Enabling complex reasoning and action with ReAct, LLMs, and LangChain

Giuseppe Zappia
Principal Solutions Architect
Amazon Web Services

Shelbee Eigenbrode
Principal AI/ML Specialist Solutions Architect
Amazon Web Services

Agenda

- Langchain Overview
- ReAct Framework: High-Level Overview
- Workshop Introduction

LangChain and Retrieval Augmented Generation (RAG) Overview

LangChain Components

Component	Function
Document Loaders	Load and manipulate documents
Vector Stores	Store and query unstructured data through vectors
Prompt Templates	Build templates to optimize LLM queries
LLMs	Interfaces for LLMs
Chains	Combine LLMs and prompt templates to build workflows
Memory	State management of chains/agents to preserve context
Agents	Use LLMs to choose which activities to perform
Tools	Used by agents to perform a specific task (Google Search, DB lookups, etc.)

ReAct Framework: High Level Overview

If you were asked you the following question, how would you solve it?

What is the 4th largest planet in our solar system,
and how many Earths can fit inside it?

Things you need to know:

- Which planet is the 4th largest in the solar system?
- What is the volume of that planet?
  - What's the radius of the planet?
  - What's the formula for the volume of a sphere?
- What is the volume of Earth?
  - What's the radius of Earth?
  - What's the formula for the volume of a sphere?
- What is the ratio of the Earth to that planet?

Wikipedia

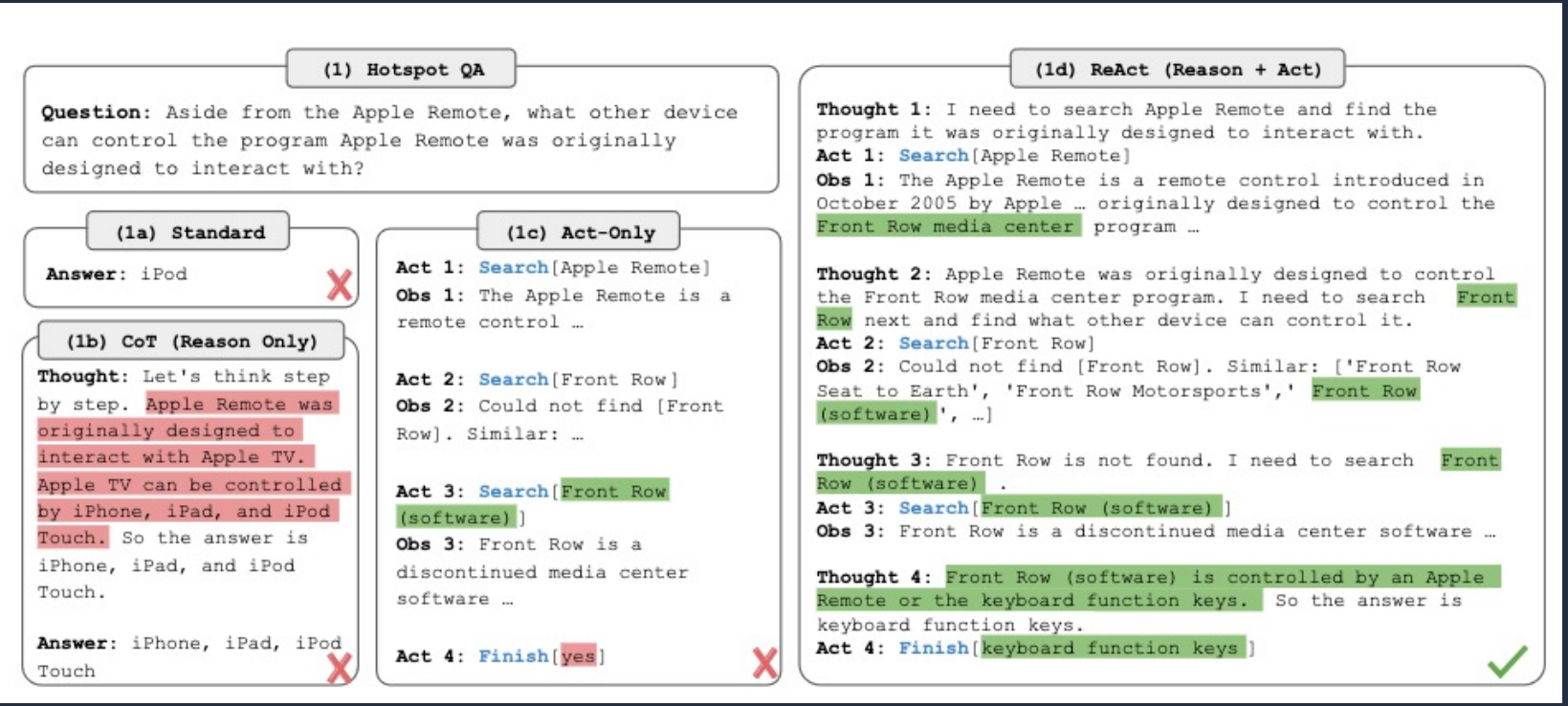
Calculations

Things you look up:

- Neptune
- V = 6.253 x 10^{13} km^3
  - Radius = 24,622 km
  - V = 4/3 π r^3
- V = 1.08321 x 10^{12} km^3
  - Radius = 6,371 km
  - V = 4/3 π r^3
- 62.53/1.08321 = ~ 57.7

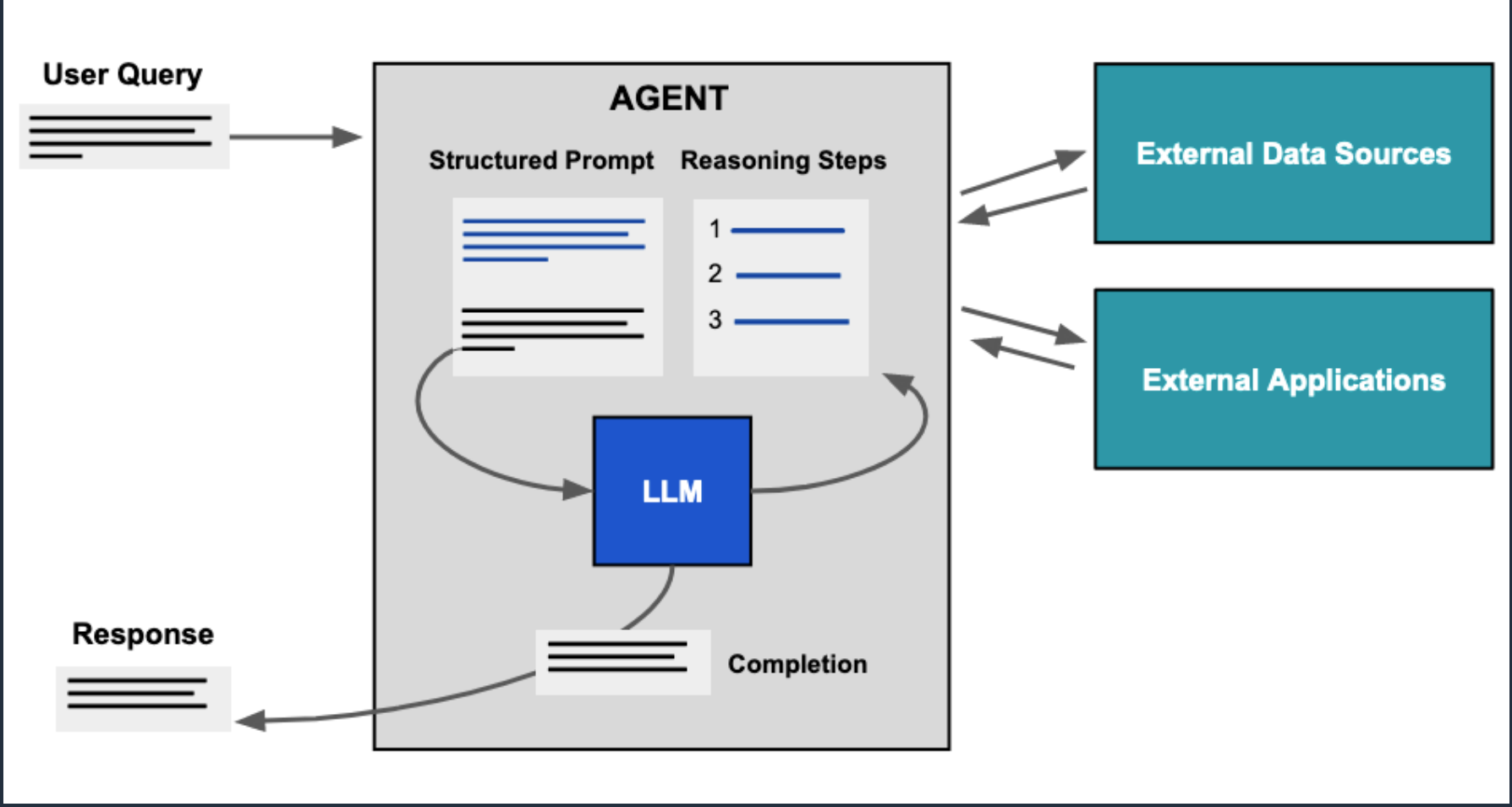
ReAct Framework

ReAct: Synergizing Reasoning and Acting in Language Models



Source: https://arxiv.org/pdf/2210.03629.pdf

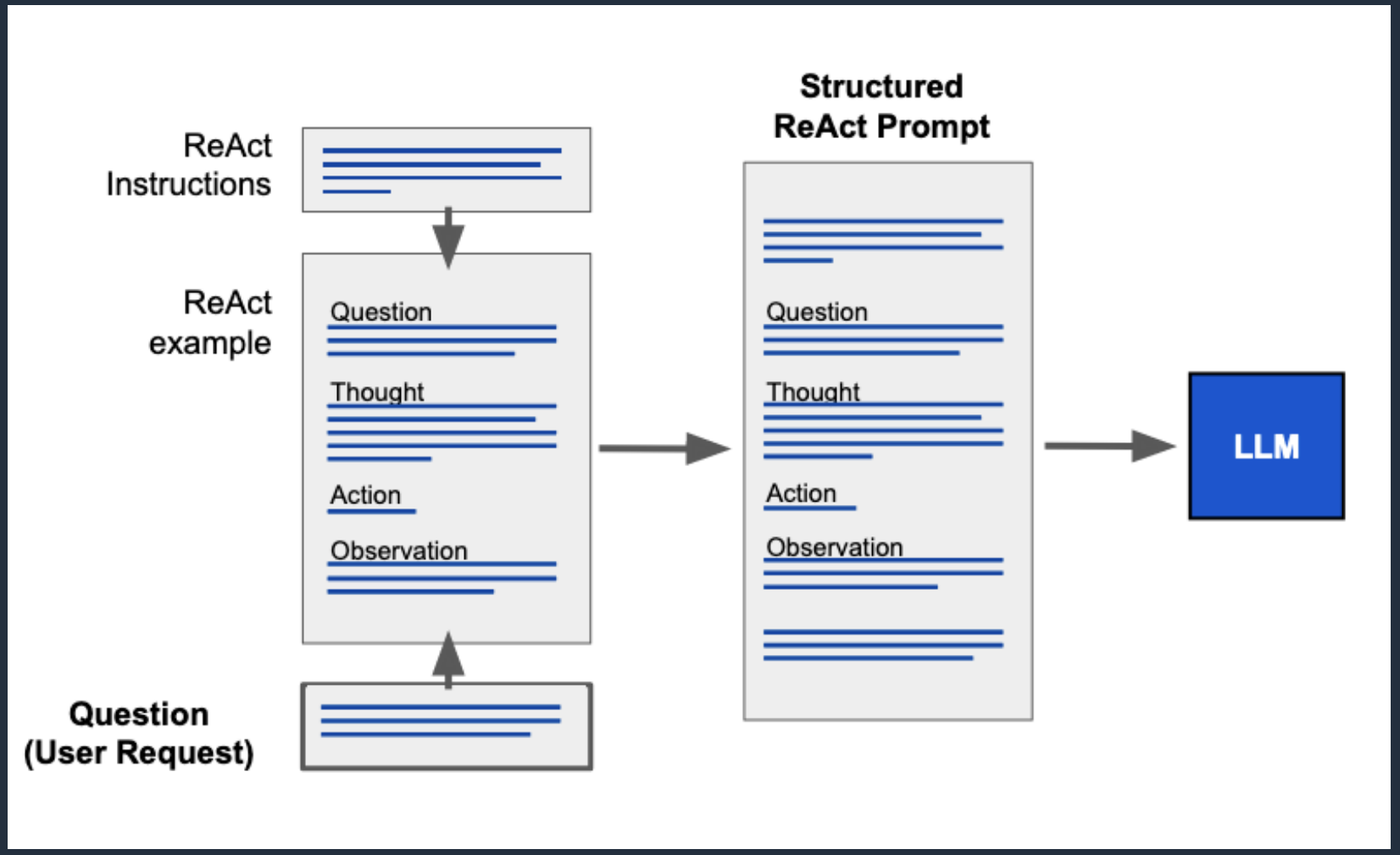
ReAct: The role of Agents



Agents orchestrate prompt-completion workflows between user requests, the foundation model, and external data sources and/or applications

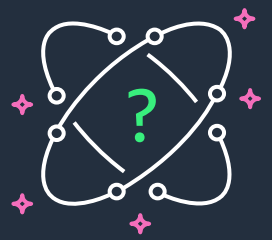
Source: Generative AI on AWS, O'Reilly

ReAct: Prompt Structure



Source: Generative AI on AWS, O'Reilly

ReAct: Prompt Breakdown

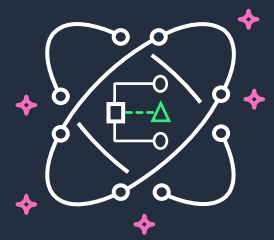


Question

Problem or query that requires reasoning and one or more steps to solve

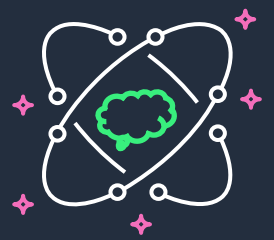
Thought

Reasoning steps to guide the model in how to solve the problem and which actions to take



Action

External actions that can be carried out through allowed actions and tools



Observation

Result of completing the action

Final Answer

ReAct: Prompt Breakdown

Question

Which candy was created first, Twix or Snickers?

```
Tools Available: [Wikipedia]
Actions Allowed:  search[entity], lookup[string], finish[answer]
```

Thought Need to search for Twix and Snickers and see which one was created first

Action search[Twix]

Observation *“The product was first produced in the United Kingdom in 1967...”*

Thought Twix was first produced in 1967. Search for Snickers next.

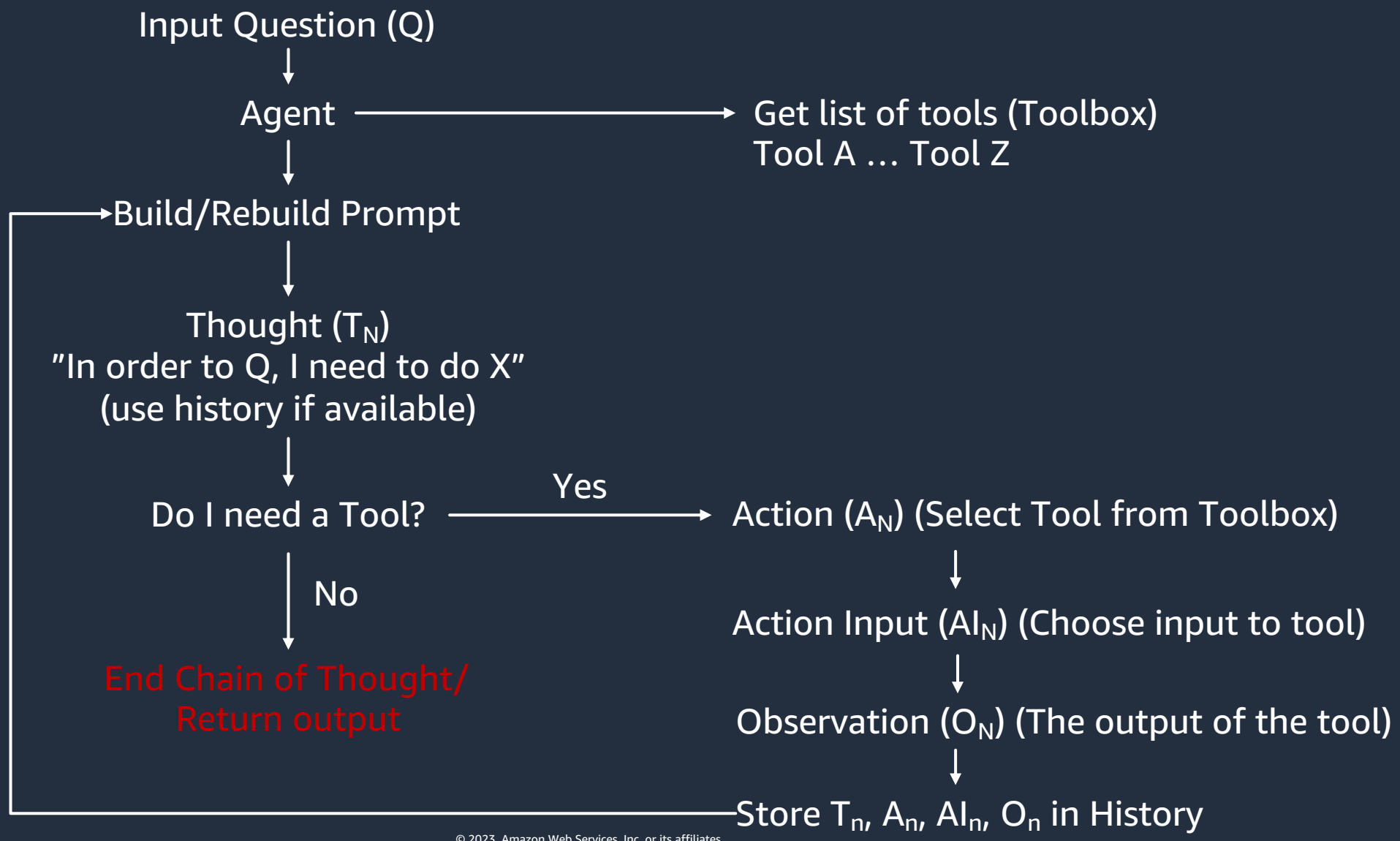
Action search[Snickers]

Observation “In 1930, Mars introduced Snickers,...”

Thought Twix was first produced in 1967. Snickers was first produced introduced in 1930 so Snickers was created first.

Action finish[Snickers]

ReAct: Visual Workflow



Workshop Introduction