AI Factory Austria AI:AT

# Large Language Models Explained

## For Non-Techies

**Dr. Daniel Lehner**
Expert AI Knowledge Transfer @ AI Factory Austria AI:AT

20.10.2026

# Dr. Daniel Lehner



**Expert AI Knowledge Transfer** @ AI Factory Austria AI:AT

AI in **manufacturing** (through digital twins)

- IT Consultant and Trainer
- Business Informatics Researcher (441 citations) [1]

AI for **general public**

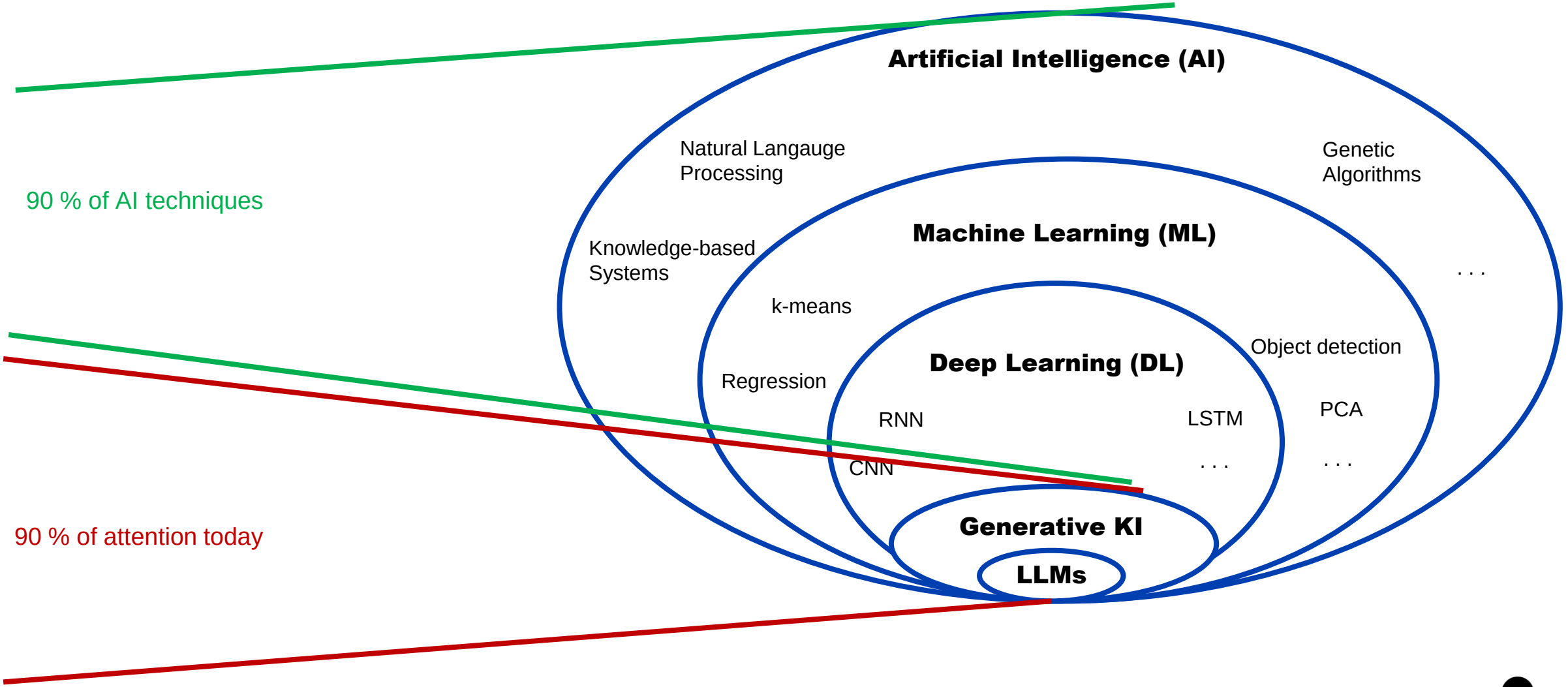
- University lectures AI for law/business administration
- VHS courses AI for general public [2]
- Certified AI Manager trainings [3]

[1] https://scholar.google.com/citations?user=TGGaQ_0AAAAJ

[2] https://vhskurs.linz.at/index.php?kathaupt=18&suchesetzen=false&kfs_dozentid=139372

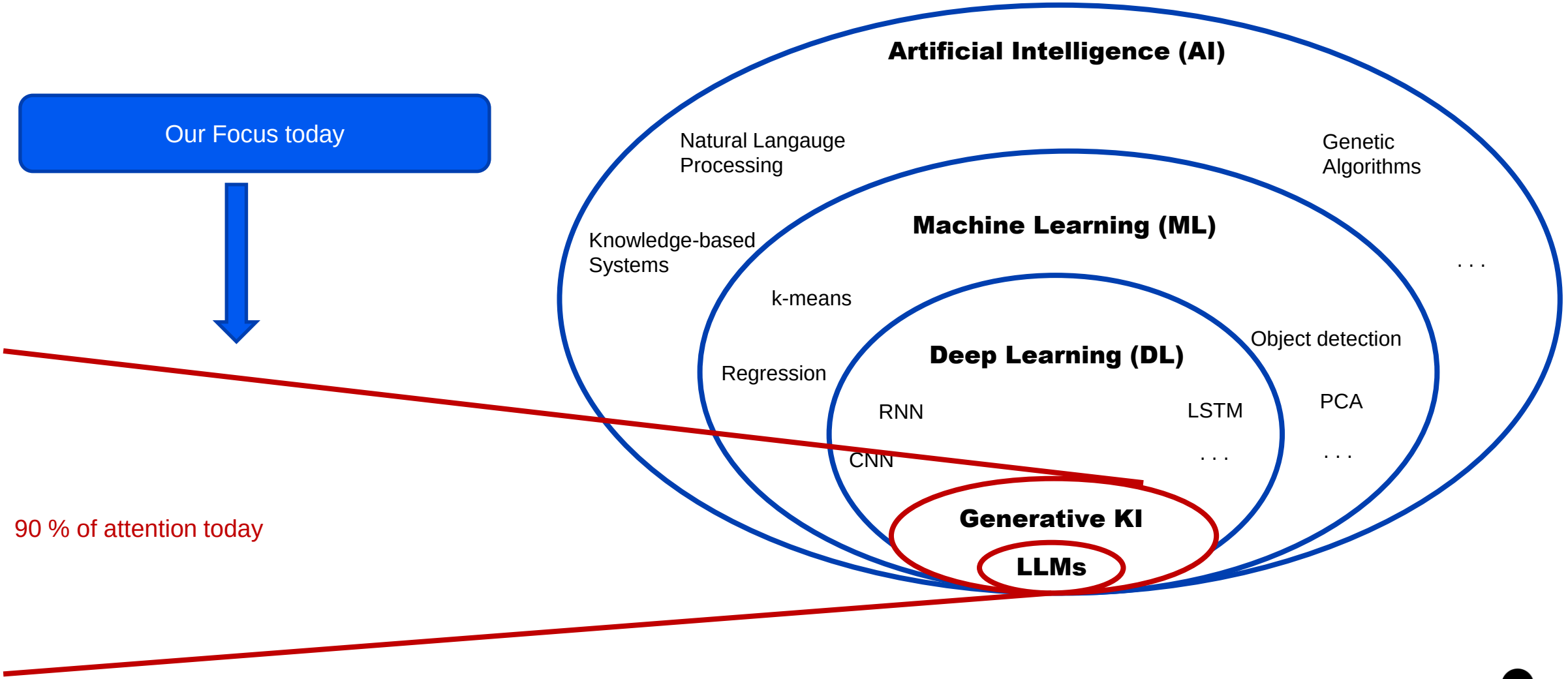
[3] https://tecnovy.com/de/tecnovy/certified-ai-manager

# „AI Tunnel View“



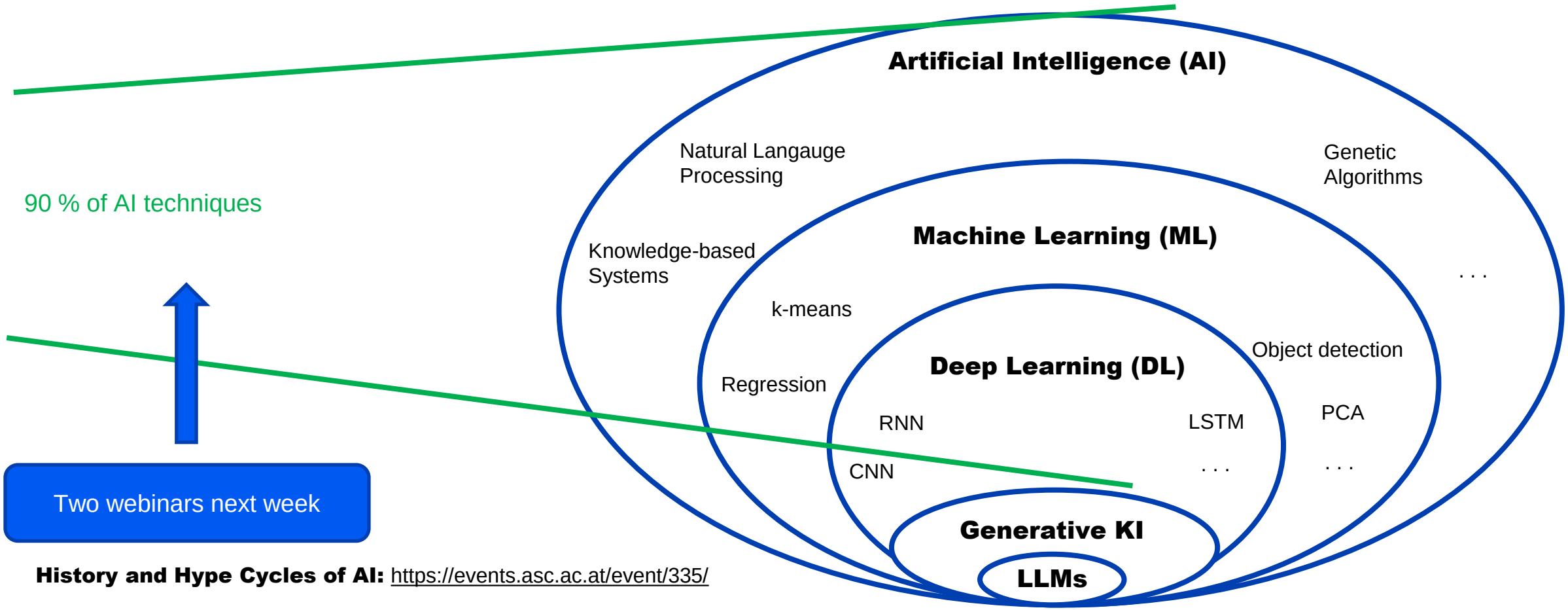

Source: https://www.linkedin.com/posts/jaylatta_llm-tunnel-vision-weve-fallen-into-the-activity-7386990451360387072-y6KI

# „AI Tunnel View“




Source: https://www.linkedin.com/posts/jaylatta_llm-tunnel-vision-weve-fallen-into-the-activity-7386990451360387072-y6KI

# „AI Tunnel View“

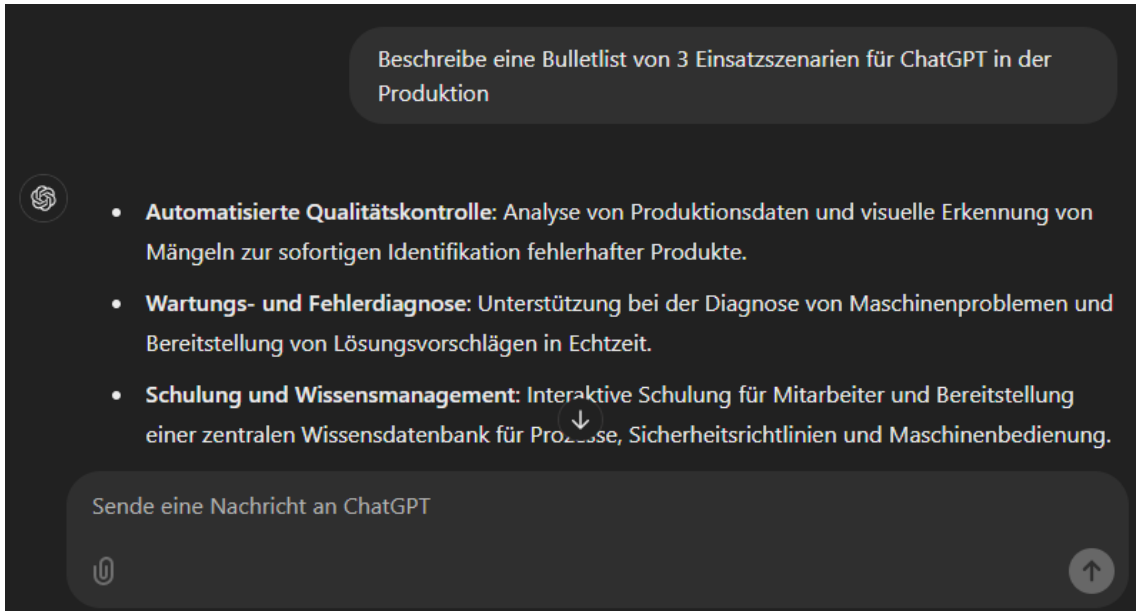


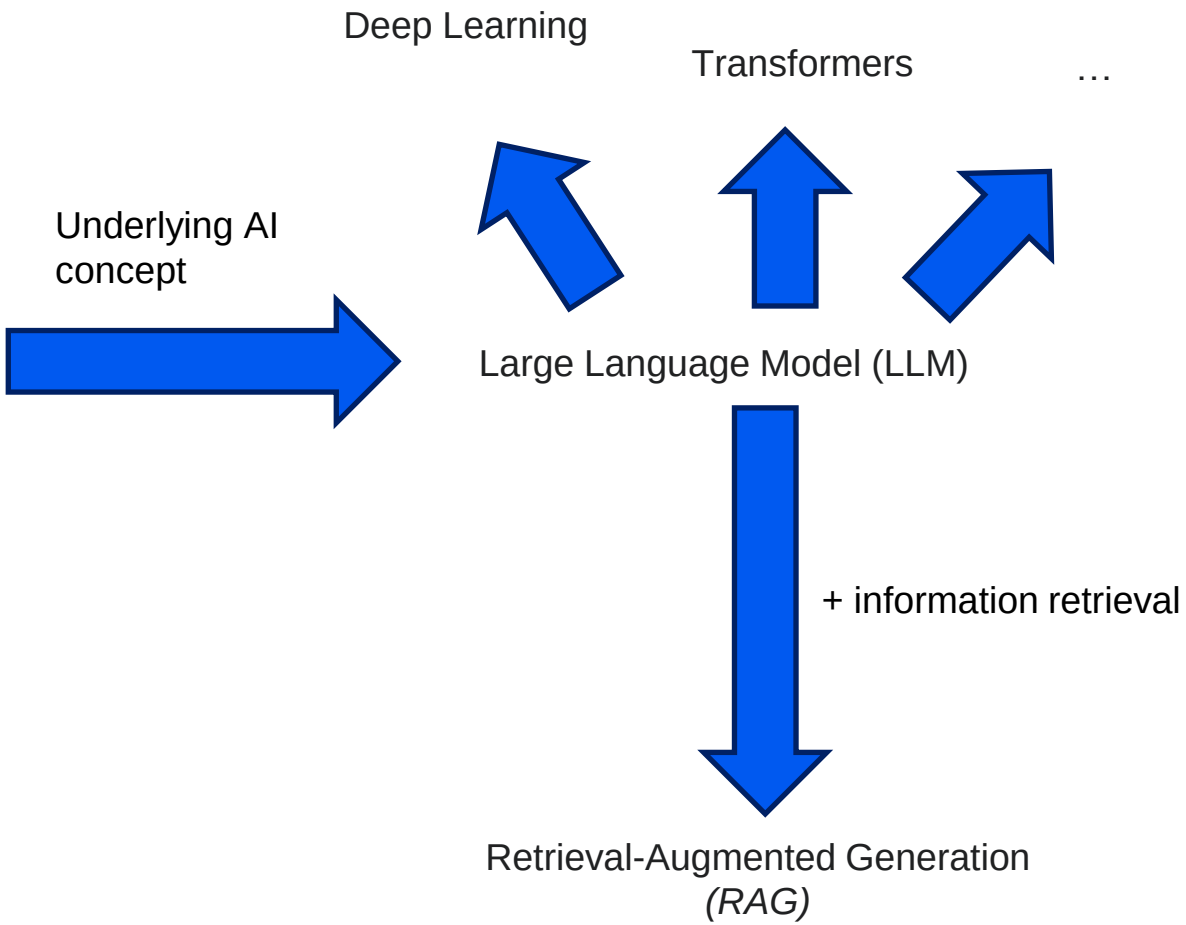

**History and Hype Cycles of AI:** https://events.asc.ac.at/event/335/

**Machine Learning Applications and Foundations:** https://events.asc.ac.at/event/336/

Source: https://www.linkedin.com/posts/jaylatta_llm-tunnel-vision-weve-fallen-into-the-activity-7386990451360387072-y6KI

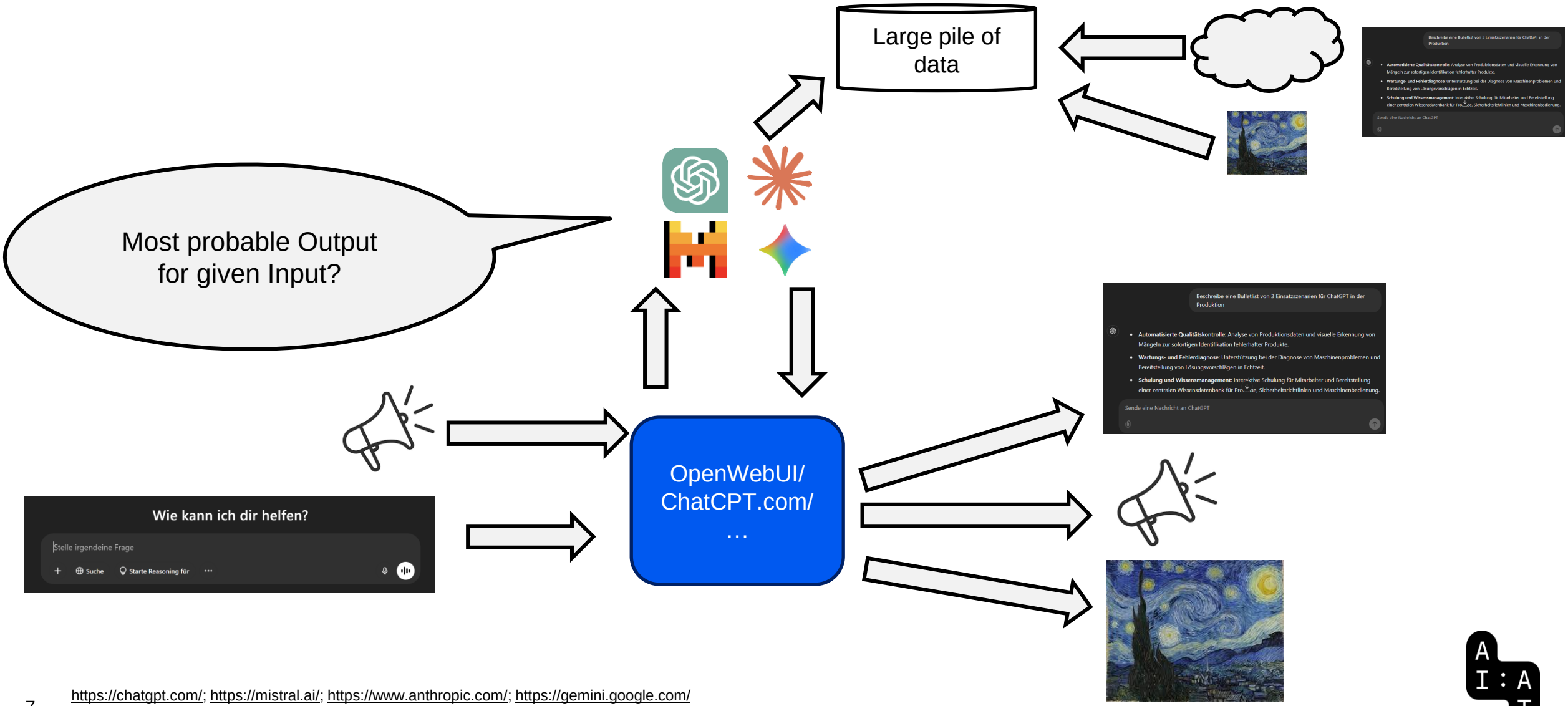
# AI Chatbots



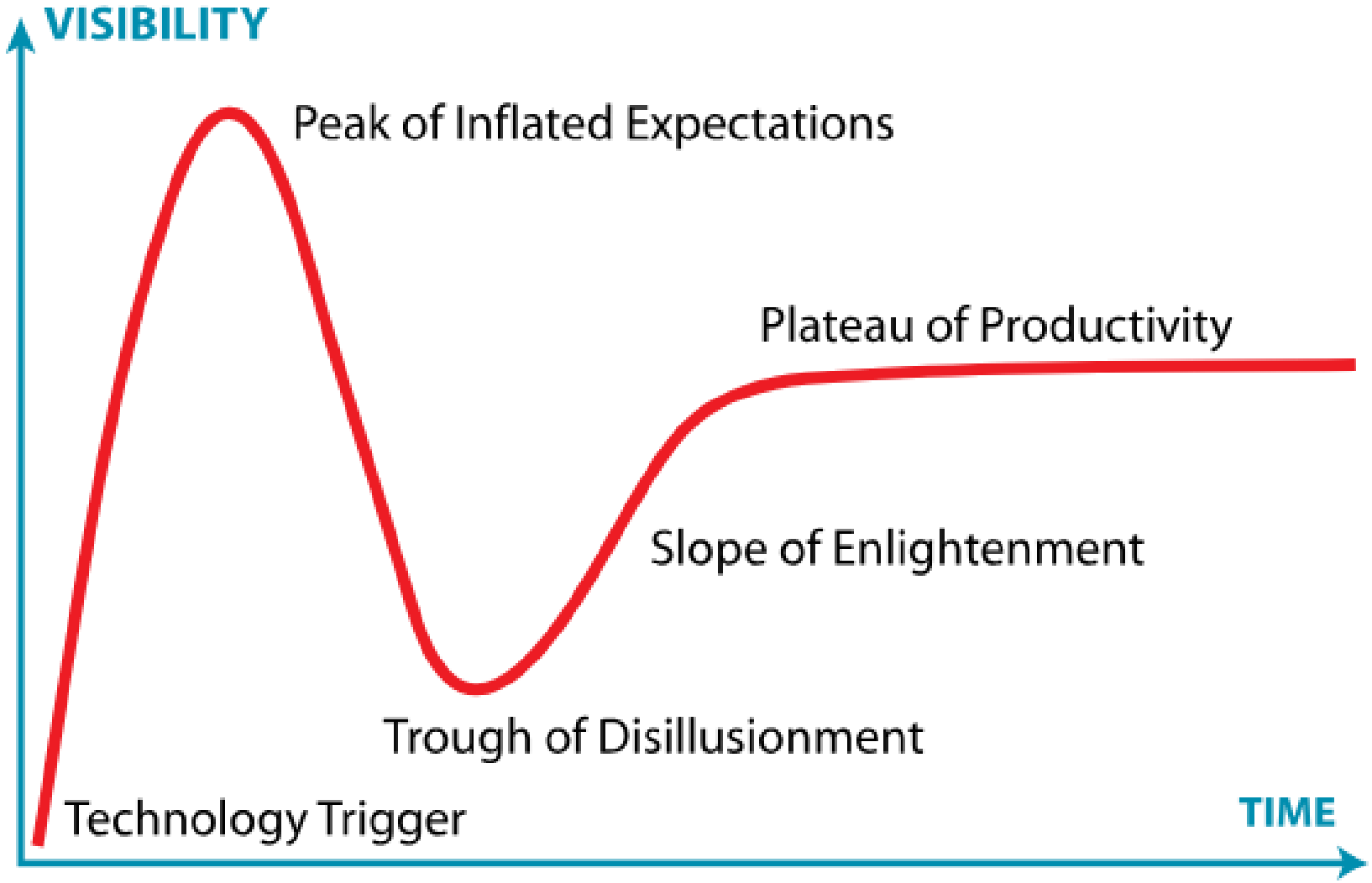

https://chatgpt.com/

# Generative AI



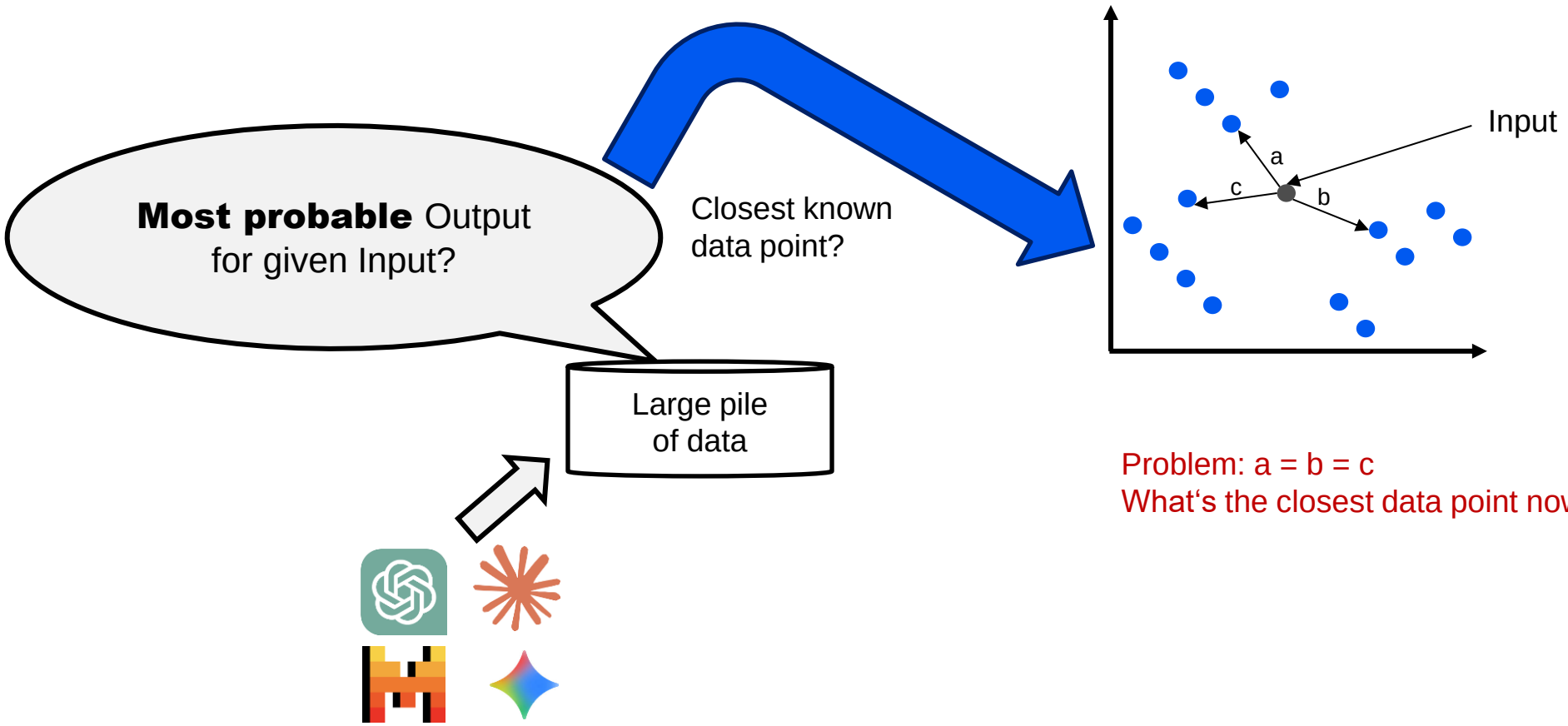

https://chatgpt.com/; https://mistral.ai/; https://www.anthropic.com/; https://gemini.google.com/

# Expectation Management
# Generative Artificial Intelligence




https://en.wikipedia.org/wiki/Gartner_hype_cycle

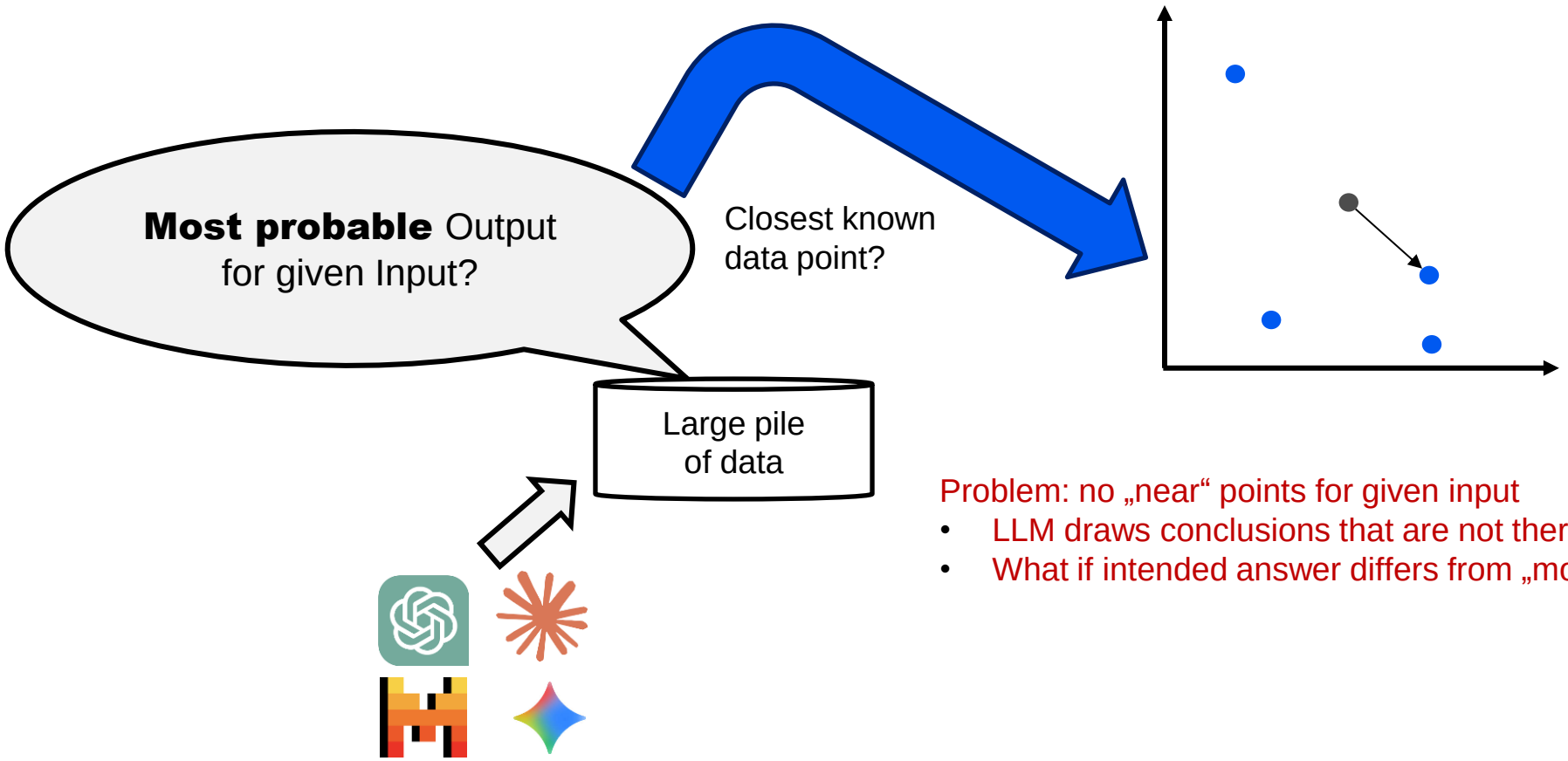
# Challenge 1: Non-Deterministic Result




Problem: a = b = c
What‘s the closest data point now?

https://chatgpt.com/; https://mistral.ai/; https://www.anthropic.com/; https://gemini.google.com/

# Challenge 2: „Halluzinations“

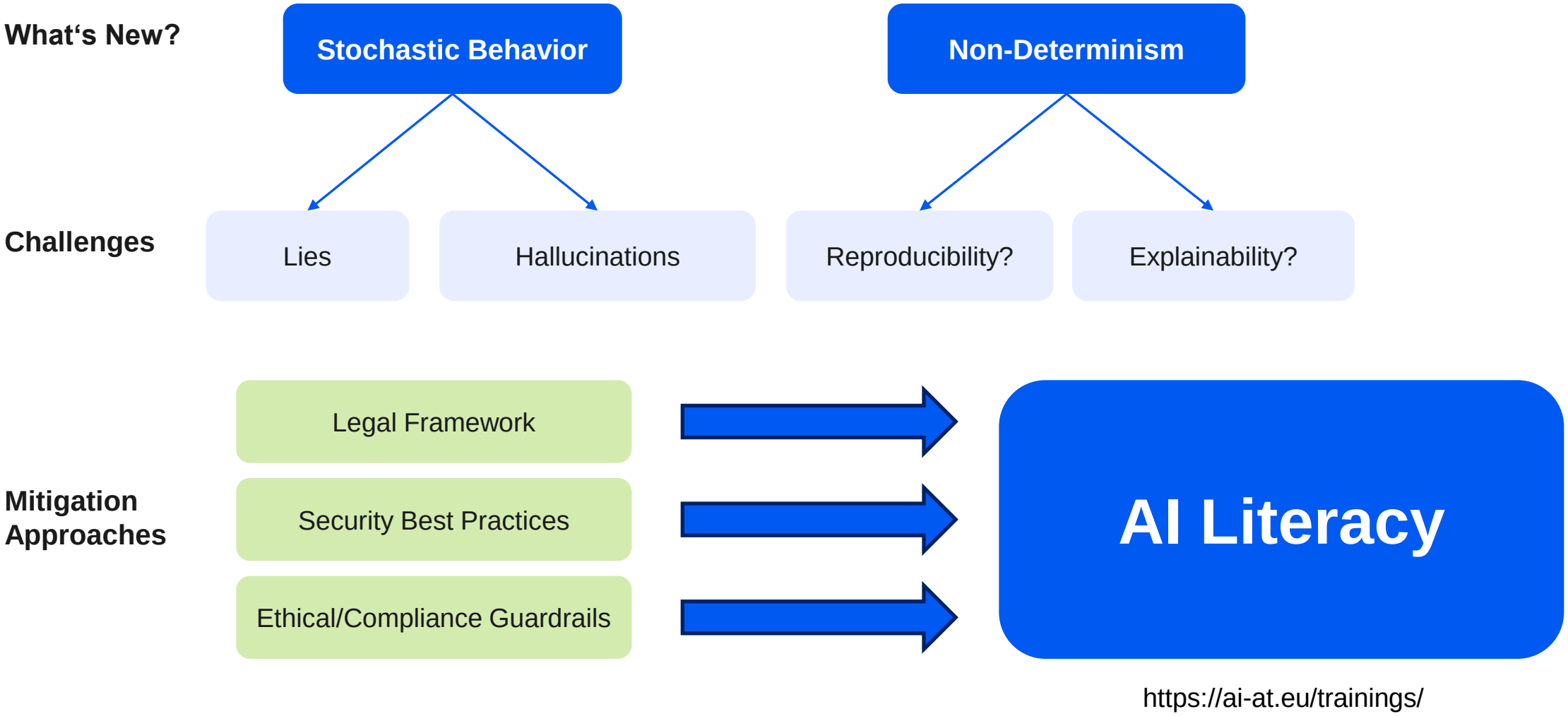



Problem: no „near“ points for given input

- LLM draws conclusions that are not there => still better than no answer
- What if intended answer differs from „most probable“ output?

https://chatgpt.com/; https://mistral.ai/; https://www.anthropic.com/; https://gemini.google.com/

# Generative AI – What‘s New?




https://ai-at.eu/trainings/

# Working with LLMs





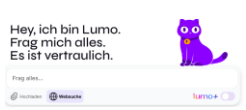
# Common AI Chatbots

Different Providers, different Philosophies

ChatGPT

Gemini

Claude






Different LLMs from each provider

- Short answers (cheaper)
- Complex answers (expensive)
- Generate Images
- Generate Videos
- Summarize Documents

Different tools building upon these LLMs

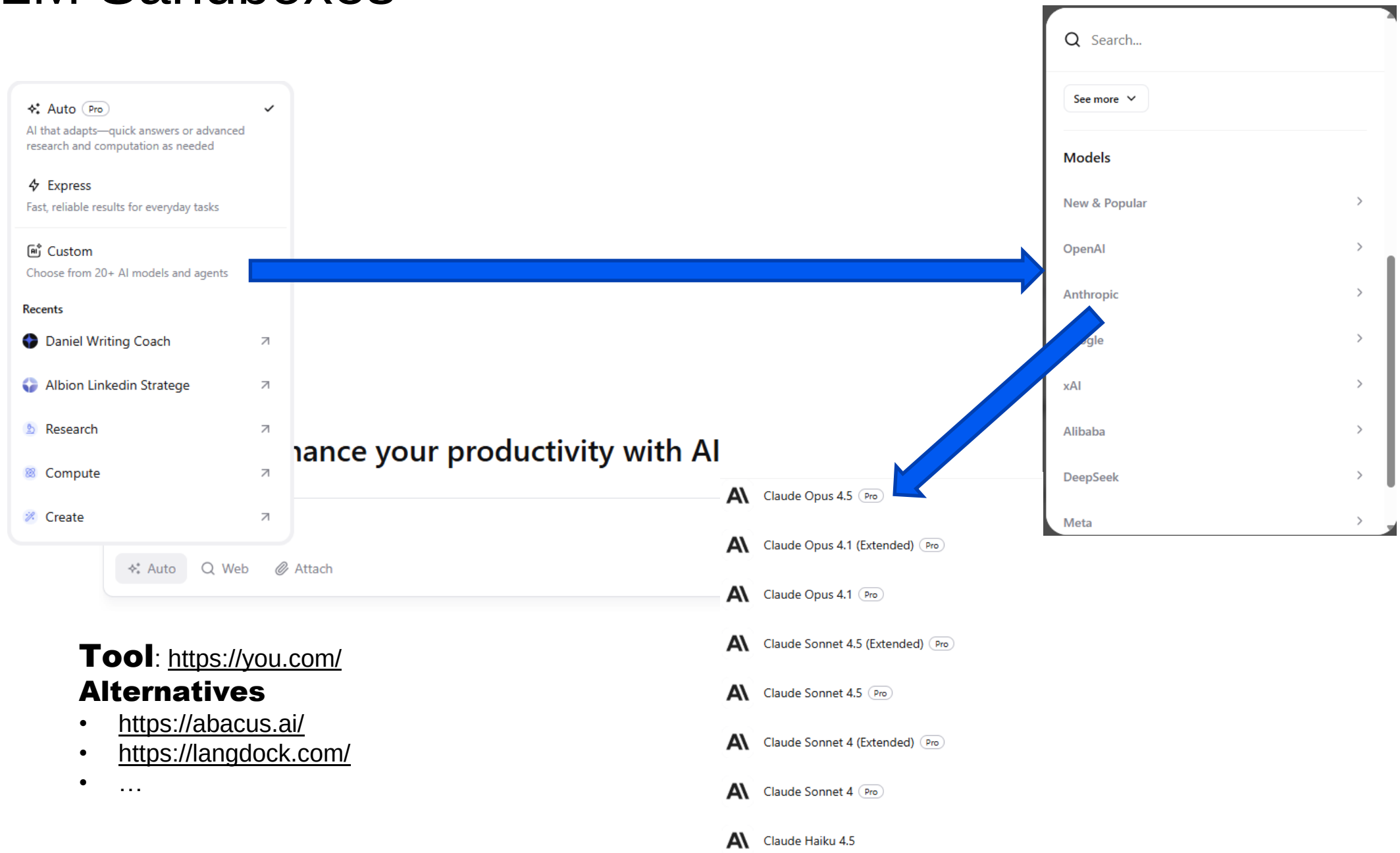
NotebookLM

fonio.ai



https://chatgpt.com/; https://mistral.ai/; https://www.anthropic.com/; https://gemini.google.com/; https://lumo.proton.me/;

# LLM Sandboxes

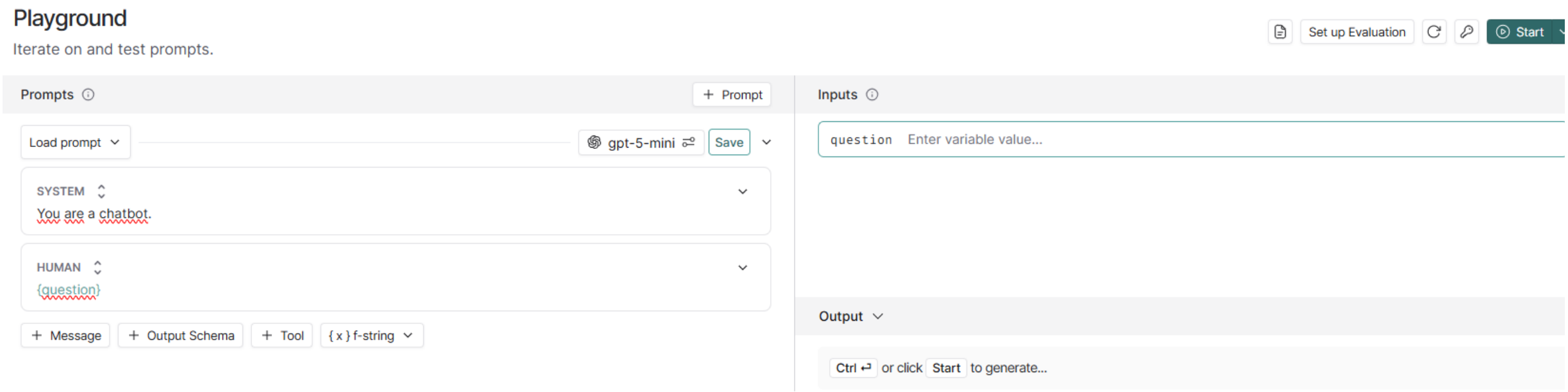



**Tool**: https://you.com/

**Alternatives**

- https://abacus.ai/
- https://langdock.com/
- …

# LLM Sandboxes for Prompt Optimization




https://smith.langchain.com/o/cd3eb3e5-dac5-46b4-840b-4846c284f7c0/playground

**Alternativs**:
https://arize.com/phoenix/
https://abacus.ai/

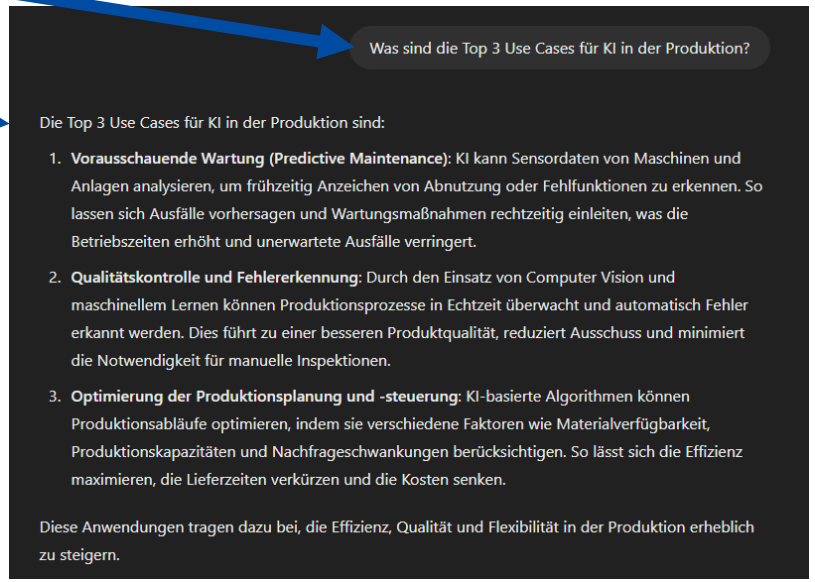
# Use LLMs within software

```
from openai import OpenAI

client = OpenAI(api_key=„…“
response = client.chat.completions.create(
    model="gpt-4o-mini",
    max_tokens=100,
    messages=[{"role": "user", "content": "IYOUR PROMT HERE"}]
)
print(response.choices[0].message.content)
```

```
pip install openai
```

Get it from: https://platform.openai.com/




https://campus.datacamp.com/courses/working-with-the-openai-api/introduction-to-the-openai-api

# Use Cases involving „LLM APIs“

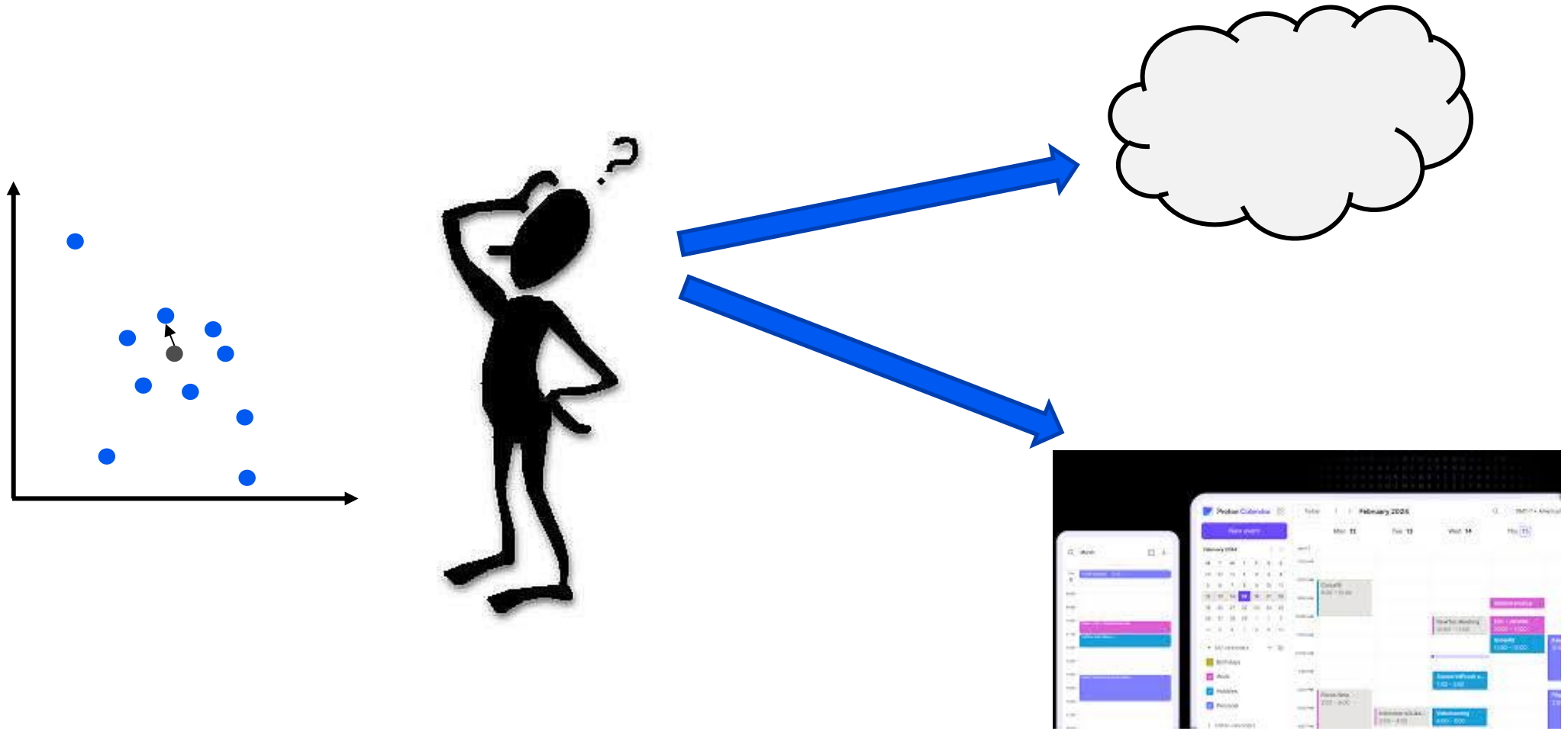
Add LLM responses to your own software (e.g., chatbots)

Improving/completing prompts („System Prompts“)

Enrich prompts with your own data („RAG“)

# Extending LLM Knowledge

# Extending LLM Knowledge



# Retrieval Augmented Generation (RAG)

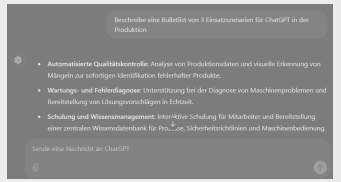




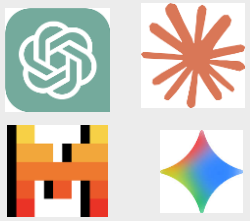
Individual Data

Large Pile of Data

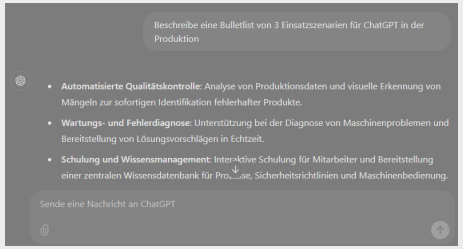


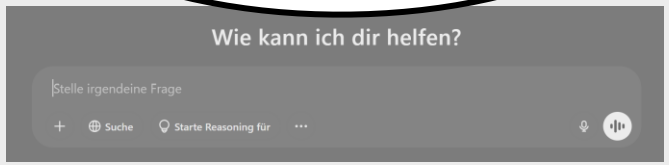



Most probable Output for given Input?



For your answer, also consider the following individual data from the following source: folgender Quelle: …



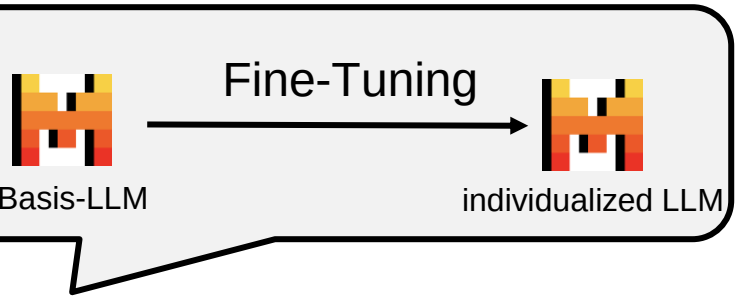

OpenWebUI/ ChatCPT.com/ …





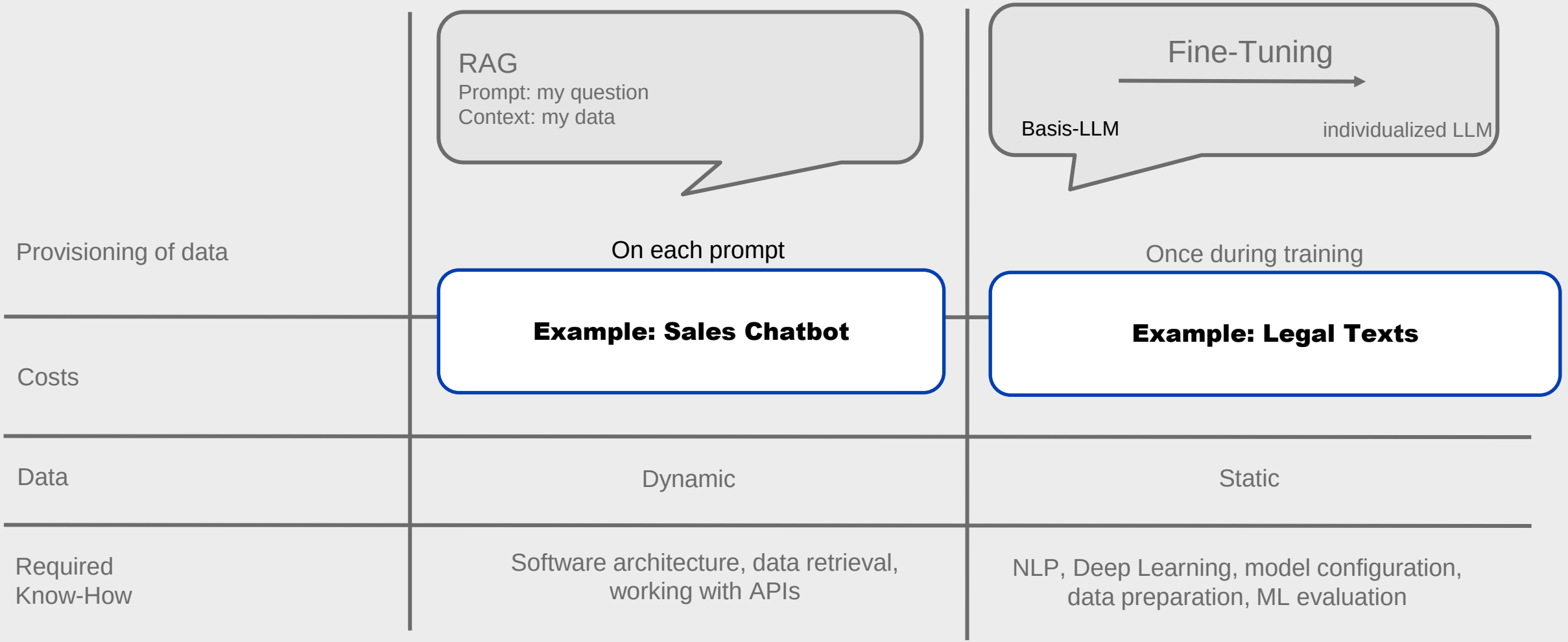
https://chatgpt.com/; https://mistral.ai/; https://www.anthropic.com/; https://gemini.google.com/

# RAG VS Fine-Tuning

| | RAG<br>Prompt: my question<br>Context: my data |  |
|---|---|---|
| Provisioning of data | On each prompt | Once during training |
| Costs | Small overhead during operations<br>Dedicated software architecure necessary | Special hardware necessary |
| Data | Dynamic | Static |
| Required Know-How | Software architecture, data retrieval, working with APIs | NLP, Deep Learning, model configuration, data preparation, ML evaluation |

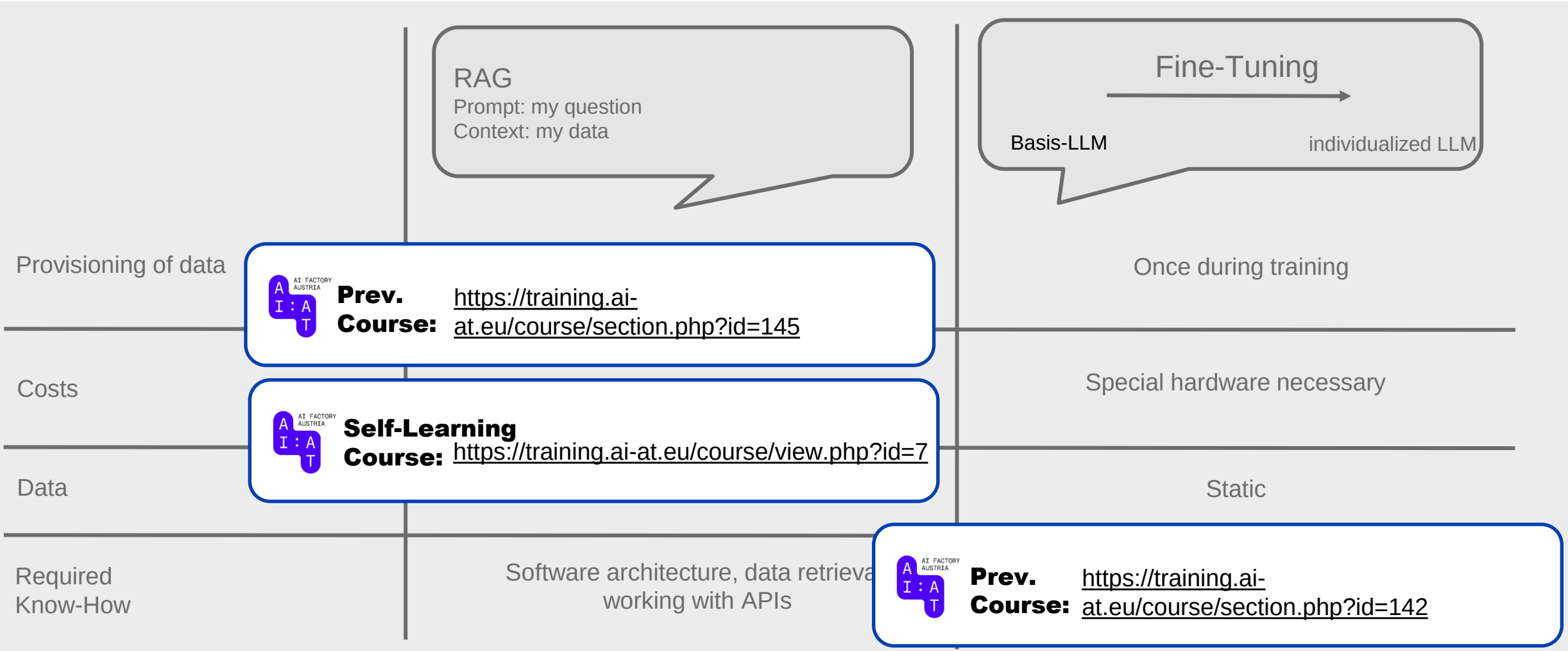
https://www.redhat.com/de/topics/ai/rag-vs-fine-tuning

# RAG VS Fine-Tuning




https://www.redhat.com/de/topics/ai/rag-vs-fine-tuning

# RAG VS Fine-Tuning

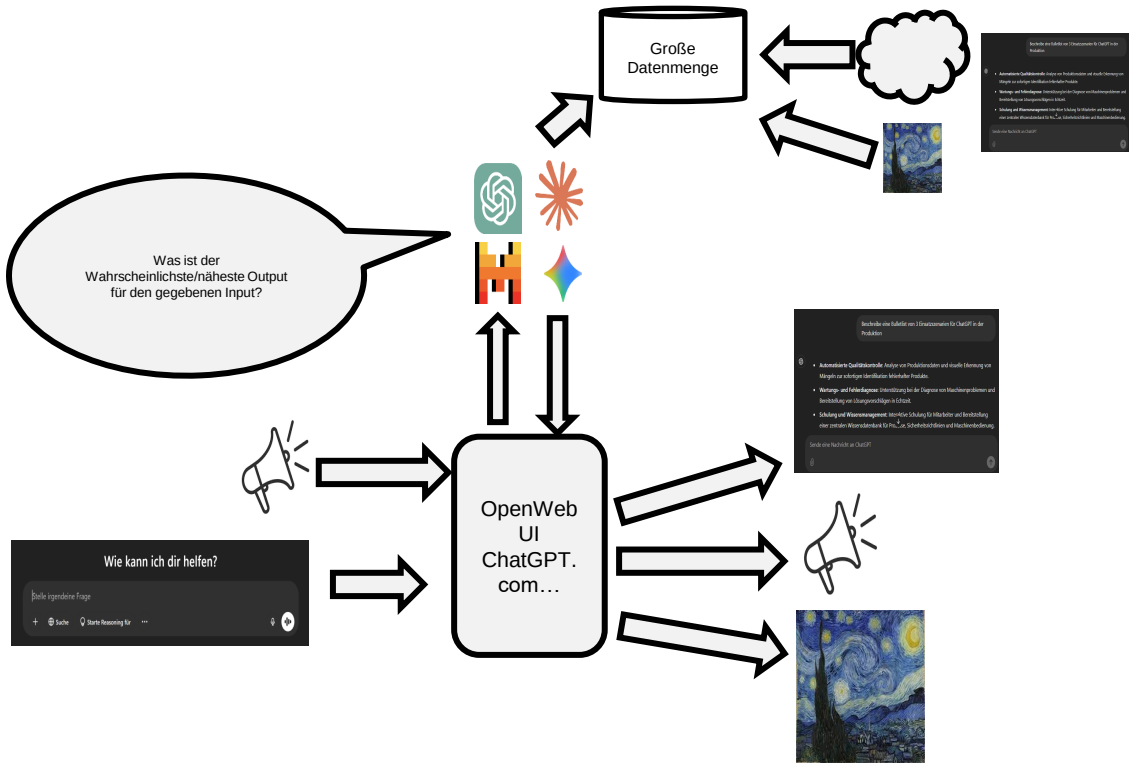



https://www.redhat.com/de/topics/ai/rag-vs-fine-tuning

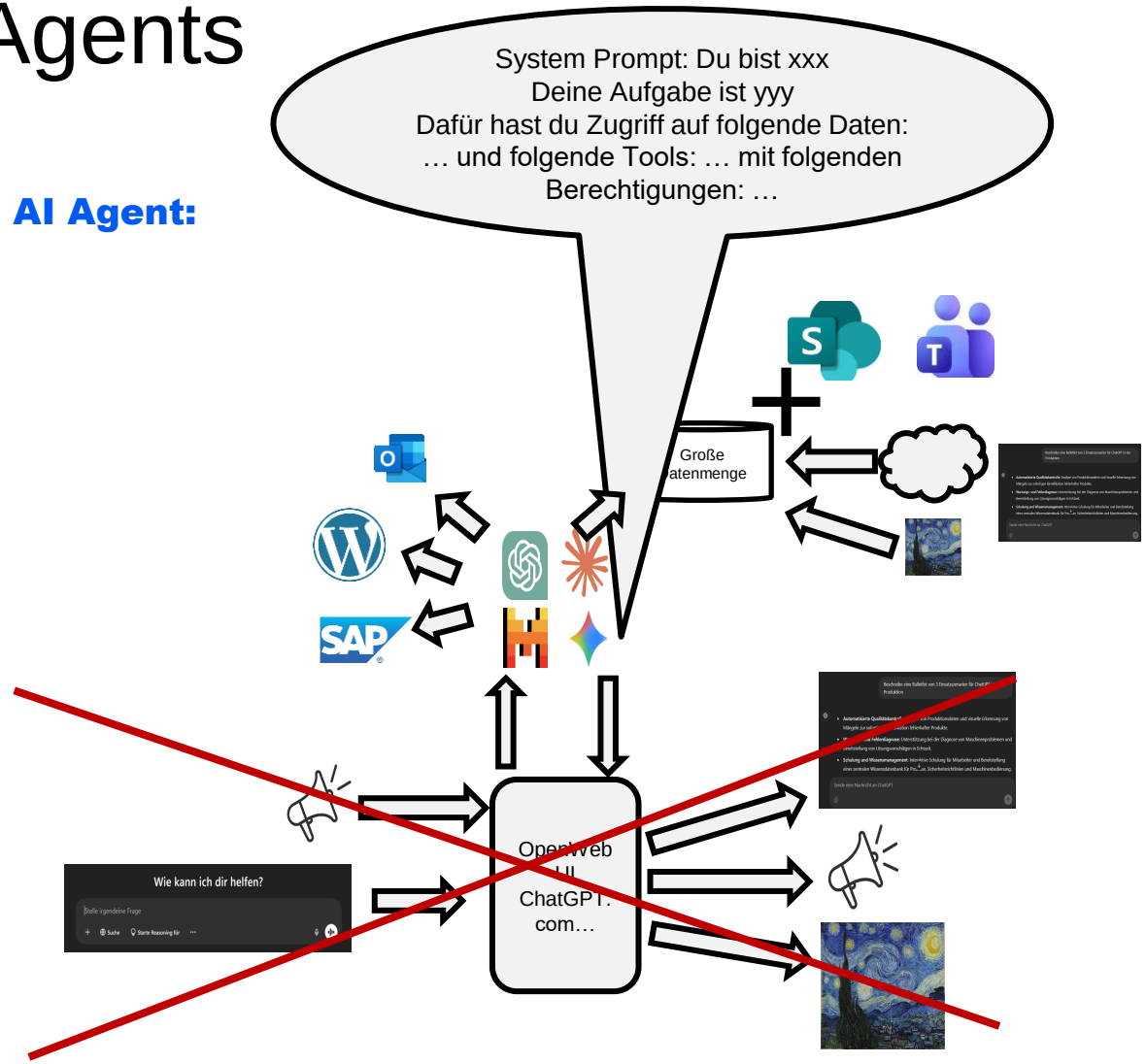
# AI Agents

# Difference AI Chatbots and AI Agents

## AI Chatbot:



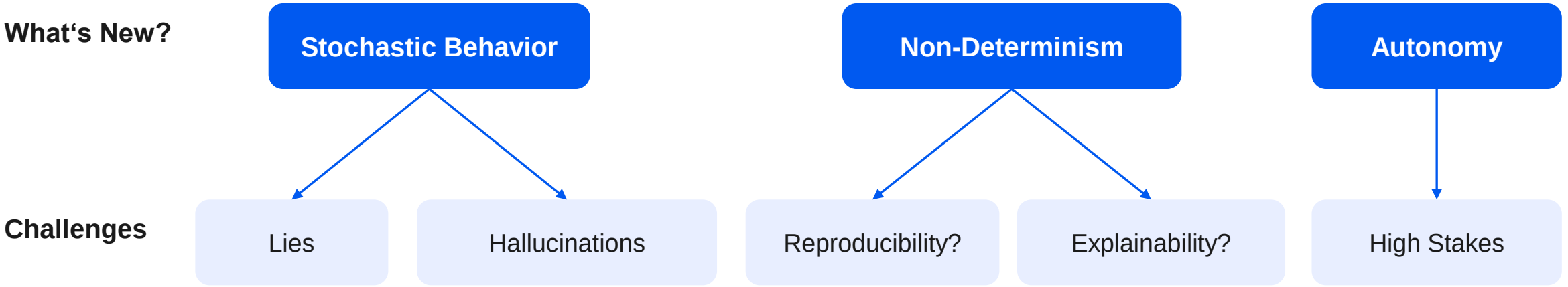

## AI Agent:

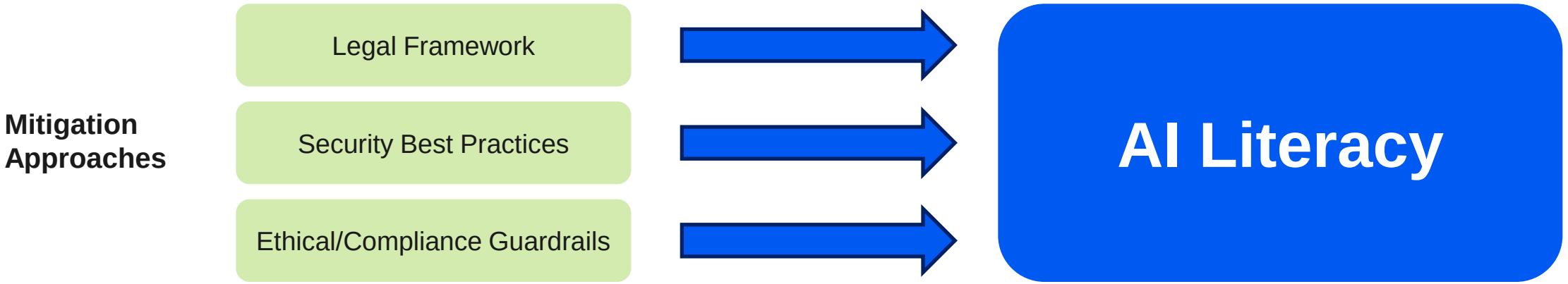





https://chatgpt.com/; https://mistral.ai/; https://www.anthropic.com/; https://gemini.google.com/

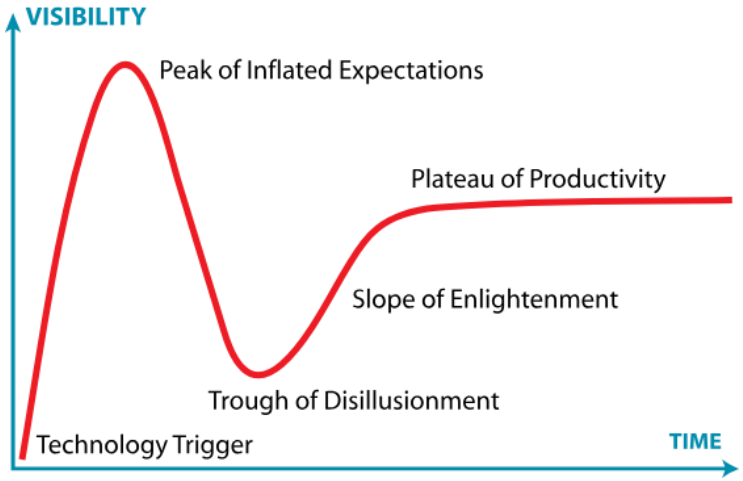
# AI Agents – What‘s New?




https://ai-at.eu/trainings/

# Summary

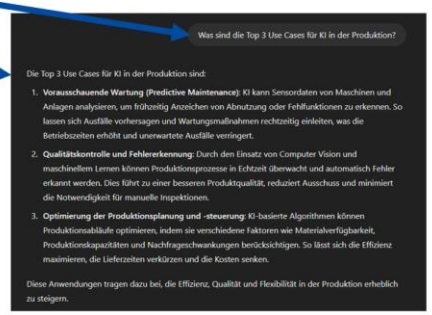



```
from openai import OpenAI

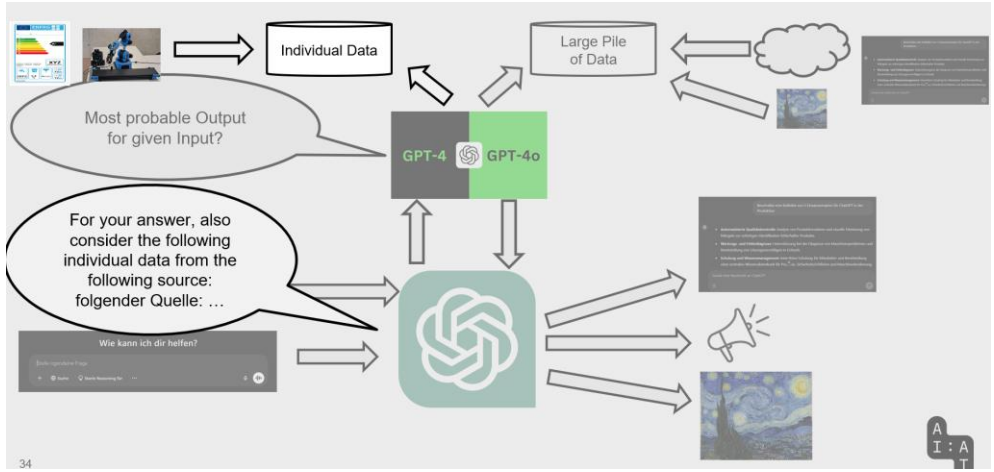
client = OpenAI(api_key="...")
response = client.chat.completions.create(
  model="gpt-4o-mini",
  max_tokens=100,
  messages=[{"role": "user", "content": "IYOUR PROMT HERE"}]
)
print(response.choices[0].message.content)
```

pip install openai

Get it from: https://platform.openai.com/




https://campus.datacamp.com/courses/working-with-the-openai-api/introduction-to-the-openai-api




 https://chatgpt.com/; https://mistral.ai/; https://www.anthropic.com/; https://gemini.google.com/; https://en.wikipedia.org/wiki/Gartner_hype_cycle

# Contact


## Dr. Daniel Lehner

Expert AI Knowledge Transfer
AI Factory Austria AI:AT

daniel.lehner@ai-at.eu

AI Factory Austria AI:AT
Schwarzenbergplatz 2
1010 Wien, Austria

training@ai-at.eu
info@ai-at.eu

ai-at.eu

@ai-factory-austria

# Funded by

under discussion with

AI Factory Austria AI:AT has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101253078. The JU receives support from the Horizon Europe Programm of the European Union and Austria (BMIMI / FFG).