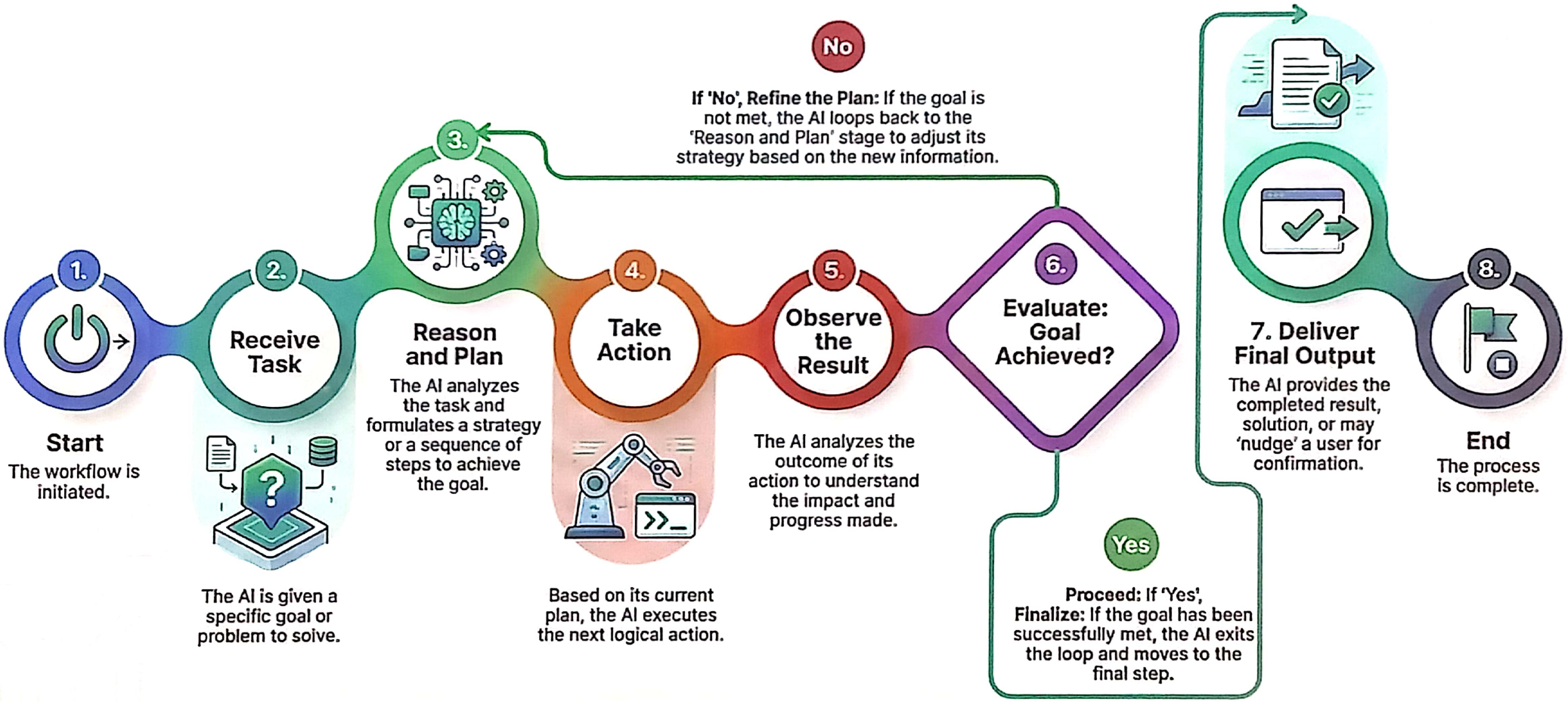
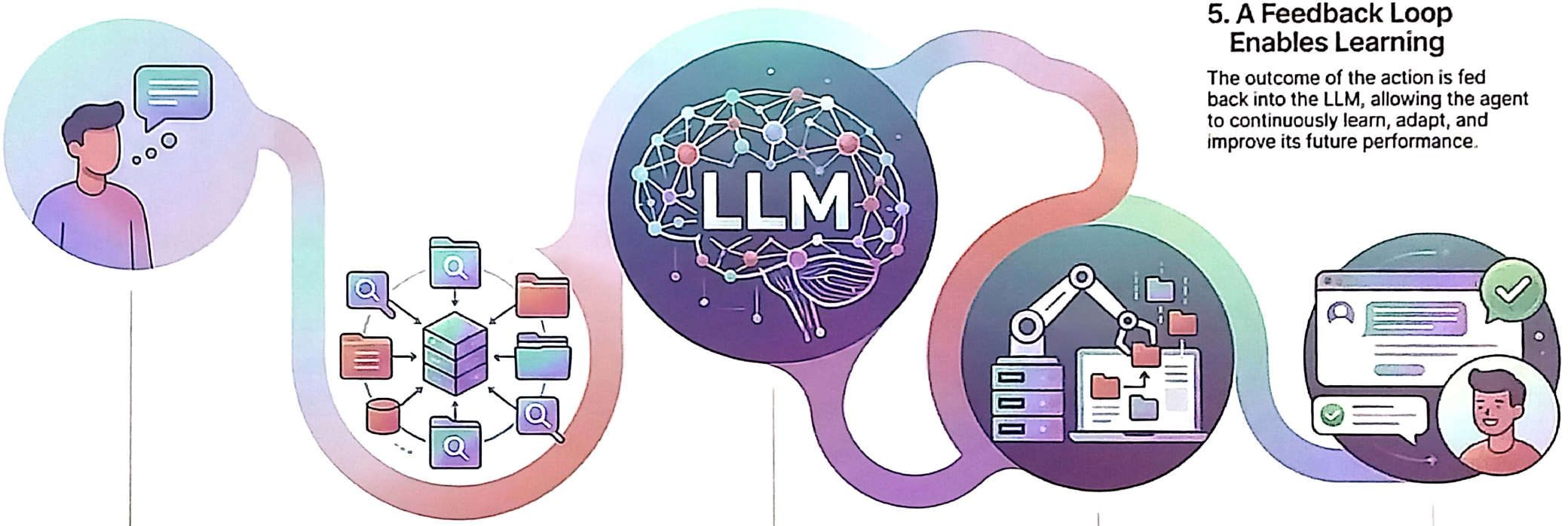
How an Agentic AI Completes a Task



How an AI Agent Works: A High-Level Flow



1. User Initiates a Request

The process begins when a user provides an input, such as a question or a command, to the system.

2. System Accesses Database

To understand the context and retrieve relevant information, the system queries its knowledge database based on the user's request.

3. LLM Processes Information

The user's request, along with the contextual data from the database, is sent to the Large Language Model (LLM) for reasoning and decision-making.

4. Agent Executes an Action

Based on the LLM's output, the system performs a specific task or action, represented by the turning gears.

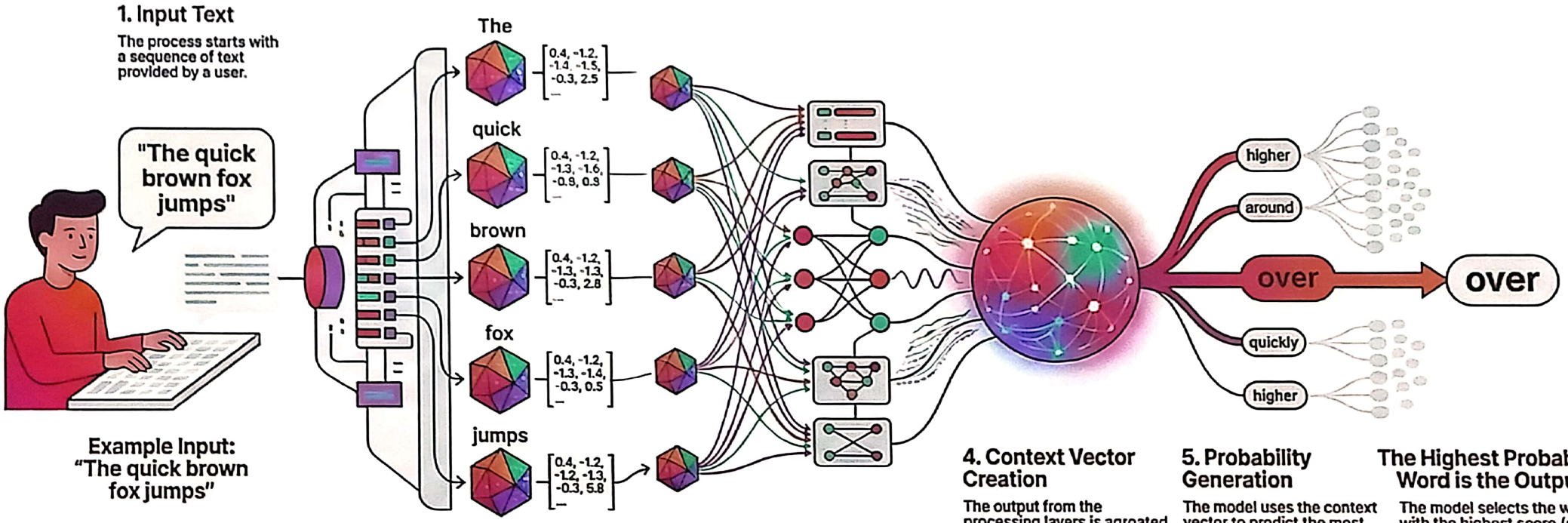
5. A Feedback Loop Enables Learning

The outcome of the action is fed back into the LLM, allowing the agent to continuously learn, adapt, and improve its future performance.

6. System Responds to the User

The user receives a response or sees the result of the action, completing the interaction cycle.

How Do LLMs Work? A Simplified View



2. Text to Embeddings

Each word is converted into a unique numerical vector, or "embedding," that the machine can process. This captures the word's semantic meaning.

3. Contextual Processing

The embeddings are processed through multiple deep learning layers, allowing the model to understand the relationships and context between the words in the sequence.

4. Context Vector Creation

The output from the processing layers is agroated into a single "Context Vector" which represents the combined meaning of the entire input sentence.

5. Probability Generation

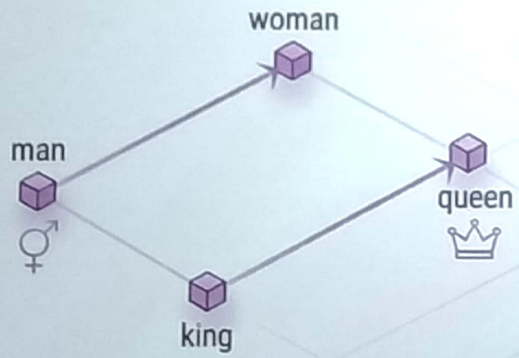
The model uses the context vector to prodict the most likely noxt word by assigning a probability score to all words in its vocabulary.

The Highest Probability Word is the Output

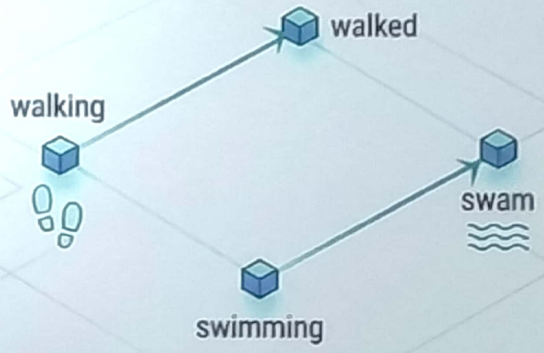
The model selects the word with the highest score (e.g., 'over', 'around', 'quickly') as its final output.

Visualizing Word Relationships: How Word Embeddings Understand Language

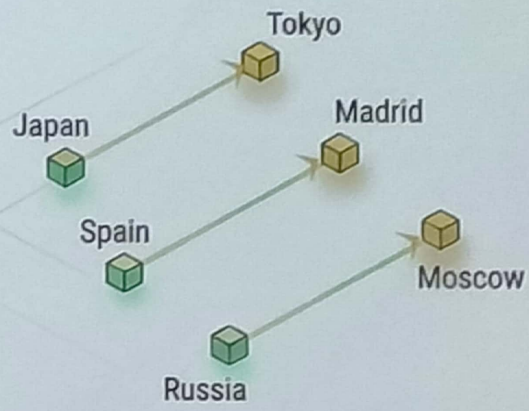
The Gender Relationship



The Verb Tense Relationship



The Geographic Relationship



Capturing Abstract Concepts like Gender

Word embeddings can learn the relationship between male and female concepts. The vector pointing from "man" to "woman" is almost identical to the vector from "king" to "queen".

Vector Math for Analogies

Vector(king) − Vector(man) + Vector(woman) ≈ Vector(queen)

Understanding Grammatical Rules

The model can also learn grammatical patterns, like the relationship between a verb's present tense and its past tense.

Consistent Tense Transformation

The vector that transforms "walking" into "walked" is the same vector that transforms "swimming" into "swam," showing a learned rule.

Learning Real-World Facts

Word embeddings can capture factual relationships, such as the one between a country and its capital city.

Parallel Vectors, Consistent Meaning

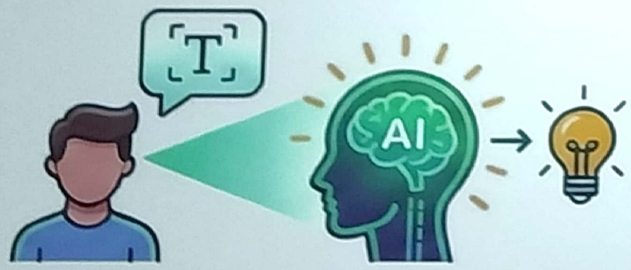
The vectors pointing from each country to its capital (e.g., Japan to Tokyo, Spain to Madrid, Russia to Moscow) are all parallel and similar in length, representing the single concept of "is the capital of".

KEY CONCEPTS IN AI INTERACTION: PROMPTING, COT & REASONING

PROMPTING

What is Prompting?

The fundamental act of giving instructions or queries to an AI model to elicit a desired response.



How is it done?

Crafting clear, concise, and effective instructions for an AI.

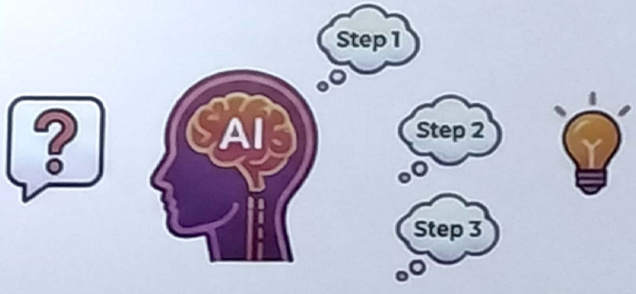
Why is it important?

Guiding AI behavior and achieving accurate results.

COT PROMPTING

What is COT Prompting?

An advanced technique (Chain-of-Thought) encouraging AI to show its step-by-step thinking process.



How does it work?

Structuring a prompt to ask the AI to "think step-by-step" before providing a final answer.

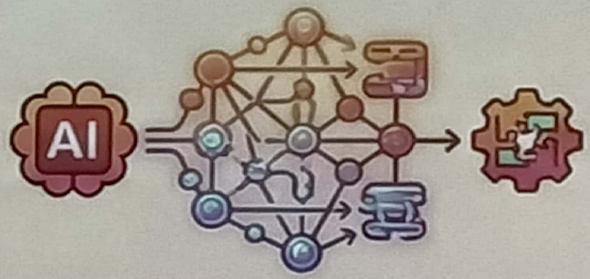
Why is it used?

Improving accuracy on complex problems.

REASONING

What is AI Reasoning?

The ability of an AI to infer, deduce, and make logical connections beyond simply pattern matching.



How is it achieved?

Enabled by model architectures and mechanisms that allow for demonstrating reasoning capabilities.

Why does it matter?

Significant for solving complex, multi-step problems and building more sophisticated AI systems.