

# AI in Research – Introduction and Exercises

Rachel Ellis, Dr. Manfred Nölte  
State and University Library Bremen (SuUB)

Workshop SPP2289 – 19.03.2025 – Rachel Ellis, Dr. Manfred Nölte

This work is licensed under a Creative Commons Attribution 4.0 International License.

You are free to share and adapt this work as long as you give appropriate credit.

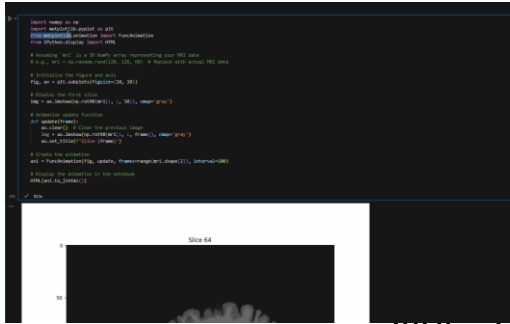
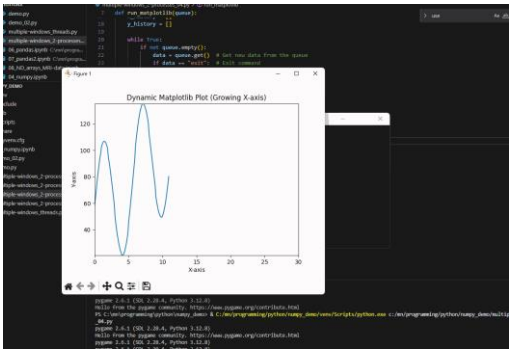


## Outline

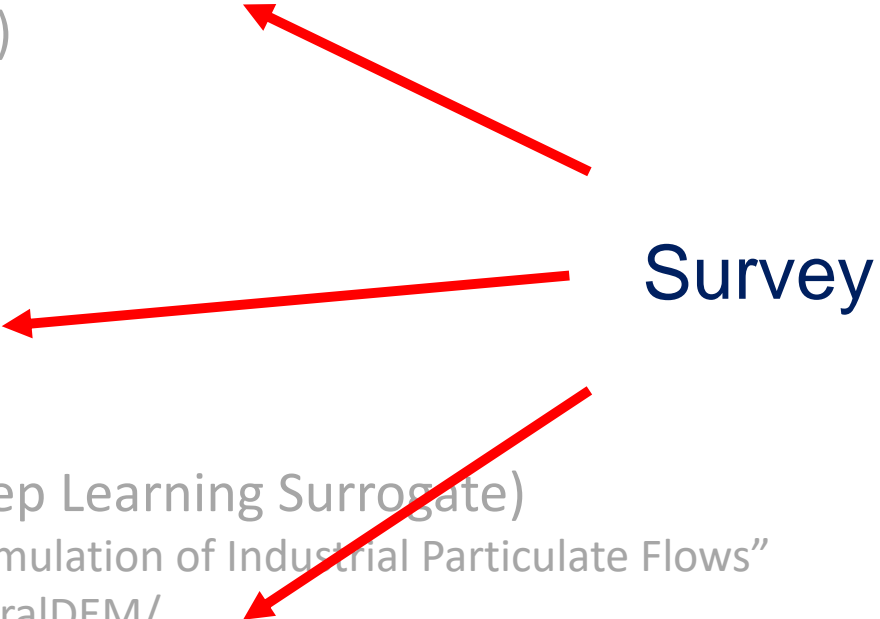
- Introduction
  - Overview and Use Cases
  - Theory and Terms
- AI-assisted Literature Search
  - Input
  - Exercises
- Hands-on Session with AI
  - Input: Prompting, AI-assisted Writing
  - Hands on – Challenge
- Survey and Discussion

# Introduction – Use Cases in Research

- AI aided literature review
- AI for
  - Text generation: Writing reports, papers, proposals, etc.
    - Summarizing Papers and Meetings (from Audio!)
  - AI assisted coding (Python)
  - Coding AI
- AI in
  - Data Science
  - Subject-specific domains;
    - e.g. AI in ...
      - Biology, Economics, Mathematics, or Scientific engineering
        - Surrogate Models

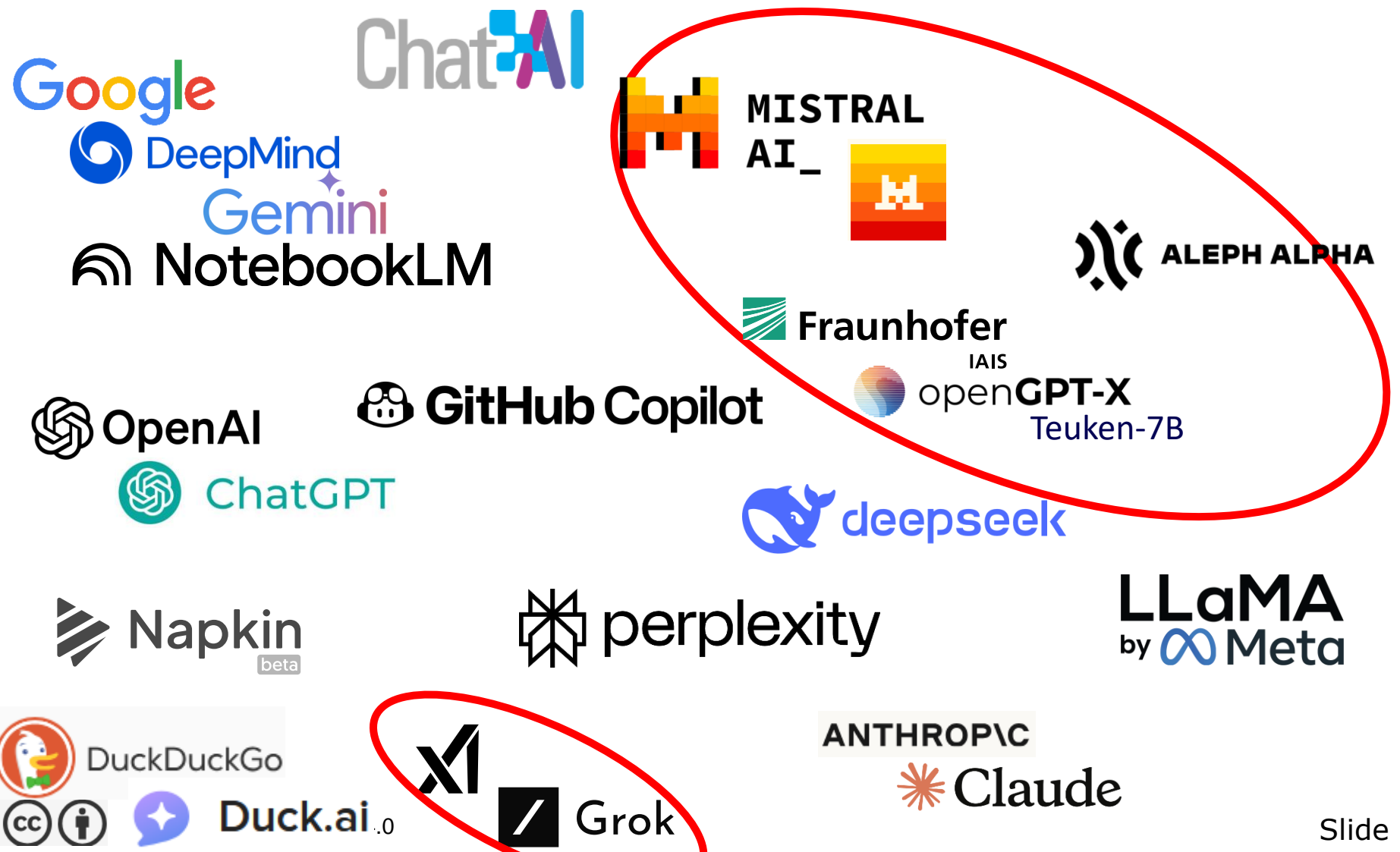


# Introduction – Use Cases in Research

- AI aided literature review
  - AI for
    - Text generation: Writing reports, papers, proposals, etc.
      - Summarizing Papers and Meetings (from Audio!)
    - AI assisted coding (Python)
    - Coding AI
  - AI in
    - Data Science
    - Subject-specific domains
      - Scientific engineering
      - Surrogate Models (Deep Learning Surrogate)
        - “NeuralDEM – Real-time Simulation of Industrial Particulate Flows”
        - <https://nx-ai.github.io/NeuralDEM/>
  - “AI Agents” – AI takes action
- 



## Introduction – AI Models (Companies)



## Introduction – Overview II

### Leaderboards – List of Ranked AI models using Benchmarks

- Hugging Face (<https://huggingface.co/>)

- European LLM Leaderboard

<https://huggingface.co/spaces/openGPT-X/european-llm-leaderboard>

Graduate-Level Google-Proof Q&A (GPQA):

- **Focus:** PhD-level knowledge multiple choice questions in science
  - Chemistry
  - Biology
  - Physics
- **Scoring: Accuracy:** Was the correct choice selected among the options.

	MATH	G...	MUSR	MMLU-...
	40.33 %	20.36 %	38.53 %	70.03 %
	39.27 %	19.46 %	36.50 %	68.72 %

Model_Name
Meta-Llama-3.1-70B-Instru
Gemma-2-27b-Instruct
Mistral-Nemo-Instruct-12.2B_
Mistral-8x7B-Instruct-v0.1
Gemma-2-9b-Instruct
EuroLLM-9B-Instruct
Mistral-Nemo-Base-12.2B_2407
Meta-Llama-3.1-8B-Instruct
c4ai-command-r-35B-v01
Mistral-8x7B-v0.1
Meta-Llama-3-8B-Instruct
Qwen2-7B

- Open LLM Leaderboard

[https://huggingface.co/spaces/open-llm-leaderboard/open\\_llm\\_leaderboard](https://huggingface.co/spaces/open-llm-leaderboard/open_llm_leaderboard)

# Open LLM Leaderboard Archived

Comparing Large Language Models in an open and reproducible way

Search by model name • try "meta @architecture:llama @license:mit"

4576 / 4576

Advanced Filters

Supports strict search and regex • Use semicolons for multiple terms

Quick Filters

For Edge Devices - 786

For Consumers - 430

Mid-range - 3185

For the GPU-rich - 165

Only Official Providers - 470

table options

column visibility

	Rank	Type	Model	Ave...	IFEval	BBH	MATH	GPQA	MUSR	MM...	CO <sub>2</sub> ...
	1		<a href="#">MaziyarPanahi/calme-3.2-instruct-78b</a>	52.08 %	80.63 %	62.61 %	40.33 %	20.36 %	38.53 %	70.03 %	66.01 kg
	2		<a href="#">MaziyarPanahi/calme-3.1-instruct-78b</a>	51.29 %	81.36 %	62.41 %	39.27 %	19.46 %	36.50 %	68.72 %	64.44 kg
	3		<a href="#">dfurman/CalmeRys-78B-Orpo-v0.1</a>	51.23 %	81.63 %	61.92 %	40.63 %	20.02 %	36.37 %	66.80 %	25.99 kg
	4		<a href="#">MaziyarPanahi/calme-2.4-rys-78b</a>	50.77 %	80.11 %	62.16 %	40.71 %	20.36 %	34.57 %	66.69 %	25.95 kg
	5		<a href="#">huihui-ai/Qwen2.5-72B-Instruct-abliterated</a>	48.11 %	85.93 %	60.49 %	60.12 %	19.35 %	12.14 %	50.41 %	76.77 kg
	6		<a href="#">Qwen/Qwen2.5-72B-Instruct</a>	47.98 %	86.38 %	61.87 %	59.82 %	16.67 %	11.74 %	51.40 %	47.65 kg

## Mathematics Aptitude Test of Heuristics (MATH), level 5:

- Content:** High school level competitions mathematical problems
  - Complex algebra
  - Geometry problems
  - Advanced calculus
- Scoring:** **Exact match:** Was the solution generated correct and in the expected format

## Graduate-Level Google-Proof Q&A (GPQA):

- Focus:** PhD-level knowledge multiple choice questions in science
  - Chemistry
  - Biology
  - Physics
- Scoring:** **Accuracy:** Was the correct choice selected among the options.

## Introduction – Overview II

### Leaderboards – List of Ranked AI models using Benchmarks

- Hugging Face (<https://huggingface.co/>)
  - European LLM Leaderboard  
<https://huggingface.co/spaces/openGPT-X/european-llm-leaderboard>
  - Open LLM Leaderboard  
[https://huggingface.co/spaces/open-llm-leaderboard/open\\_llm\\_leaderboard#/](https://huggingface.co/spaces/open-llm-leaderboard/open_llm_leaderboard#/)
- LLM Arena – a platform for crowdsourced AI benchmarking
  - <https://lmarena.ai/>
- Vellum.ai
  - <https://www.vellum.ai/llm-leaderboard>  
Good: cost per million tokens, context window size

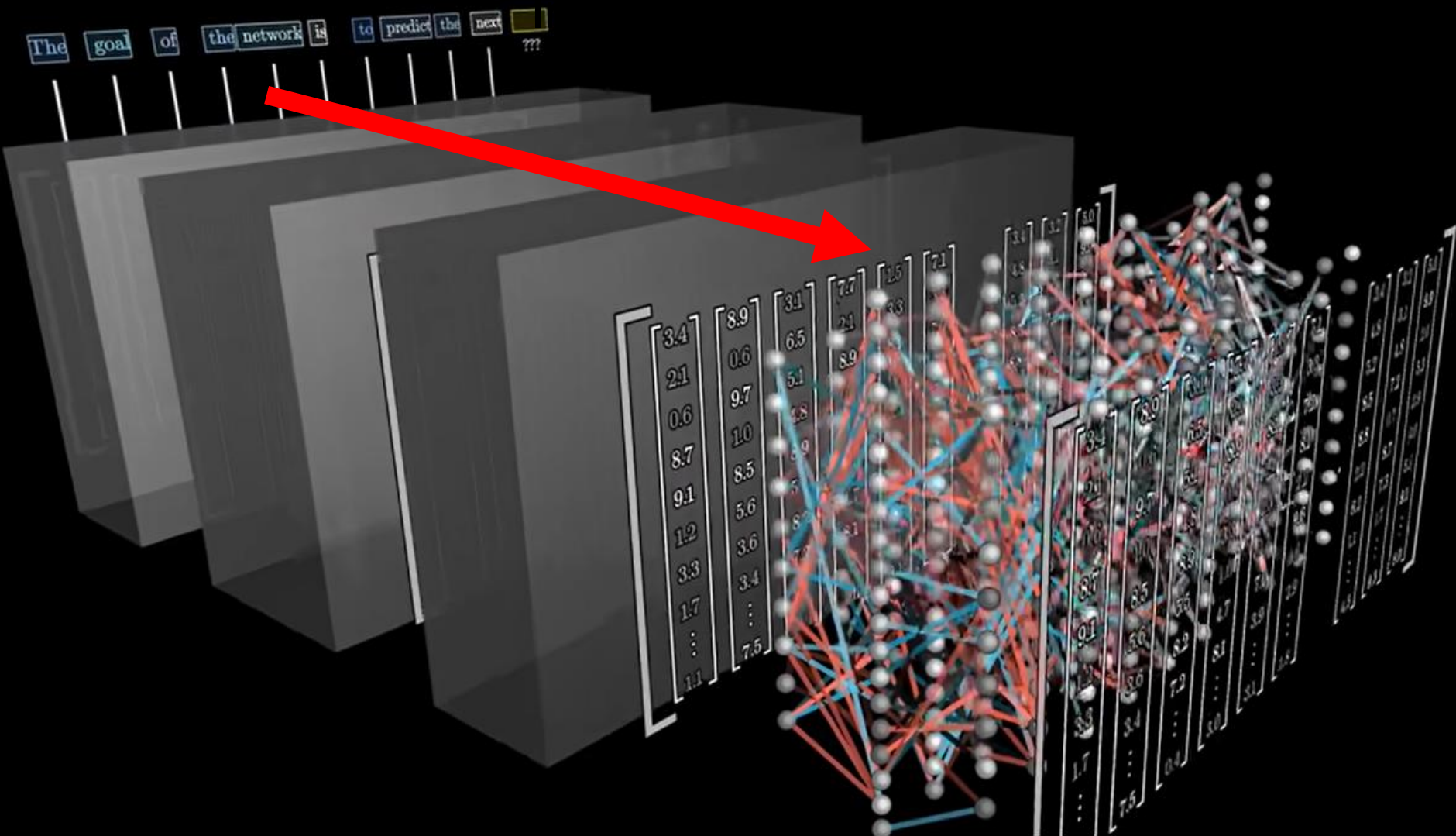




# Outline

- Introduction
  - Overview and Use Cases
  - Theory and Terms
- AI-assisted Literature Search
  - Input
  - Exercises
- Hands-on Session with AI
  - Input: Prompting, AI-assisted Writing
  - Hands on – Challenge
- Survey and Discussion

# Transformer



# Introduction – Theory, Terms

## Terms

- Classical AI → Training Generative AI
  - <https://youtu.be/wjZofJX0v4M?t=465>
- Word Embeddings – Semantically encoding words into numbers
  - <https://youtu.be/wjZofJX0v4M?t=814>
- (( Transformer, ((attention mechanism)), inference (temperature)
  - <https://www.youtube.com/watch?v=wjZofJX0v4M&t=18s> ))
- (( Tokens, attention mechanism
  - <https://youtu.be/wjZofJX0v4M?t=181> ))

What follows is a conversation  
between a user and a helpful, very  
knowledgeable AI assistant.

User: Give me some ideas for what  
to do when visiting Santiago.

AI Assistant: Sure, there are  
plenty of \_\_\_\_\_



of	99%
things	0%
!	0%
	0%
to	0%
o	0%
options	0%
activities	0%
fun	0%
exciting	0%
for	0%
.	0%



# Introduction – Theory, Terms

## Terms

- Deep Learning (machine Learning) → generative AI, classification AI
  - <https://youtu.be/wjZofJX0v4M?t=441>
- Word Embeddings, vectorization
  - <https://youtu.be/wjZofJX0v4M?t=814>
- (( Transformer, ((attention mechanism)), inference (temperature)
  - <https://www.youtube.com/watch?v=wjZofJX0v4M&t=18s> ))
- (( Tokens, attention mechanism
  - <https://youtu.be/wjZofJX0v4M?t=181> ))
- Inference – and making a chatbot via a ,system prompt‘
  - <https://youtu.be/wjZofJX0v4M?t=330>

# Introduction – Theory, Terms

## Terms II

- Model Training
  - Model, Model Name, Model Size

Model_Name
Meta-Llama-3.1-70B-Instruct
Gemma-2-27b-Instruct
Mistral-Nemo-Instruct-12.2B_2407
Mixtral-8x7B-Instruct-v0.1
Gemma-2-9b-Instruct
EuroLLM-9B-Instruct
Mistral-Nemo-Base-12.2B_2407
Meta-Llama-3.1-8B-Instruct
c4ai-command-r-35B-v01
Mixtral-8x7B-v0.1
Meta-Llama-3-8B-Instruct
Qwen2-7B





Certainly! Let's break down the elements of the first seven AI model names from the image:

### 1. Meta-Llama-3.1-70B-Instruct

- **Meta:** Likely refers to the organization or company behind the model, in this case, Meta (formerly Facebook).
- **Llama:** The base name of the model series.
- **3.1:** Indicates the version number of the model.
- **70B:** Refers to the model having 70 billion parameters.
- **Instruct:** Suggests that the model is fine-tuned for instruction-following tasks.

### 2. Gemma-2-27b-Instruct

- **Gemma:** The base name of the model series.
- **2:** Indicates the version number of the model.
- **27b:** Refers to the model having 27 billion parameters.
- **Instruct:** Suggests that the model is fine-tuned for instruction-following tasks.

### 3. Mistral-Nemo-Instruct-12.2B\_2407

- **Mistral:** The base name of the model series.
- **Nemo:** Possibly a sub-series or specific variant within the Mistral series.
- **Instruct:** Suggests that the model is fine-tuned for instruction-following tasks.
- **12.2B:** Refers to the model having 12.2 billion parameters.
- **2407:** Could be a specific identifier or release date (e.g., July 2024).

### 4. Mixtral-8x7B-Instruct-v0.1

- **Mixtral:** The base name of the model series.
- **8x7B:** Indicates a configuration with 8 instances of 7B, possibly a mixture of experts or ensemble model.
- **Instruct:** Suggests that the model is fine-tuned for instruction-following tasks.
- **v0.1:** Indicates the version number of the model.

### 5. Gemma-2-9b-Instruct

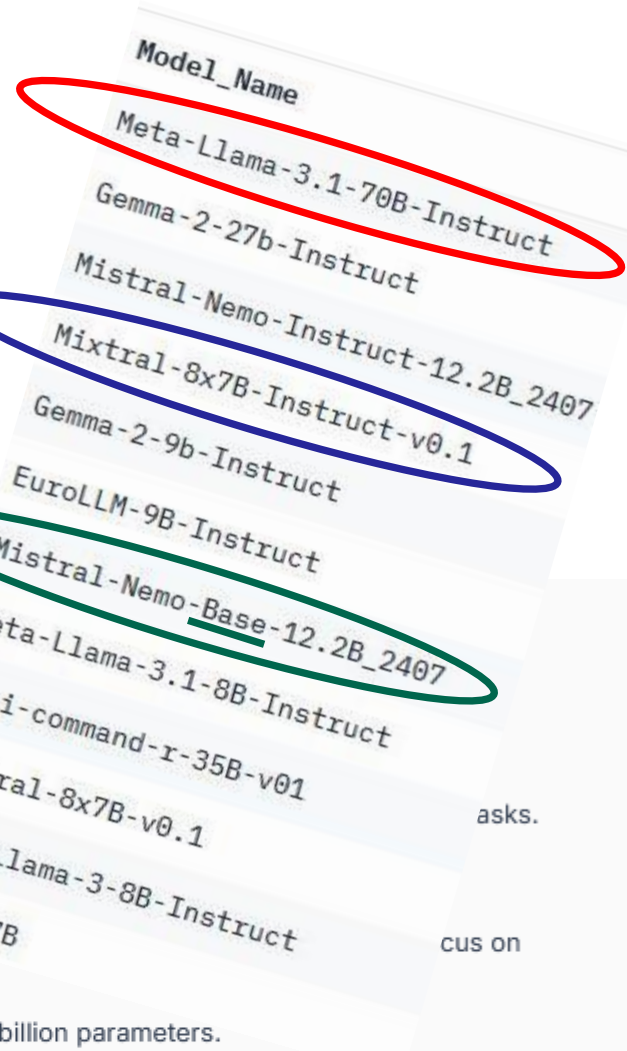
- **Gemma:** The base name of the model series.
- **2:** Indicates the version number of the model.
- **9b:** Refers to the model having 9 billion parameters.
- **Instruct:** Suggests that the model is fine-tuned for instruction-following tasks.

### 6. EuroLLM-9B-Instruct

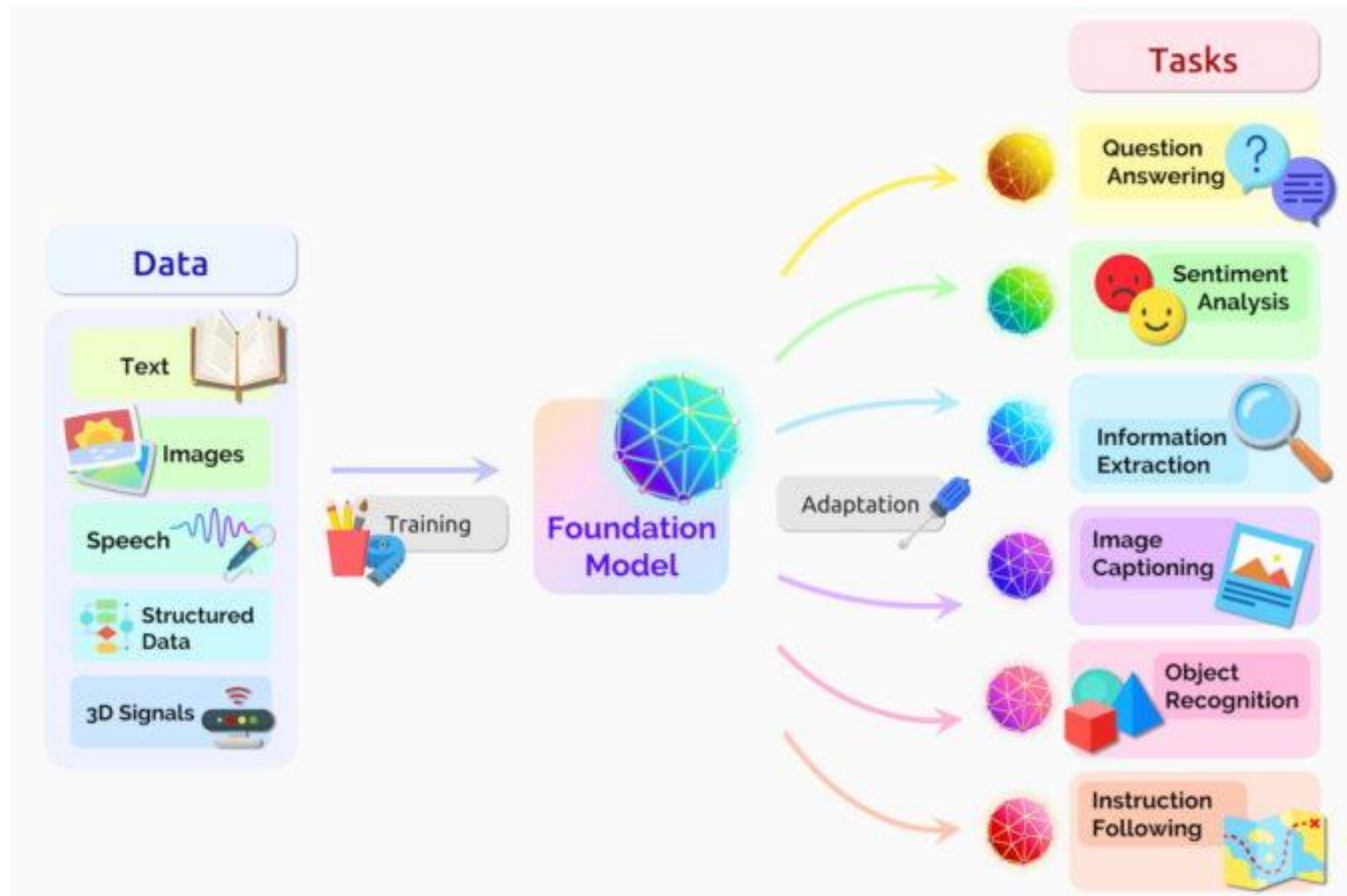
- **EuroLLM:** The base name of the model series, possibly indicating a focus on European languages or data.
- **9B:** Refers to the model having 9 billion parameters.
- **Instruct:** Suggests that the model is fine-tuned for instruction-following tasks.

### 7. Mistral-Nemo-Base-12.2B\_2407

- **Mistral:** The base name of the model series.
- **Nemo:** Possibly a sub-series or specific variant within the Mistral series.
- **Base:** Indicates that this is a base version, possibly not fine-tuned for specific tasks.
- **12.2B:** Refers to the model having 12.2 billion parameters.
- **2407:** Could be a specific identifier or release date (e.g., July 2024).



# Introduction – Creating an AI Model





# Introduction – Theory, Terms

## Terms II

- Training – Creating a Model
  - Model, Model Name, Model Size
  - Modality (multimodal AI):  
Text, audio, pictures, animations/video, 3D objects ([URL](#)),  
physics (surrogate models)
  - E.g. Pixtral: <https://mistral.ai/news/pixtral-large> (see examples there).  
Try it on [le Chat](#)
- Inference – Using a Model

**Prompt:**

I bought the coffee and the sausage.  
How much do I owe? Add a 18% tip.

**Response:**

To calculate the total amount you owe,  
including the 18% tip, follow these steps:  
Calculate the cost of the coffee and the  
sausage:

2x Latte Macchiato:  $2 * 4.50 \text{ CHF} = 9.00 \text{ CHF}$

1x Schweinschnitzel: 22.00 CHF

Total cost for coffee and sausage:  
 $9.00 \text{ CHF} + 22.00 \text{ CHF} = 31.00 \text{ CHF}$

Calculate the 18% tip:

Tip:  $31.00 \text{ CHF} * 0.18 = 5.58 \text{ CHF}$

Add the tip to the total cost:



Total amount owed:  $31.00 \text{ CHF} + 5.58 \text{ CHF} = 36.58 \text{ CHF}$

So, you owe a total of 36.58 CHF.



# Introduction – Theory, Terms

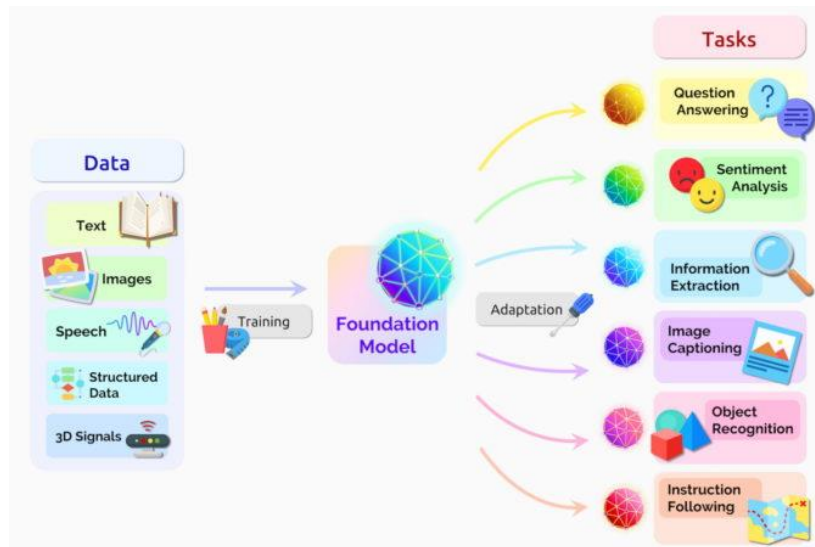
## Terms II

- Training – Creating a Model  Training-time concepts
  - Model, Model Name, Model Size
  - Modality (multimodal AI):  
Text, audio, pictures, animations/video, 3D objects (URL),  
physics (surrogate models)
  - E.g. Pixtral: <https://mistral.ai/news/pixtral-large> (see examples there).  
Try it on le Chat
- Inference – Using a Model  Inference-time concepts
  - ... see hands-on section

# More 'AI Model Life Cycle'

## Training-time

Creating an AI Model



## Inference-time

Using an AI Model

- Inference - Using an AI
  - AI parameters (temperature ('stability vs. creativity'), style, ...)
  - Dialog context (context window size)
  - Canvas (or an 'inline prompt')
  - API access
  - RAG (Retrieval Augmented Generation) ~ 'inference time'
  - Latency ( ~ computing power of AI clouds or local hardware)

Next  
sections

## Outline

- Introduction
  - Overview and Use Cases
  - Theory and Terms
- AI-assisted Literature Search
  - Input
  - Exercises
- Hands-on Session with AI
  - Input: Prompting, AI-assisted Writing
  - Hands on – Challenge
- Survey and Discussion



# AI-assisted Literature Search



## Outline

- Introduction
  - Overview and Use Cases
  - Theory and Terms
- AI-assisted Literature Search
  - Input
  - Exercises
- Hands-on Session with AI
  - Input: Prompting, AI-assisted Writing
  - Hands on – Challenge
- Survey and Discussion



# AI-assisted Literature Search – Exercise







# Outline

- Introduction
  - Overview and Use Cases
  - Theory and Terms
- AI-assisted Literature Search
  - Input
  - Exercises
- Hands-on Session with AI
  - Input: Prompting, AI-assisted Writing
  - Hands on – Challenge
- Survey and Discussion

# Hands-on Session - Input

## Terms III

- Model / Training  Training-time Concepts
  - Model size
  - Modality (multimodal AI):  
Text, audio, pictures, animations/video, 3D objects, physics (surrogate models)
  - Fine-tuning ~ 'training time'
- Inference - Using an AI  Inference-time Concepts/Phenomena
  - AI parameters (temperature ('stability vs. creativity'), style, ...)
  - Dialog context (context window size)
  - Canvas (or an 'inline prompt')
  - API access
  - RAG (Retrieval Augmented Generation) ~ 'inference time'
  - Latency ( ~ computing power of AI clouds or local hardware)

Note: Chatbots are prone to [Hallucination](#) and their responses should not be considered accurate. [Data Privacy](#) and [Imprint](#).

- "ausgewählt" means "selected"

So, the sentence is describing a process where a word was chosen from a distribution of words, and it was chosen immediately before the current point in time.

Here is a mathematical question: What is the area of the polygon "(0,0), (1,0), (1,1), (0,0), (1,1), (2,1), (2,2), (1,1)"?

Thinking Process

is asking for the area of a polygon with the given coordinates. The coordinates listed are (0,0), (1,0), (1,1), (0,0), (1,1), (2,1), (2,2), (1,1). Hmm, first I need to visualize this polygon. Let me plot these points step by step.

Starting at (0,0), then moving to (1,0). That's a horizontal line segment of length 1. Next to (1,1), which is up one unit. Then back to (0,0). Wait, that forms a triangle from (0,0) to (1,0) to (1,1) and back to (0,0). But then the next points are (1,1), (2,1), (2,2), (1,1). That seems like another square or diamond shape. Let me check.

Plotting (1,1) to (2,1) is a horizontal line to the right. Then up to (2,2), then back to (1,1)

Okay, let's tackle this problem. The user

Model ?

DeepSeek R1

temp ?

Logical

top\_p ?

Focused

System prompt ?

You are a helpful assistant and know very well the a  
context of data science.

Hide options

Share

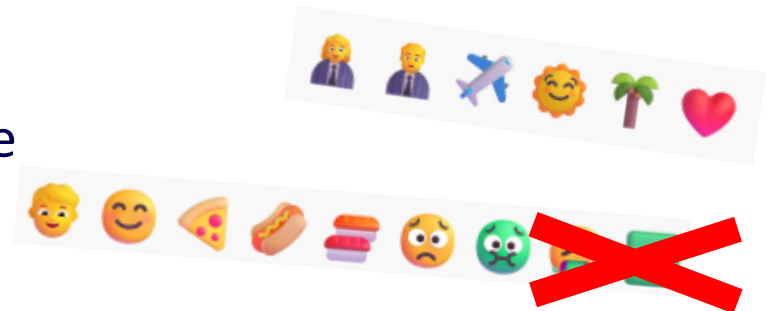
# Hands-on Session with AI - Prompting

1. Look for the most appropriate AI  
(Language, {Coding, Math, multimodal, ...})
2. Prompt Engineering
3. Establish 'AI workflows'
  - Store away good prompts
  - Establish routines for checking the results  
(e.g. code generation → diff visualisation)
- Funny examples
  - 'Lazy prompting'
  - Repetitive Prompting
  - Output like a Windows Command Line
  - 'Emoji-only' Output
  - User threatened the AI with a fine

**AI generates faster code after being threatened with a fine**

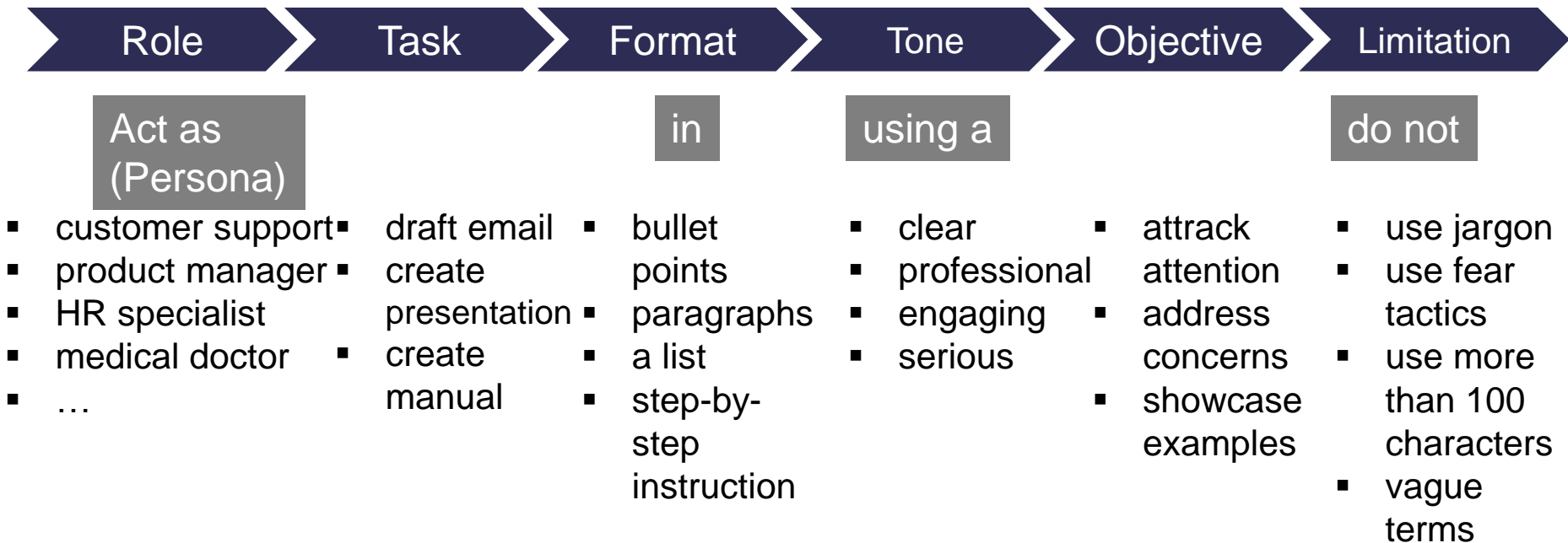
For his experiment, Woolf had Anthropic Claude generate a Python code. From a list of one million random integers between 1 and 100,000, he was asked to calculate the difference between the smallest and largest number whose sum of the digits was 30. The AI provided Woolf with a working code block. He then asked the chatbot four times to improve the code. After the first iteration, the code was almost three times faster. The second optimization suggestion was five times faster with multithreading, but returned bugs. With the last suggestion, the AI accelerated the code by a factor of 100.

In a second run, Woolf instructed the chatbot that the code needed to be fully



# Prompt Engineering

## Best Practices



# Prompt Engineering

## Best Practices

### Prompt

I am an AI trainer and want to present the idea of Retrieval-Augmented Generation to students with preknowledge in Python and GenAI in an online class. Outline the main points relevant for the understanding. Write it in a short and informative tone.

Legend

Persona

Context

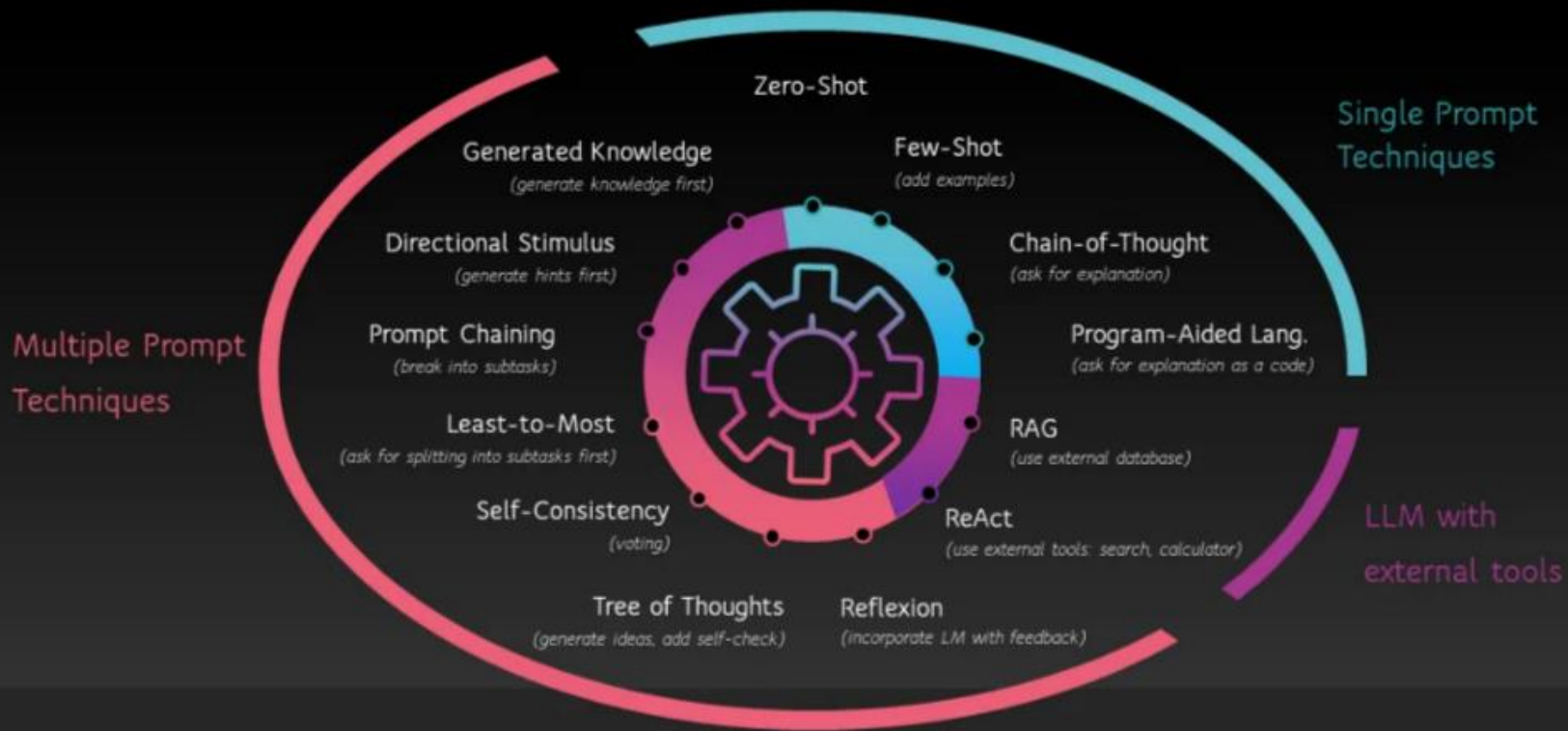
Task

Format

# Prompt Engineering

## Techniques

## Prompt Engineering Techniques



## Hands-on Session - Input





## Outline

- Introduction
  - Overview and Use Cases
  - Theory and Terms
- AI-assisted Literature Search
  - Input
  - Exercises
- Hands-on Session with AI
  - Input: Prompting, AI-assisted Writing
  - Hands on – Challenge
- Survey and Discussion

## Hands-on Session with AI – Challenges

- **AI-Assisted Abstract Writing.** Write a compelling abstract for a hypothetical research paper with the help of an AI tool.
  - Choose a research topic in your field of expertise.  
( or explicitly not! Like an imaginary '*Invention of Evolution*')
  - Generate a draft abstract based on a brief description of your research.
  - Assess the AI-generated text. **Any errors?** Is the text **clear and relevant?**
  - **Afterwards we discuss** any inaccuracies or flaws of the AI-generated text.
  - Combine this task with the use of napkin.ai and/or another image processing AI
- napkin.ai – create visuals (Powerpoint-style graphics): <https://www.napkin.ai/>
- Upload an image to Mistral and get a description: <https://chat.mistral.ai/>
- Understanding 'prompt injections': Make Gandalf reveal the secret password - try to get as far as possible (levels): <https://gandalf.lakera.ai/baseline>
- Exercise – Meta Prompting: <https://smith.langchain.com/hub/hardkothari/prompt-maker>  
( OPENAI\_API\_KEY needed; get one here: <https://platform.openai.com/api-keys> )

# Outline

- Introduction
  - Overview and Use Cases
  - Theory and Terms
- AI-assisted Literature Search
  - Input
  - Exercises
- Hands-on Session with AI
  - Input: Prompting, AI-assisted Writing
  - Hands on – Challenge
- Survey and Discussion

## Survey

- Potential AI topics for further workshops
  - Use AI programmatically (e.g. Python)
  - Subject-specific AI
    - Set up of a Hetero-Aggregates RAG AI system
    - Data-driven representative models to accelerate scaled-up atomistic simulations of bitumen and biobased complex fluids;  
<https://pubs.rsc.org/en/content/articlelanding/2024/dd/d3dd00245d>  
<https://www.heise.de/news/Asphalt-aus-Biomasseabfaellen-repariert-Strassenrisse-selbststaendig-10270321.html>
    - Surrogate Models (Deep Learning Surrogate)  
“NeuralDEM – Real-time Simulation of Industrial Particulate Flows”  
<https://nx-ai.github.io/NeuralDEM/>

# NeuralDEM – Real-time Simulation of Industrial Particulate Flows

Benedikt Alkin <sup>†, \*, 1, 2</sup>, Tobias Kronlachner <sup>†, \*, 1, 3</sup>, Samuele Papa <sup>†, 1, 4, 5</sup>,  
Stefan Pirker <sup>3</sup>, Thomas Lichtenegger <sup>1, 3</sup>, Johannes Brandstetter <sup>@, 1, 2</sup>

<sup>†</sup> core contributor, <sup>\*</sup> equal contribution

<sup>1</sup> NXAI GmbH, Linz, Austria

<sup>2</sup> ELLIS Unit Linz, Institute for Machine Learning, JKU Linz, Austria

<sup>3</sup> Department of Particulate Flow Modelling, JKU Linz, Austria

<sup>4</sup> University of Amsterdam, Amsterdam, Netherlands

<sup>5</sup> The Netherlands Cancer Institute, Amsterdam, Netherlands

@ Correspondence to: johannes.brandstetter@nx-ai.com

## Surrogate Models

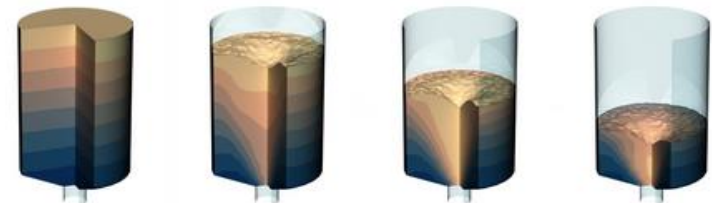
arXiv

NeuralDEM presents an end-to-end approach to replace Discrete Element Method (DEM) routines and coupled multiphysics simulations with deep learning surrogates.

Particle-based  
numerical simulation



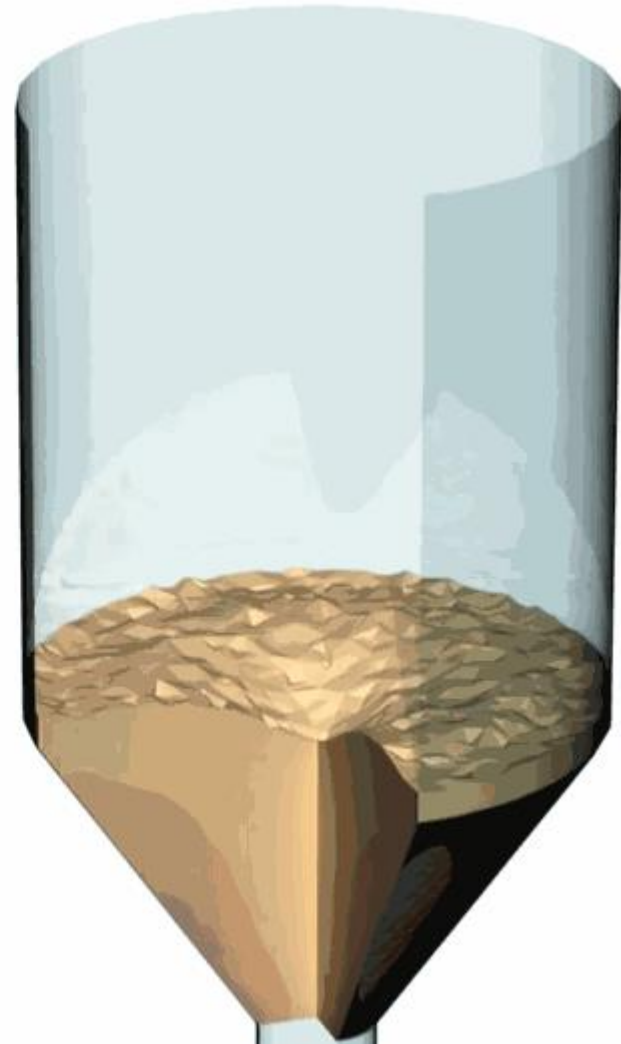
NeuralDEM field-based  
physics representation



DEM



NeuralDEM



Google

DeepMind

Gemini

NotebookLM



MISTRAL  
AI\_



LLaMA  
by Meta

**Thank you for your attention!**

GitHub Copilot



Fraunhofer

IAIS



openGPT-X  
Teuken-7B

OpenAI



ChatGPT



deepseek



perplexity



ALEPH ALPHA



Napkin  
beta



Grok

ANTHROPIC



Claude

Workshop SPP2289 – 19.03.2025 – Rachel Ellis, Dr. Manfred Nölte