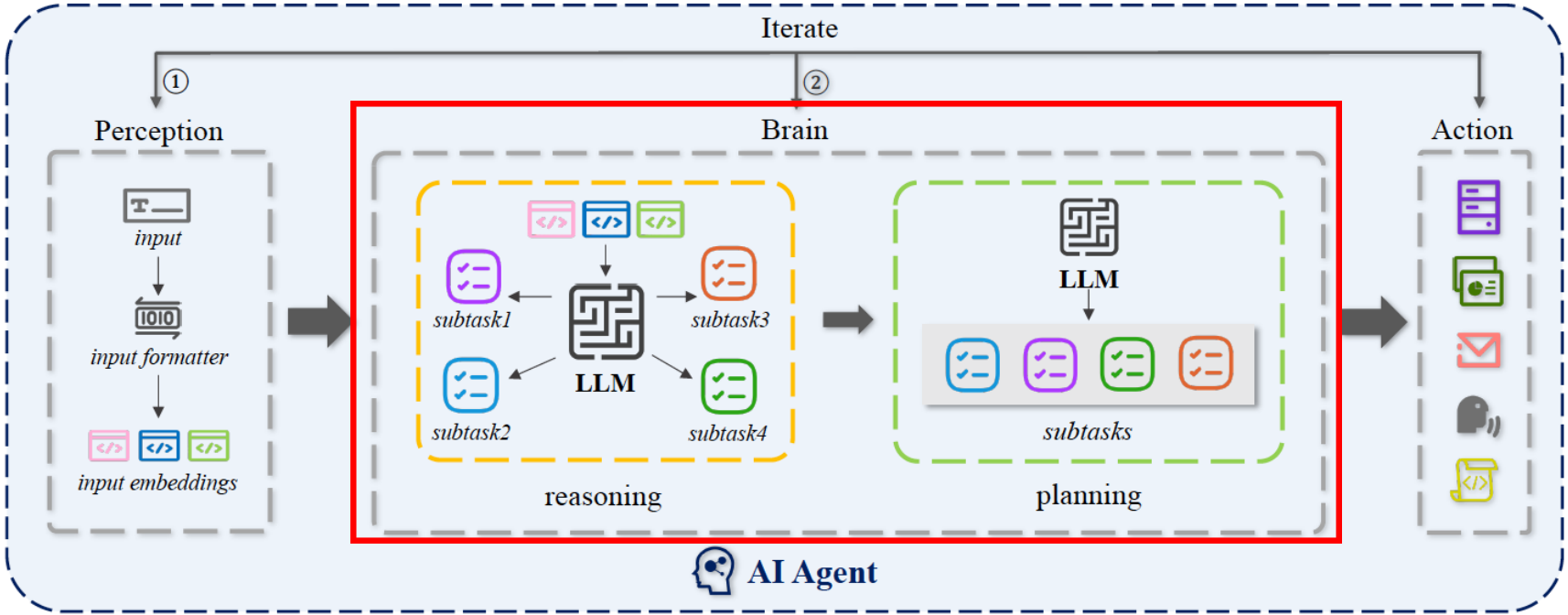


Swinburne University of Technology. AI Agents Under Threat: A Survey of Key Security Challenges and Future Pathways.

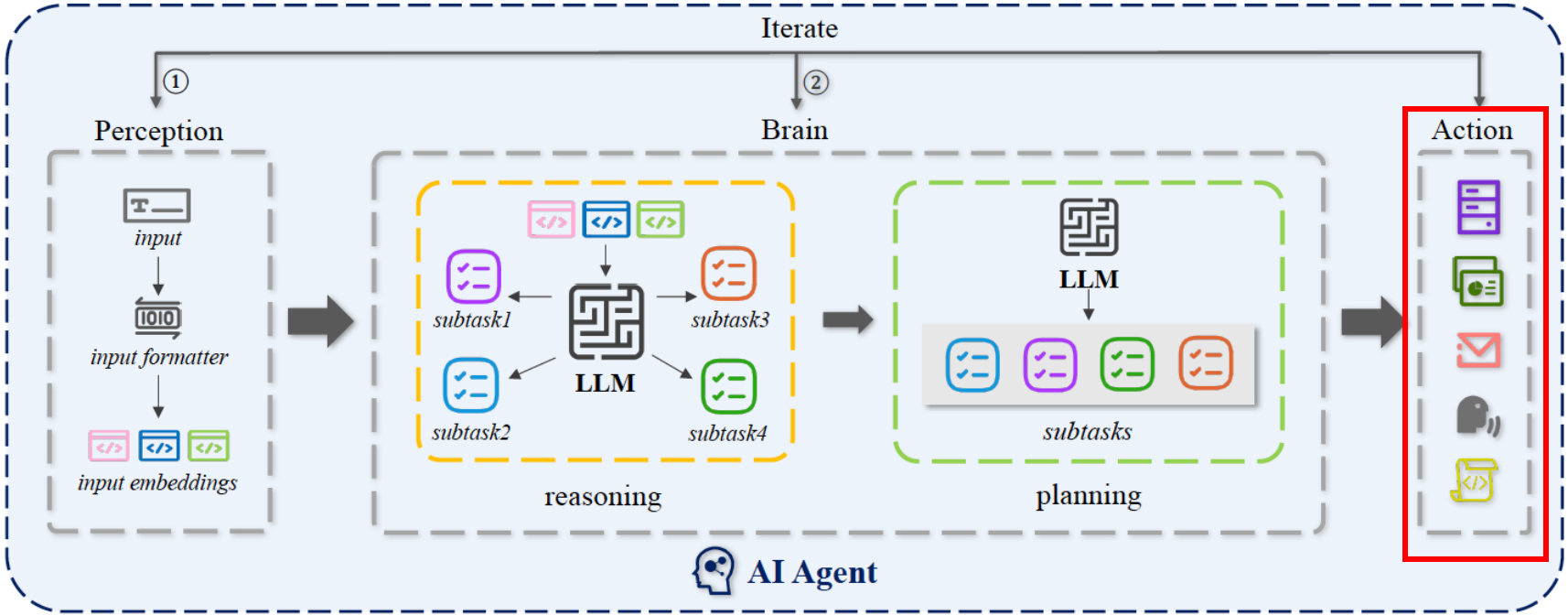
Conceptual Framework of AI Agent



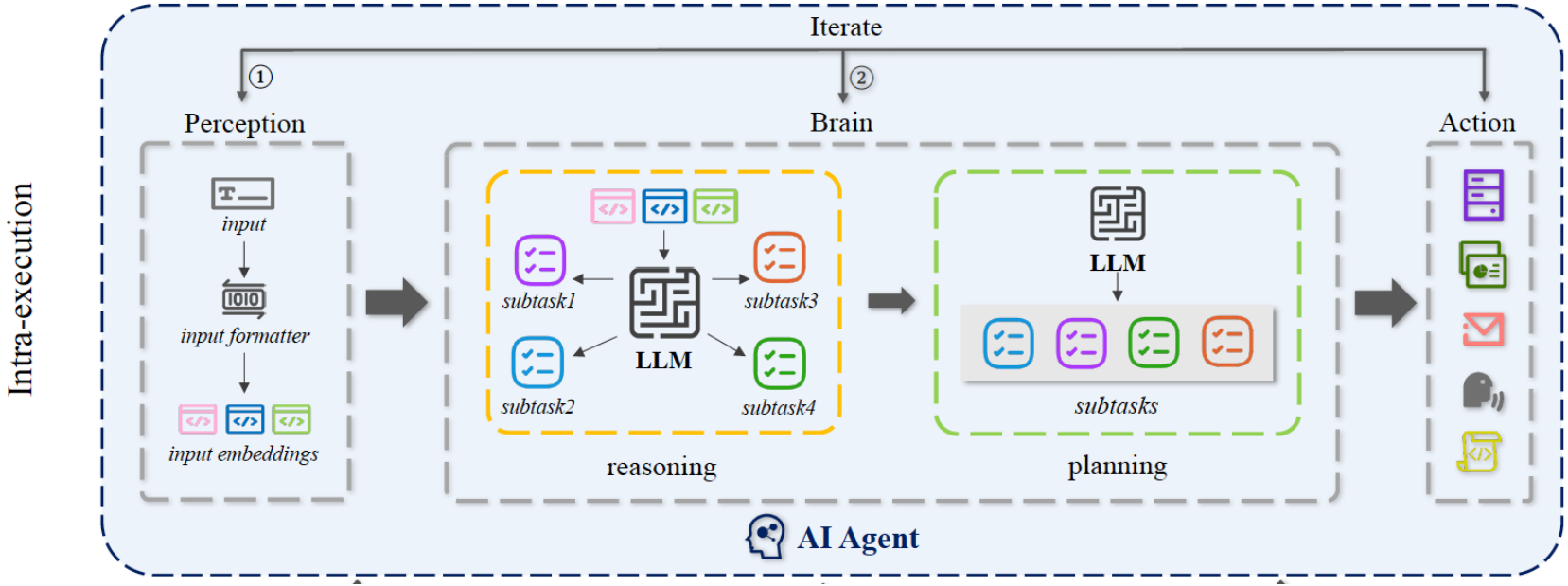
Conceptual Framework of AI Agent



Conceptual Framework of AI Agent



Conceptual Framework of AI Agent



Conceptual Framework of AI Agent