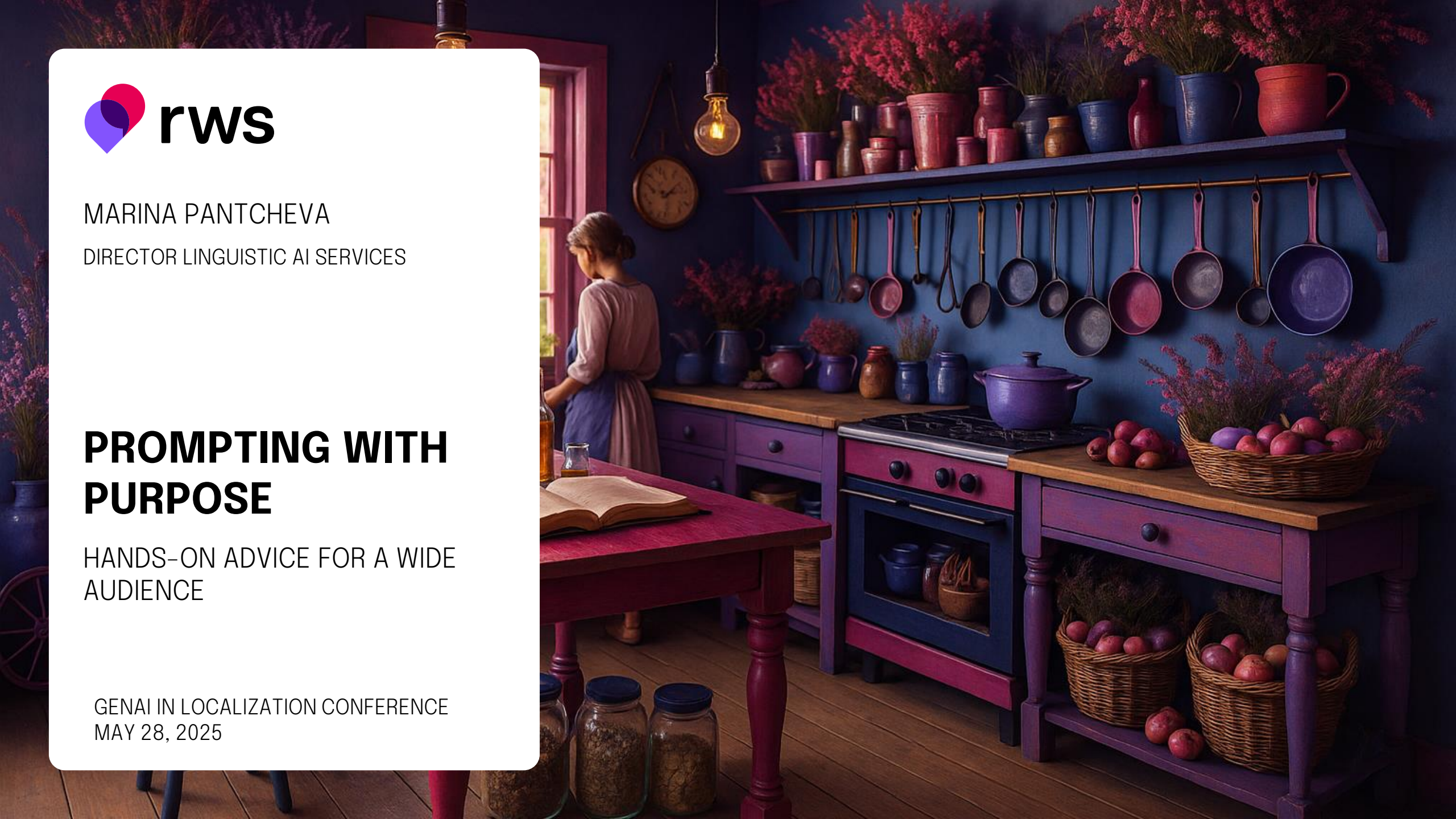


rws

MARINA PANTCHEVA

DIRECTOR LINGUISTIC AI SERVICES

PROMPTING WITH PURPOSE

HANDS-ON ADVICE FOR A WIDE AUDIENCE

GENAI IN LOCALIZATION CONFERENCE
MAY 28, 2025

6 BASIC CONCEPTS IN AI

A prompt is your input into the AI system to obtain specific results.

PROMPT

LLMs break down words into **tokens**. The relation between words and tokens is not straightforward.

TOKENS

The response of an LLM to a prompt is called **completion** (or sometimes more generically **output**).

COMPLETION

A fancy name for '*examples*' in machine learning.

SHOTS

The limit on how many tokens an LLM can "keep in mind" when performing a task.

CONTEXT WINDOW

This is the content you want the LLMs to analyze

PRIMARY CONTENT

rws

LLM EVOLUTION

The great imitators

GPT 3, Llama 1, Mistral 7B

Instruction-tuned

GPT 3.5, Llama 2,
Claude 3.5, Gemini 1.0

Agentic AI

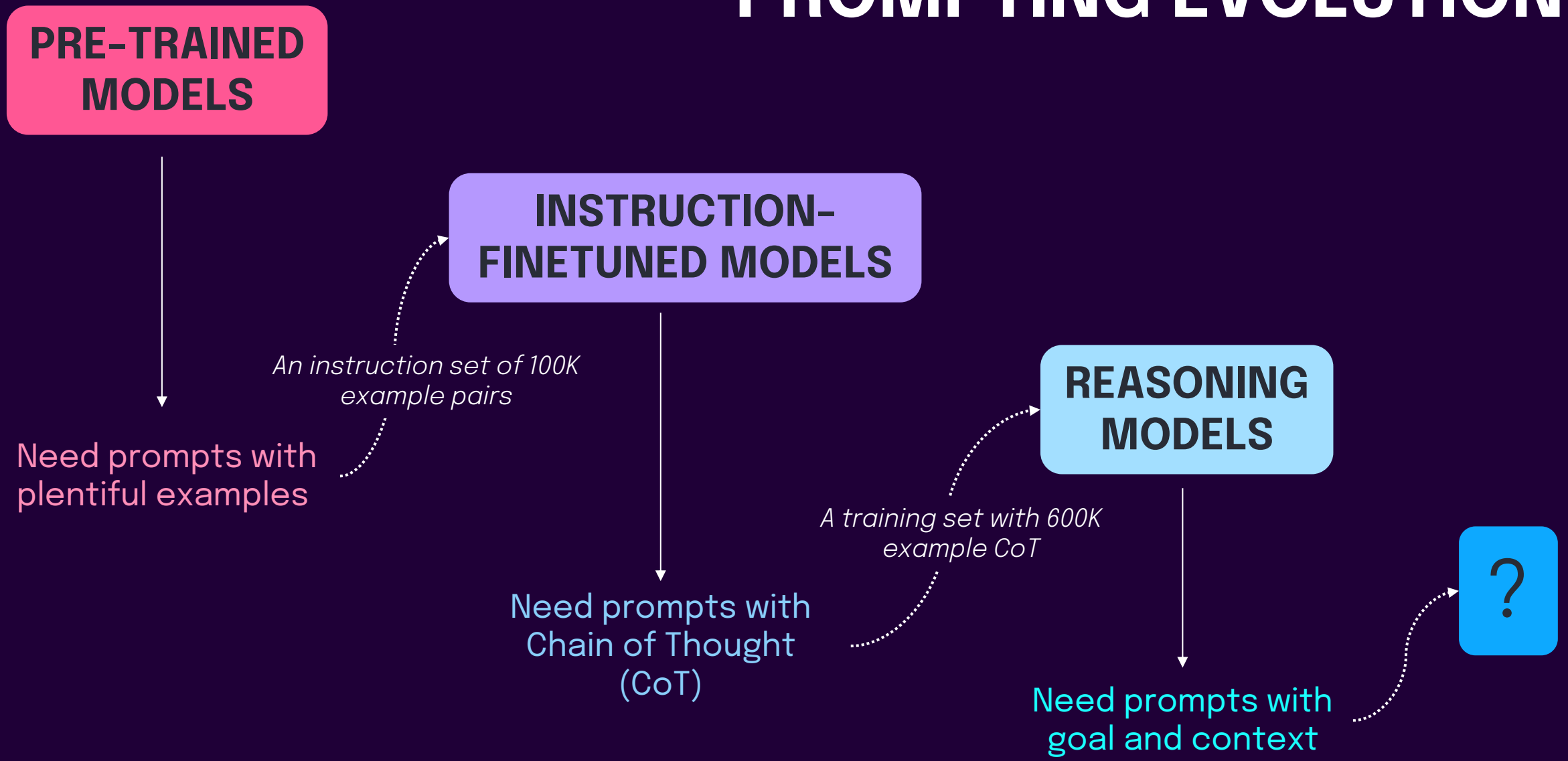
AutoGPT, Operator,
Deep Research

Reasoning models

GPT o1/o3, Gemini Flash
DeepSeek R1/2

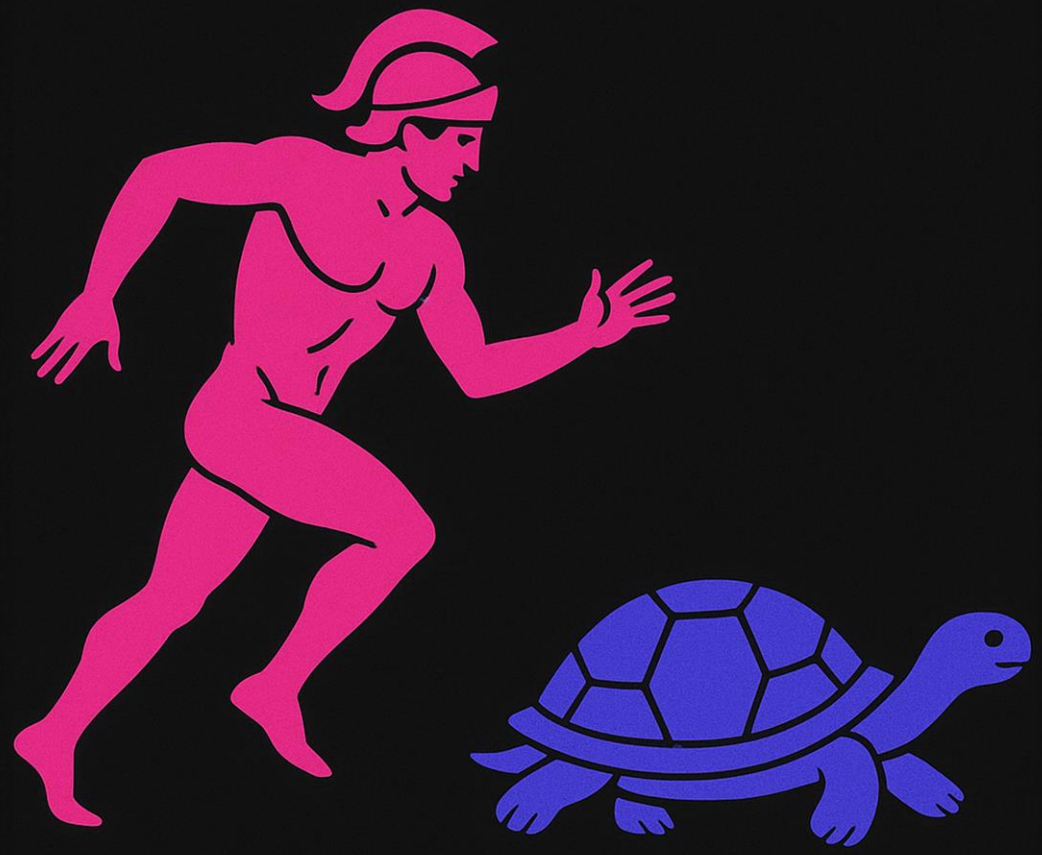
INFANCY CHILDHOOD ADOLESCENCE ADULTHOOD

PROMPTING EVOLUTION



ZENO's PARADOX OF CATCHING UP WITH AI

Chasing AI developments only **halves the distance** to the **latest AI** frontier model.



rws

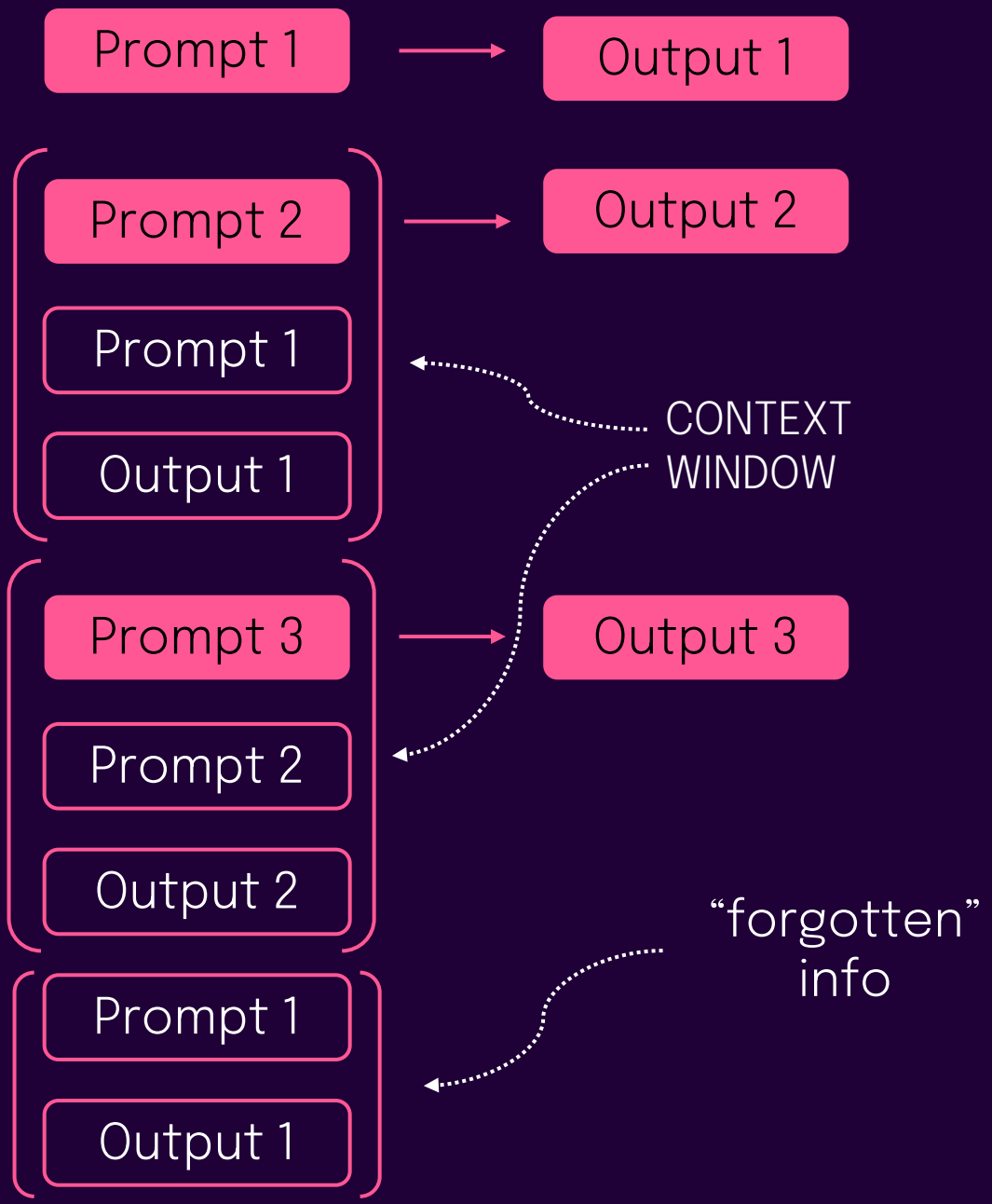
AI PILOTS


Cook (prompt designer)
Manual tasting (evaluation)
Tested recipes (prompts)
Selected ingredients (data)
Small portions (datasets)

CHATTING WITH AI

- Natural language dialogue
- **System prompt imposed**: No control over the actual prompt and parameters.
- **Contextual continuity**: the generated prompt contains all messages in the chat session that fit into the context window.
  This incremental prompting may lead to degradation of response adequacy.

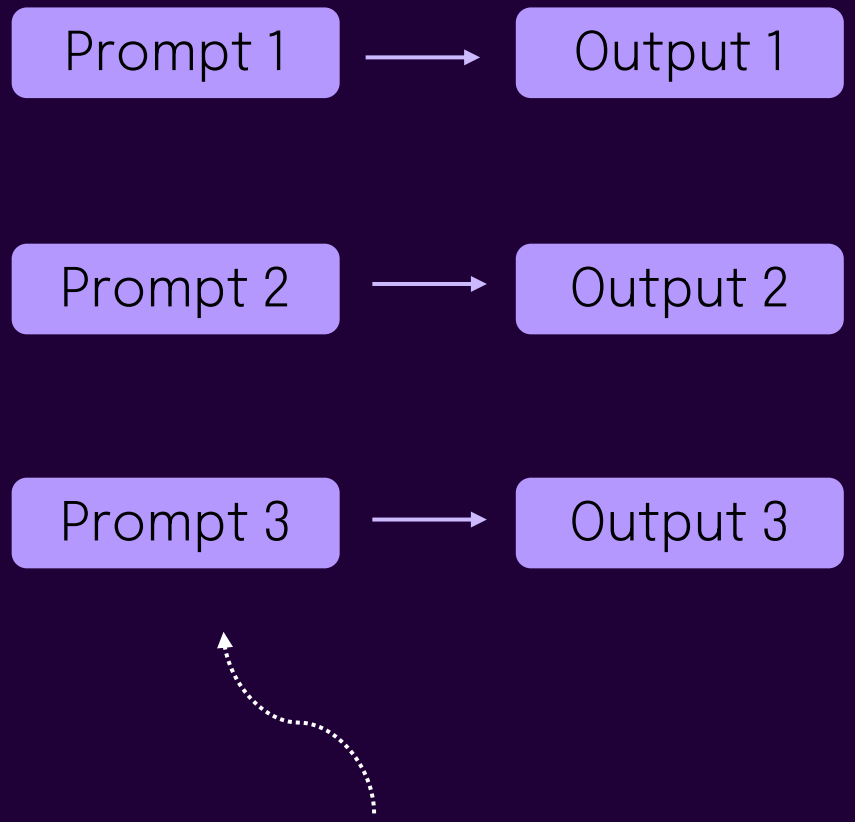
Brainstorming, question answering, assistance, quick proof of concepts



AI IN PRODUCTION

PROMPTING VIA API

- Technical knowledge needed to manage rate limits, authentication, etc.
- Full **control over the prompt** and its parameters
- **No contextual continuity**: context information is included in each call



Automated workflows, repeatable tasks, large scale operations, AI integration in apps

THE TOKEN OVERHEAD PROBLEM

PROBLEM

The token count of the prompt outweighs by far the tokens of the text you want to process.

SOLUTION 1

Group the text segments into **batches** of 30-50.

- Wrap each segment in tags to avoid leakage.
- If concatenating, use unique delimiter that the LLM preserves.
- If using IDs, beware of ID shift.

SOLUTION 2

Extract into the prompt only the reference information that is **relevant** to the processed segments:

- relevant style guide rules
- terms that are indeed present in the segments

Query to an assistant

Prompting non-agents

- Step-by-step instruction
- Context and references
- Desired output format
- A few examples

LLM output:

A **response** –
an answer or converted input

Primary failure:

- wrong/irrelevant answers
- Incorrectly formatted/converted segments.

Standard operating procedures

Prompting agents

- Specification of goal and success
- Catalogue of available tools
- Method and memory
- Guardrails and exit criteria

LLM output:

A **plan**
a JSON scratchpad

An **action**:
a tool call

A **decision**:
a termination signal

Primary failure:

- Wrong actions
- Overspending tokens

USE AI TO PROMPT AI

Have the LLM you plan to use for execution generate the prompt text.

This increases the chance that the prompt aligns with the instructions the LLM has seen in its training set.

Ask the model you use for task execution produce a detailed prompt. Where it fails at instructing, it will also fail at executing.

Use the most advanced model available to draft the prompt. **Reasoning** models excel at this.

PROMPT TROUBLESHOOTING

1. Break down the instructions into discrete steps within the same prompt.
2. Use prompt daisy-chaining where the output from prompt 1 is the input of prompt 2, etc.
3. Pull in more primary content to provide context (RAG).
4. Clean and restructure the primary content (your data).
5. Use a more advanced LLM model by the same provider.
6. Try a different LLM provider.

   (The prompt may have to be adjusted)

Let’s catch the AI tortoise!