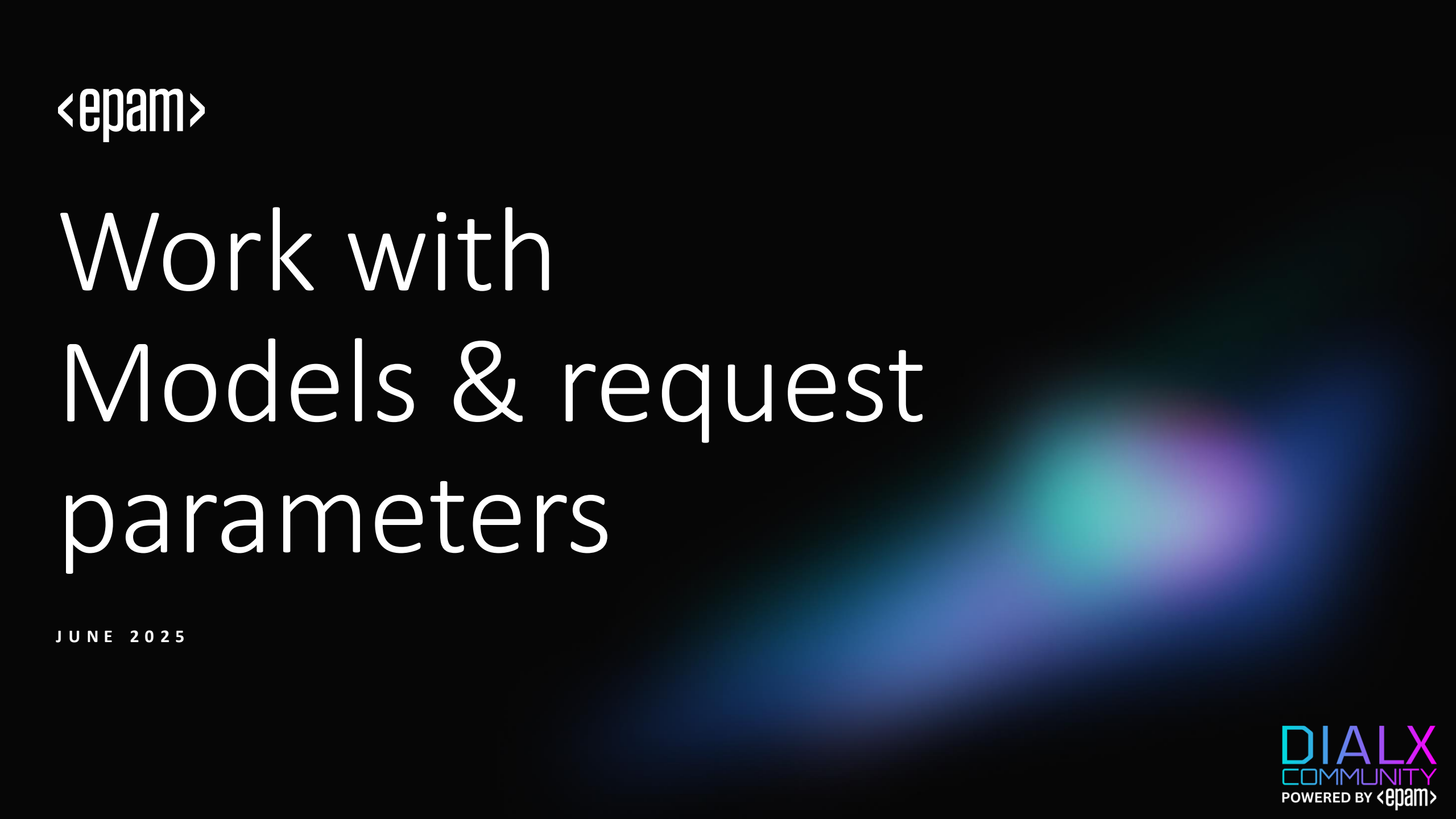

epam
Work with Models & request parameters
JUNE 2025
DIALX COMMUNITY
POWERED BY epam

There are no magic! It is simple REST API!

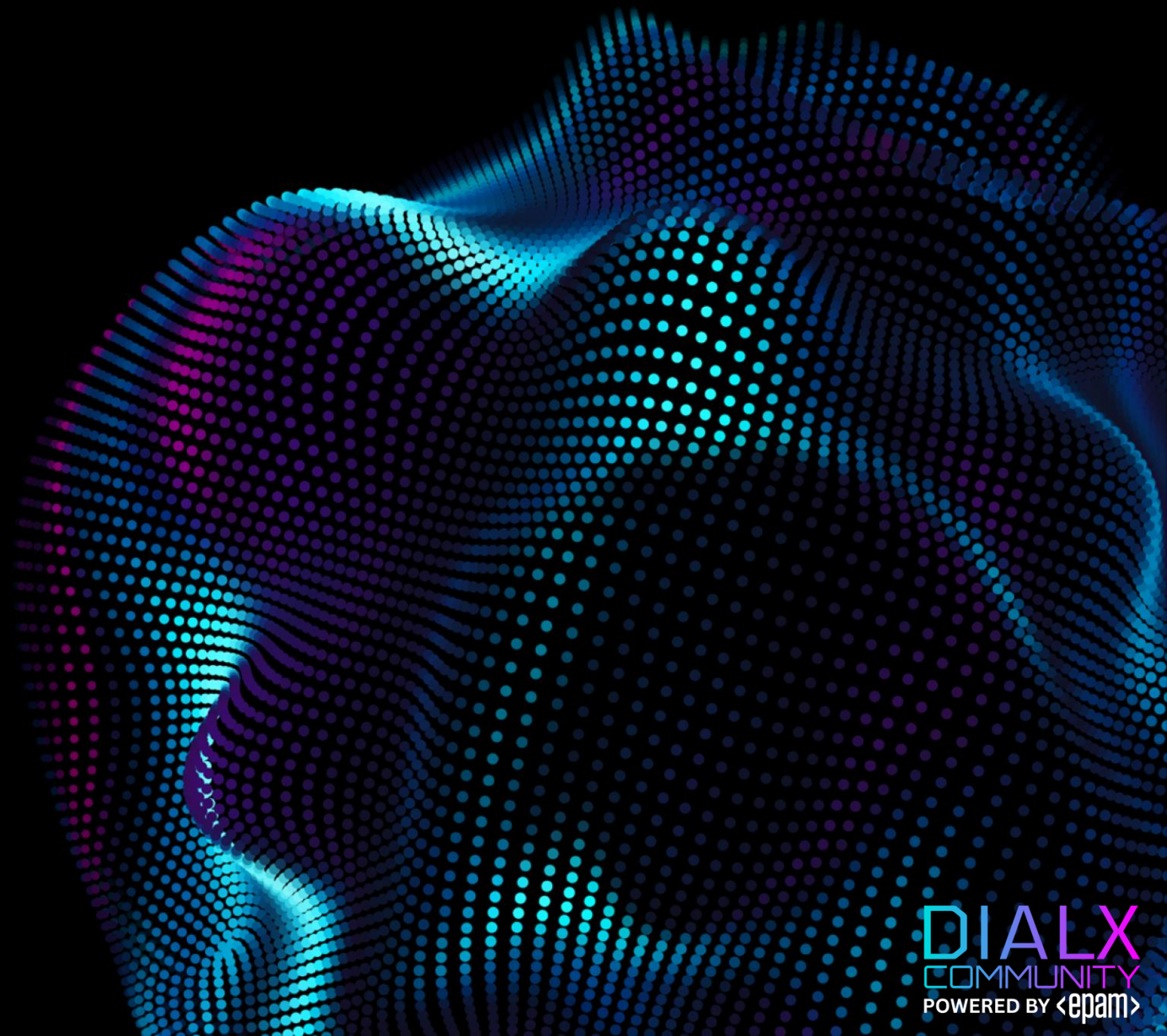
Before we start:

- 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:**
  - **Models**
  - **Request parameters**
- **Workshop:**
  - **Work with different models via DIAL API**
  - **Work with different request body parameters and explore what impact they have on the output**

Models



Models:

DIAL API, request body: https://dialx.ai/dial_api#operation/sendChatCompletionRequest

Get available models: https://ai-proxy.lab.epam.com/openai/models

OpenAI

https://platform.openai.com/docs/models

- General-purpose Agent: **ChatGPT**
- Models:
  - ***gpt-4.1-2025-04-14***
  - ***gpt-4o-2024-05-13***
  - ***gpt-4o-mini-2024-07-18***
  - ***gpt-35-turbo-1106***

Anthropic

https://anthropic.com/models

- General-purpose Agent: **Claude**
- Models:
  - ***anthropic.claude-v3-5-sonnet-v1***
  - ***anthropic.claude-sonnet-4-20250514-v1:0***
  - ***anthropic.claude-3-7-sonnet-20250219-v1:0***
  - ***anthropic.claude-v3-5-haiku***

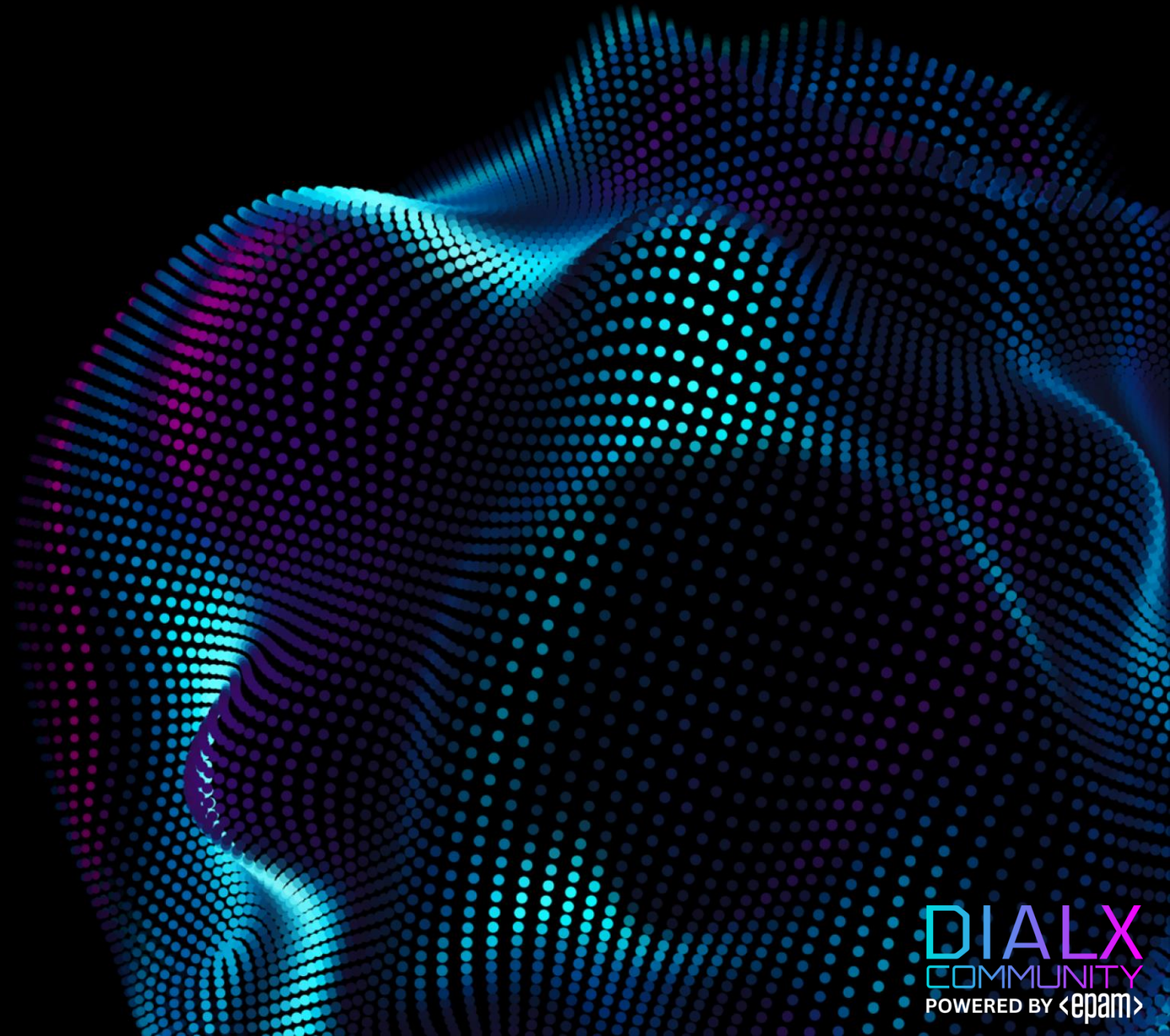
Google

https://google.com/generative-ai/models

- General-purpose Agent: **Gemini**
- Models:
  - ***gemini-2.5-pro-preview-03-25***
  - ***gemini-1.5-pro-google-search***
  - ***gemini-2.5-flash-preview-04-17***
  - ***gemini-2.5-pro-preview-03-25-google-search***

Parameters

<Request and its params>



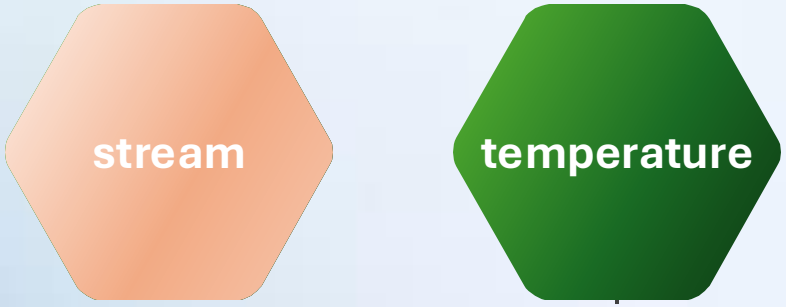
DIALX COMMUNITY POWERED BY <epam>

Parameters:

Enables response streaming, where the server sends back partial responses as they are generated.
This is useful for applications needing real-time updates.

Default value: false

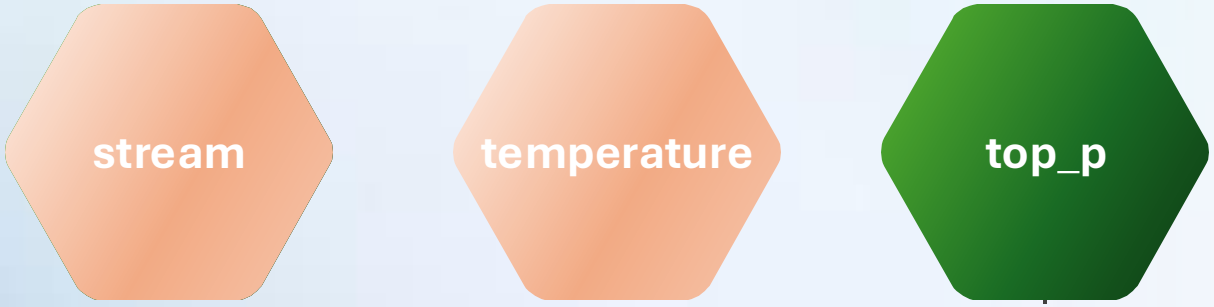
Parameters:



Controls the randomness of the output.
It’s a parameter for balancing creativity and determinism.

Range: 0.0 to 2.0
Default value: 1.0

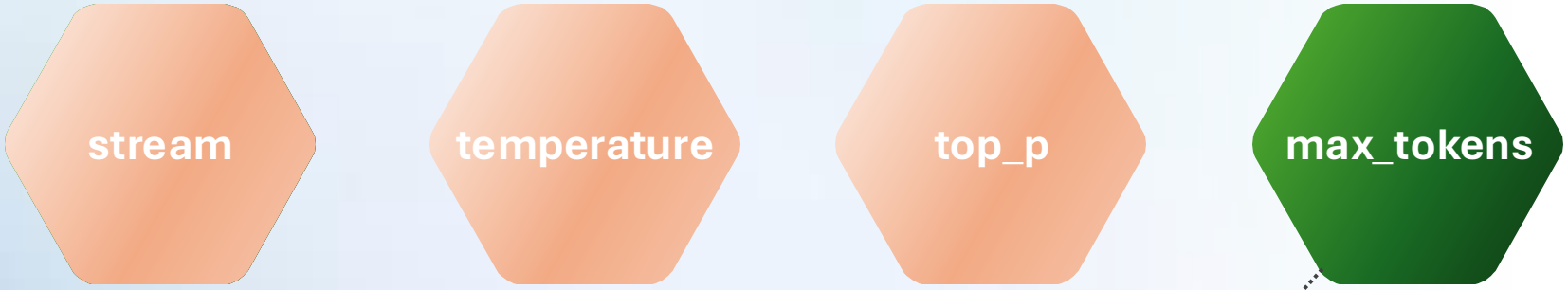
Parameters:



An alternative to “temperature” for controlling diversity
Via nucleus sample.

Range: 0.0 to 1.0
Default value: 1.0
Recommended to use “top_p” OR “temperature” not both

Parameters:



Sets the maximum number of tokens to generate in the response.

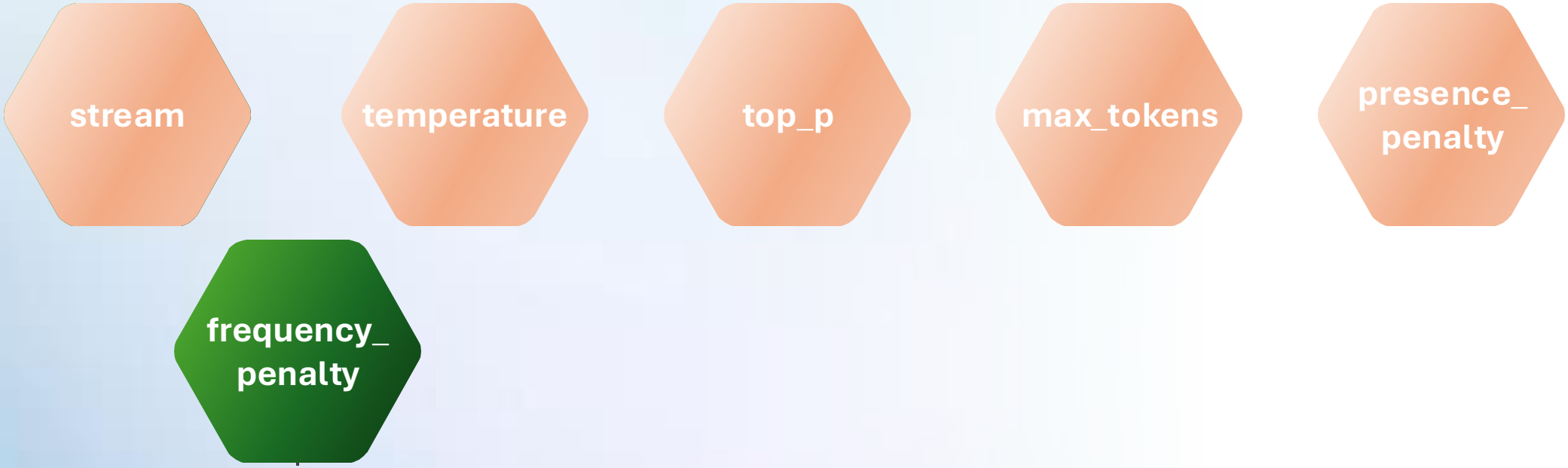
Default value: ---

Parameters:

Penalizes new tokens based on whether they appear in the text so far, encouraging the model to discuss new topics.

Range: -2.0 to 2.0
Default value: 0.0
Positive: Drives novelty and creativity.
Negative: Keeps the model focused on familiar ideas, potentially leading to more repetitive outputs.

Parameters:



Penalizes new tokens based on their frequency in the text so far, reducing verbosity or repetitiveness.

Range: -2.0 to 2.0
Default value: 0.0
Negative values encourage more repetition.
Positive values reduce token repetition.

Parameters:

stream

temperature

top_p

max_tokens

presence_ penalty

frequency_ penalty

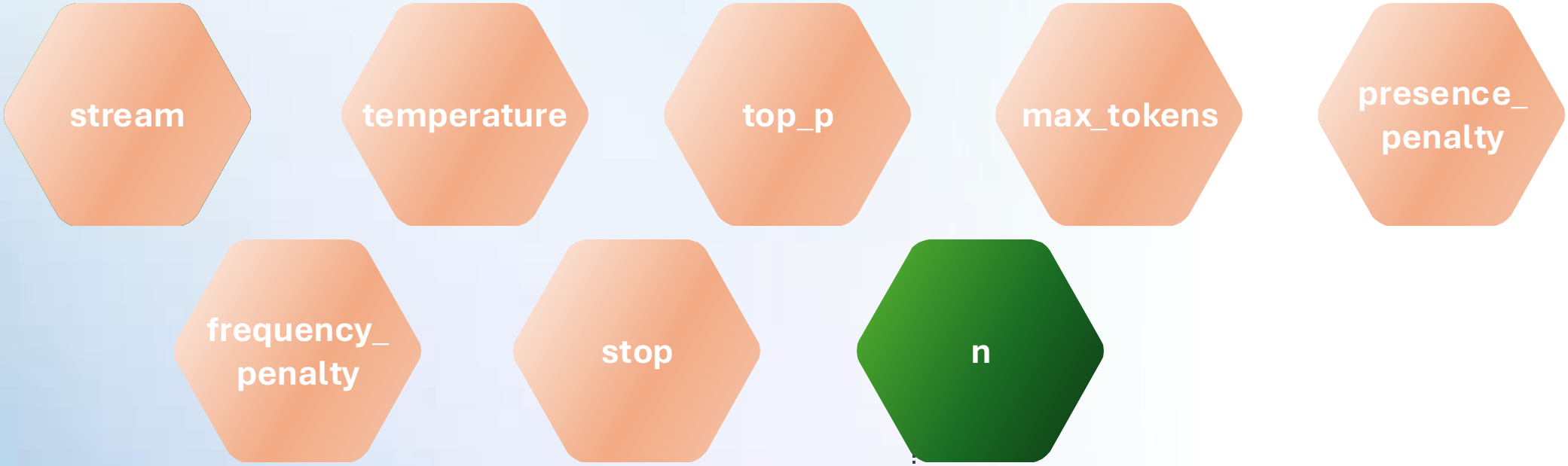
stop

Specifies one or more sequences where the API should stop generating further tokens.

[“\n”, “DIALX”]
Default value: null

Specifies one or more sequences where the API should stop generating further tokens.
Can include multiple stop conditions.

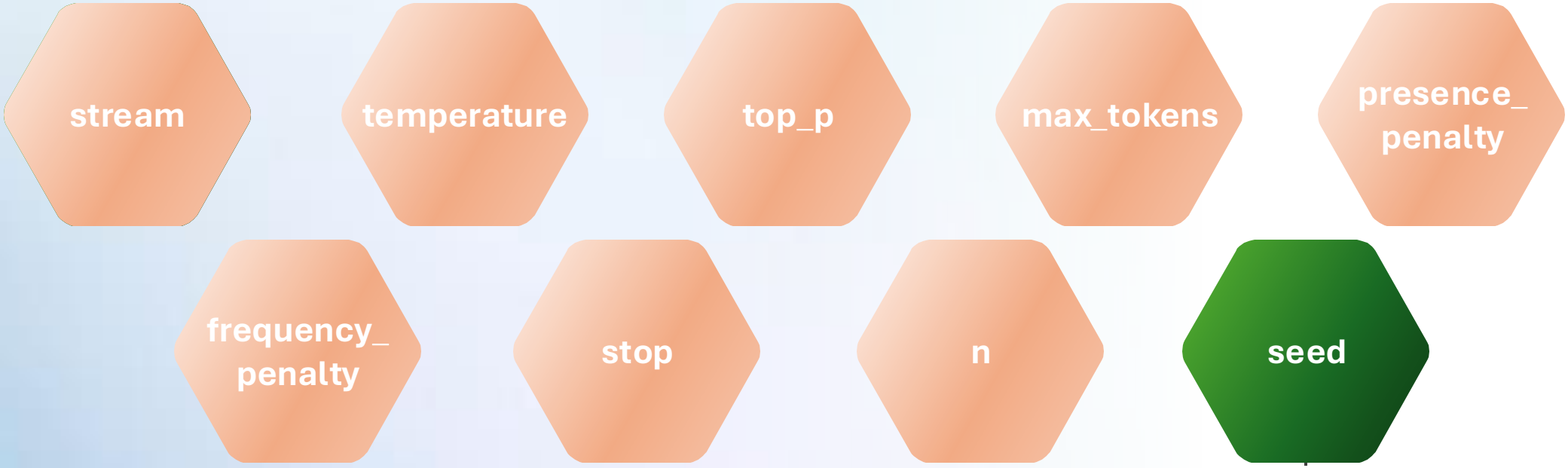
Parameters:



Specifies amount of assistant messages that should be generated.

Default value: 1

Parameters:



Ensures deterministic results for the same input and settings by making the randomness reproducible

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!