
epam
Promps;
Grounding
JUNE 2025
DIALX
COMMUNITY
POWERED BY epam

There are no magic! It is simple REST API!

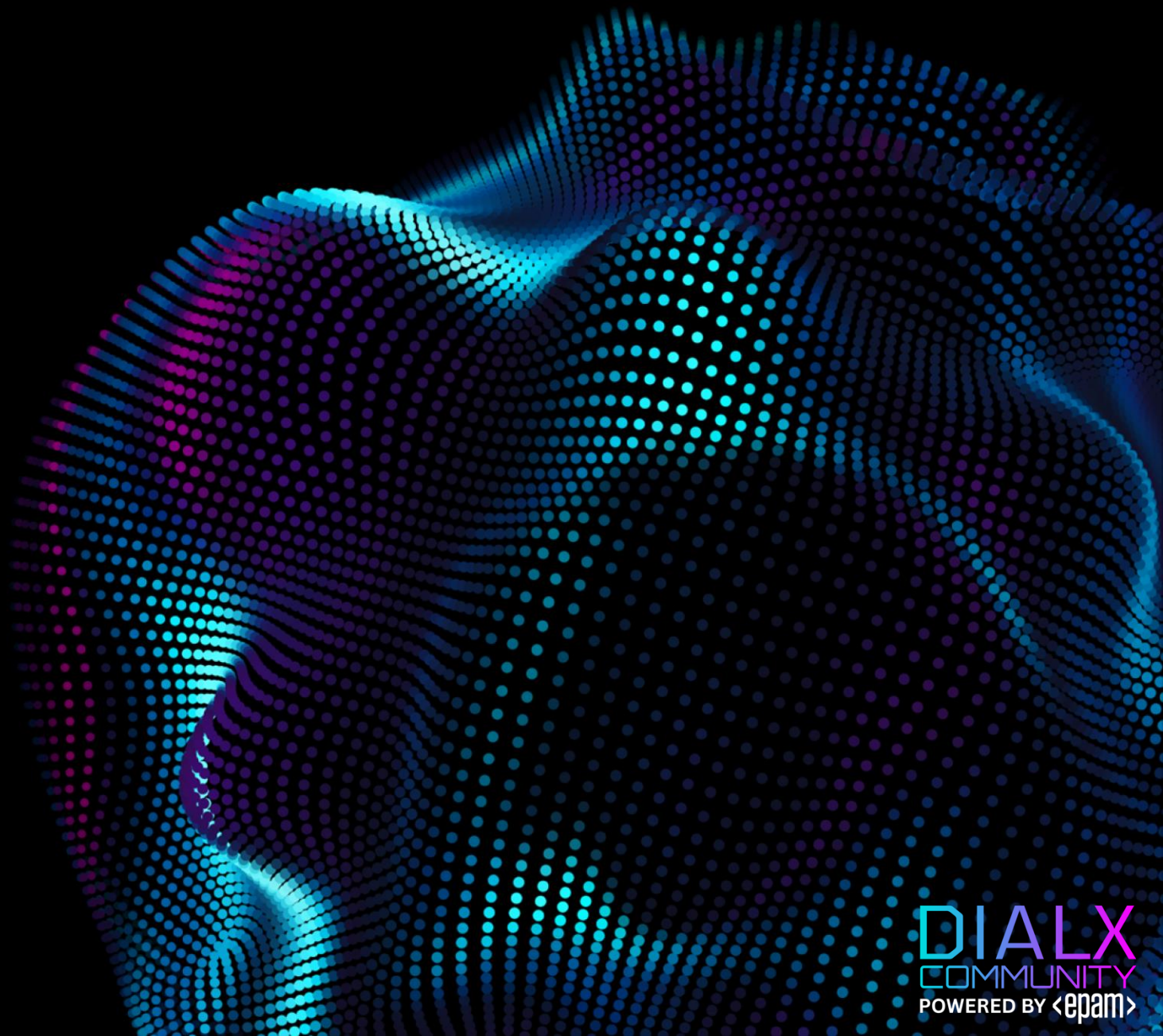
Before we start:

- Raise your hand and ask questions if you have any
- It is better to ask questions when you have
- Also, type them in chat
- We will need DIAL API key for this session

Agenda:

- **Presentation:**
  - **Prompt patterns**
  - **Grounding**
- **Workshop:**
  - **Write prompts with different patterns**
  - **Implement Input RAG-based Grounding**

Prompts



Zero-shot:

Definition:
The model is asked to perform a task without being given any examples. You rely solely on clear instructions

Use When:
- The task is simple or well-known.
- You want minimal prompt length.

Sample:
Is the number 19 a prime number? Answer "Yes" or "No".

One-shot:

Definition:

The model is given exactly one example of the task before being asked to complete a similar task.
It's a middle ground between zero-shot (no examples) and few-shot (multiple examples).

Use When:

- When just one high-quality example is enough to demonstrate the desired structure or behavior.
- When you want to minimize token usage but still guide the model.
- When the model might otherwise misinterpret the task format or intent.

Sample:

Q: Is 11 a prime number?
A: Yes

Q: Is 19 a prime number?
A:

Few-shot:

Definition:
The model is given a few input-output examples before being asked to perform the task. This helps it generalize the pattern.

Use When:
- The task requires learning from examples.
- Output format is non-obvious or needs consistency.

Sample:
Q: Is 4 a prime number?
A: No

Q: Is 7 a prime number?
A: Yes

Q: Is 19 a prime number?
A:

Chain-of-Thought:

Definition:
You guide the model to reason step-by-step before giving the answer. This improves performance on complex or reasoning-based tasks.

Use When:
- The task involves reasoning, calculation, or logic.
- You want more accurate and interpretable answers.

Sample:
Is 19 a prime number? Explain your reasoning.
Let's think step by step.

Role based:

Definition:

A prompt that assigns the AI a specific role with clear context, structured instructions, and defined parameters for how to approach and format the response.

Use When:

- You need domain-specific expertise and knowledge
- The task requires a particular perspective or professional approach
- You want responses that match a specific professional standard or methodology
- The context requires understanding of role-specific responsibilities and constraints

Sample:

Role: You are a professional email writer.
Context: You work for a tech startup that values clear, concise communication.
Task: Transform casual messages into professional emails.
Guidelines:
- Use formal language
- Include proper greeting and closing
- Maintain the original message intent
Format: Professional email format
Input: "hey can u send me the report by friday thx"

Persona pattern:

Definition:

A prompt that creates a detailed character persona for the AI, including specific personality traits, expertise background, communication style, and behavioral characteristics that shape how the AI approaches and responds to tasks.

Use When:

- You want the AI to adopt a specific personality and communication style
- The task benefits from a particular character's unique perspective or expertise
- You need responses that feel more human and relatable
- The context requires understanding of character-specific traits and behaviors

Sample:

Act as a medieval historian who specializes in European castle architecture.
You have a passion for storytelling and often use analogies from daily medieval life.
You're enthusiastic about sharing knowledge and tend to paint vivid pictures with your words.
Explain the defensive features of a motte-and-bailey castle.

Constraint pattern:

Definition:

A prompt that sets explicit, specific limitations and boundaries that restrict how the AI can respond, including constraints on length, format, content, style, audience, or scope.

Use When:

- You need responses that meet specific requirements or limitations
- The output must fit within particular boundaries (length, format, style)
- You want to ensure the AI focuses on specific aspects while avoiding others
- The task requires balancing multiple competing constraints

Sample:

Act as a medieval historian who specializes in European castle architecture.
Write a product description for a smartphone.
Constraints:
- Exactly 50 words
- Include 3 key features
- Use active voice only
- Target audience: tech-savvy millennials
- Avoid technical jargon

Template pattern:

Definition:

A prompt that provides a rigid, predefined structural template with specific sections, labels, and formatting that the AI must follow exactly without deviation. Can be used markdown, XML, JSON and other structured data formats.

Use When:

- You need consistent, structured output across multiple uses
- The response must follow a specific format or standard
- You want to ensure all necessary sections are included
- The task requires systematic organization of information

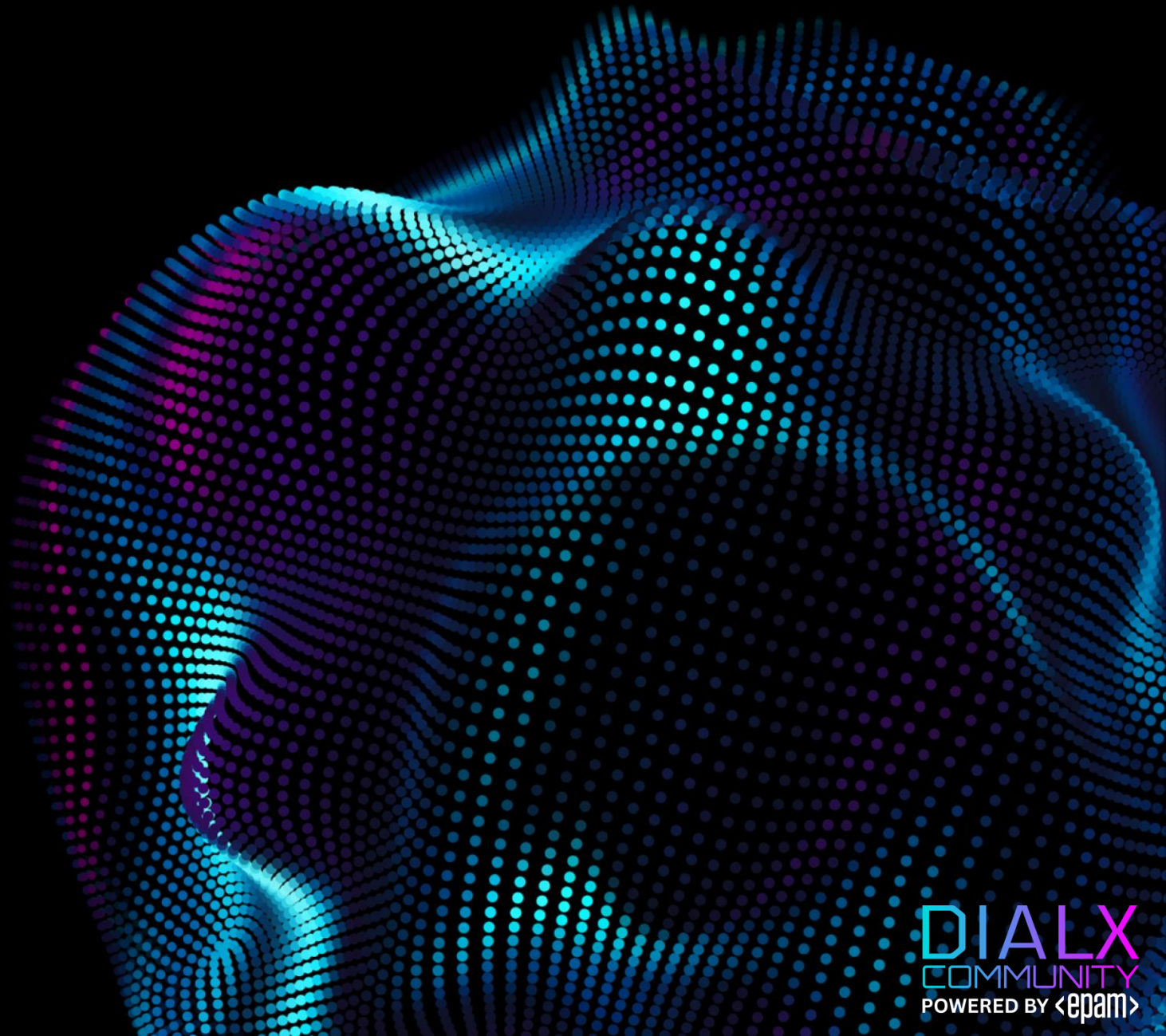
Sample:

Analyze the following business scenario using this template:

SITUATION: [Brief description]
STAKEHOLDERS: [List key parties involved]
CHALLENGES: [Identify main problems]
OPPORTUNITIES: [Potential benefits]
RECOMMENDATIONS: [Specific actions]
NEXT STEPS: [Immediate actions needed]

Scenario: A local restaurant wants to start online delivery

Grounding



Grounding:

Definition:

Grounding is a prompt engineering technique used in LLMs where the user provides specific and relevant context within the prompt. This directly increases the accuracy and relevance of the response by connecting the model's output to verifiable data sources.

Benefits:

- ***Reduces Hallucinations:*** **By grounding responses in factual data, RAG reduces the chances of generating incorrect or fabricated information**
- ***Updated Knowledge:*** **Without the need for fine-tuning, updating the information contained in the knowledge base is sufficient for the model to access new and verified information**
- ***Domain Specificity:*** **Improves LLM precision when working with documents it has never seen before and in highly specific domains**

Grounding:

RAG-based

Augmenting the prompt with relevant documents retrieved from a knowledge base (e.g., database, vector store, web).

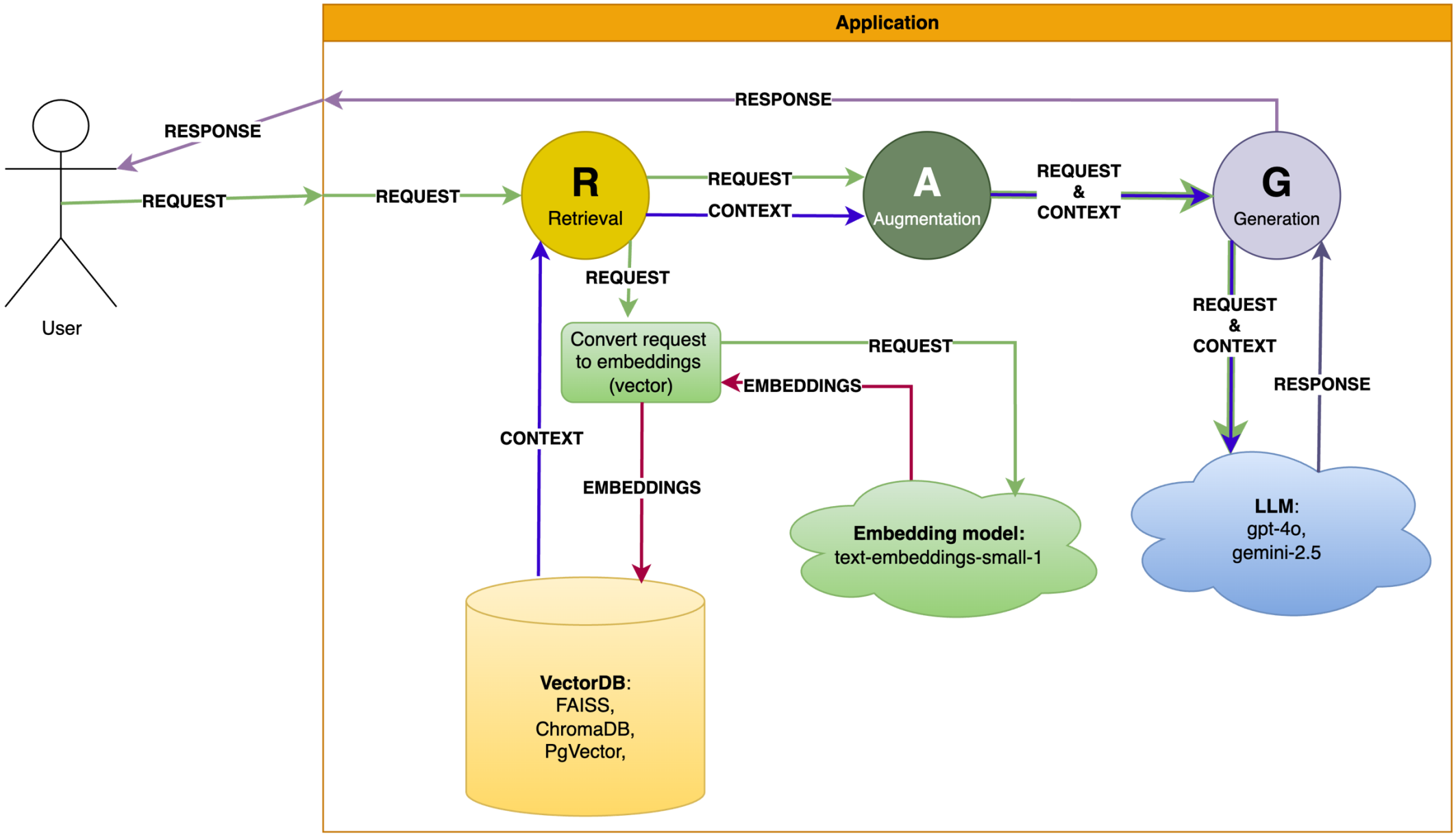
Data grounding

Using structured data from trusted databases, APIs, or business systems.

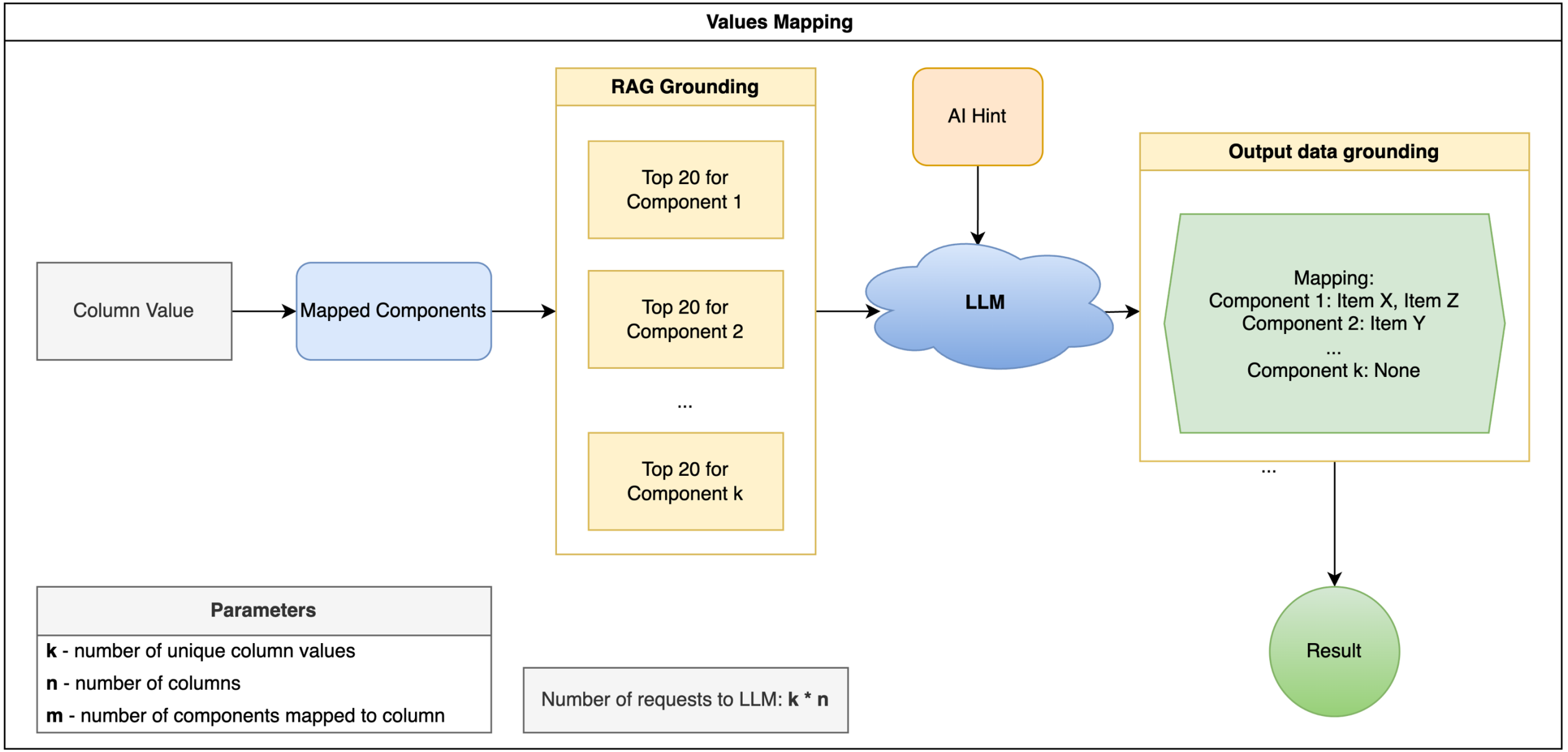
Human-in-the-Loop

Having human reviewers or validators approve or correct model outputs.

RAG-based grounding:



In&Out grounding:



DIALX COMMUNITY
POWERED BY epam

Join us:

Subscribe to WeAreCommuntiy

https://wearecommunity.io/communities/dial

Keep in touch with our latest updates. Here you find webinars, workshops and articles about DIALX features and products.

Subscribe to YouTube

https://www.youtube.com/@TeamDIALX

Here we publish videos about our newest products and features.

Join our Discord community

https://discord.gg/jvTCQv4E4q

✨AI DIALX Community✨ is the place where you can find help with your questions about DIALX, direct communication with DIALX team and contributors.


DIALX
COMMUNITY
POWERED BY epam
Thank you!