

AI 4 Science Training @ UFZ Leipzig

Robert Haase

These slides can be reused under the terms of the [CC-BY 4.0](#) license unless mentioned otherwise.

Introduction

- 2010 Diplom (FH) Computer Science
- 2016 Dr. rer. medic. in Medical Image Processing
- Teaching [BiolImaging] Data Analysis and Data Management since 2019 @ TU Dresden and since 2024 @ Uni Leipzig
- Training Coordinator @ScaDS.AI since 2023



Note: One of these images was generated with Generative Artificial Intelligence.

Introduction

Center for Scalable
Data Analytics and
Artificial Intelligence



UNIVERSITÄT
LEIPZIG



TECHNISCHE
UNIVERSITÄT
DRESDEN

Missions:

Excellent Research scalable Data Science and Artificial Intelligence

Responsible, Trustworthy, Efficient

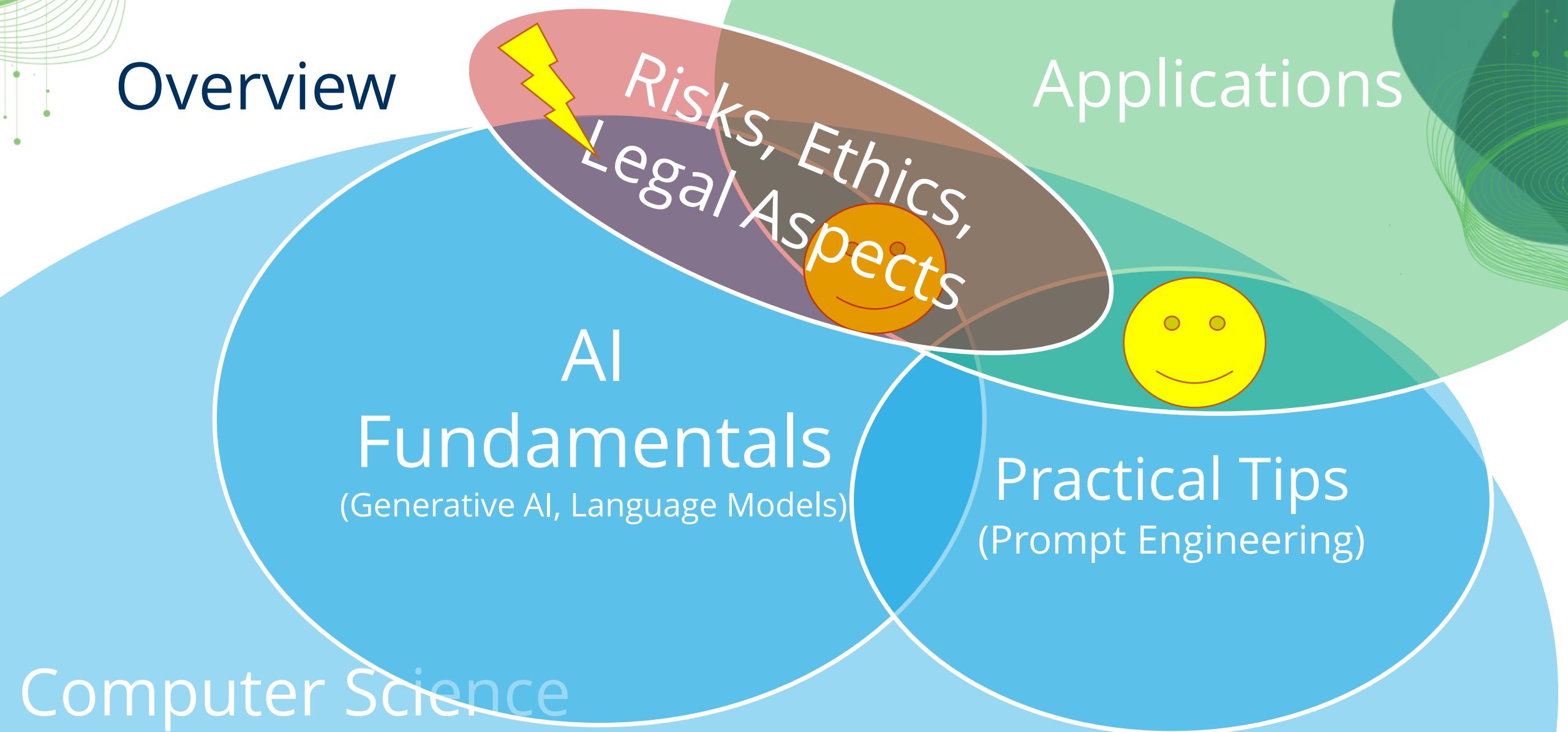
Transfer of AI-methods in applications

Training of the next generations of AI experts

Facts

- We are one of 6 AI centers in Germany
- Funded 50:50 by BMFTR and SMWK
- > 330 employees and associates in Dresden and Leipzig
- **Local partner for research in**
 - Fundamental AI research
 - AI & Data Science in real-world applications (Transfer + Service)
- **History**
 - Founded Oct. 2014 as Big Data center „ScaDS Dresden/Leipzig“
 - Since Nov. 2019: AI center ScaDS.AI Dresden / Leipzig





Quiz: Dealing with AI

How often do you use AI (ChatGPT, etc)?

< 1x per week



approx. 1x per day



Several times per day



Constantly



AI-Competency Training according to EU-AI Act

‘AI literacy’ means skills, knowledge and understanding that allow providers, deployers and affected persons, taking into account their respective rights and obligations in the context of this Regulation, to make an informed deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause; (EU AI Act, Art 3 (56))

Basics / Functionality

Proper Application

Risks / Legal Framework

Part 1

Part 1/2

Part 3

Artificial Intelligence (AI)

Weak AI

- Application-specific
- Typically trained with annotated data
- Reflexive tasks
- Extrapolation not possible

Ideal for data analysis tasks

Strong AI

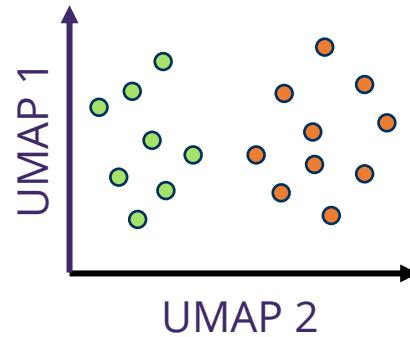
- Human capabilities
- Access to the knowledge of humanity, beyond individuals
- Can create new solutions *through creative work*

Artificial Intelligence (AI)

Explorative

Unsupervised Machine Learning (ML)

- Dimensionality reduction
- Clustering
- Outlier detection
- Hypothesis generation



Analytic

Supervised ML

- AI/ML learns tasks that need to be done by humans otherwise
- Models *are trained* using annotated data



Generative

Generative AI (*GenAI*)

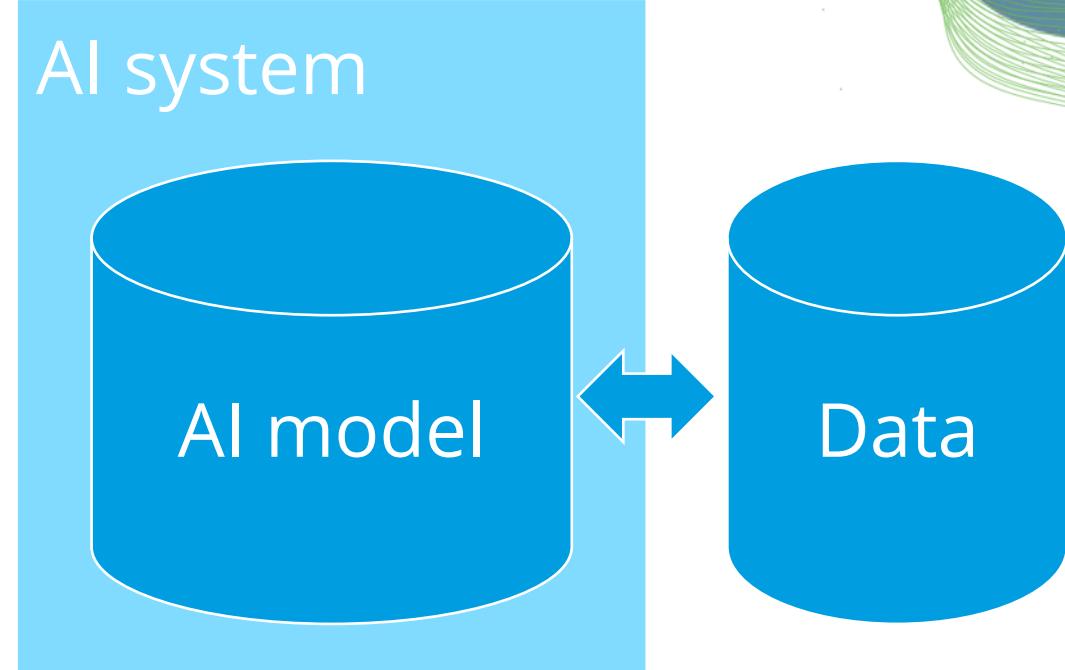
- Generates data from a prompt
- Large Language Models
- Training models is too expensive for individual projects
- Hyped since 2022, with yet unclear limitations.

Certainly!



Artificial Intelligence (AI)

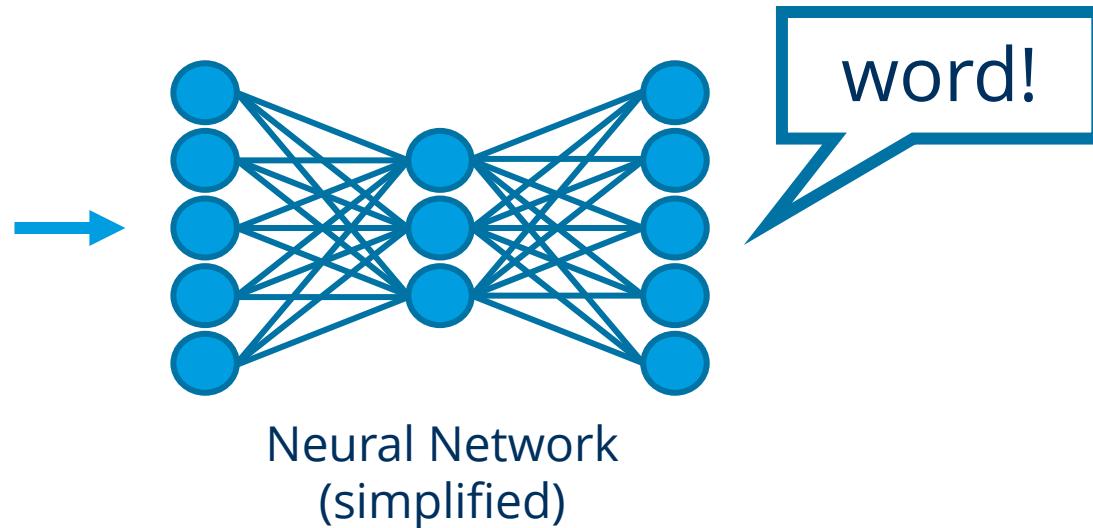
- AI models (e.g. language models):
 - Text-to-text
 - image generation, -interpretation,...
 - Examples: GPT, Gemini, llama, DeepSeek, Teuken...
- AI systems: Combination of [language] models with:
 - web-search,
 - document management,
 - databases, ...
 - Examples: ChatGPT, Perplexity, You.com



Large Language Models (LLMs)

- Advanced Architecture or Deep Neural Networks
- Statistical model for next-word prediction and translation tasks.

Predicts the next ...

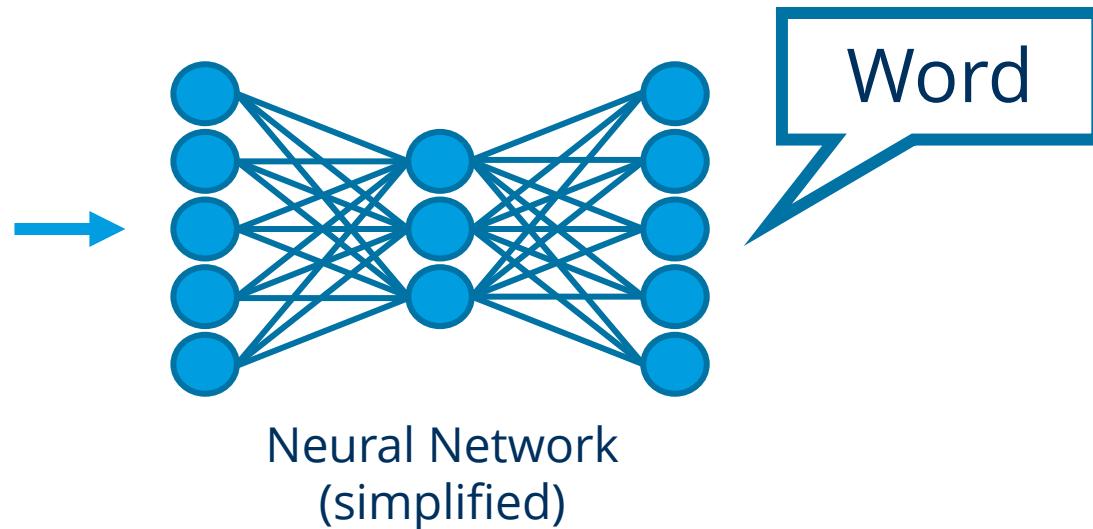


Large Language Models (LLMs)

- Advanced Architecture or Deep Neural Networks
- Statistical model for next-word prediction and translation tasks.

How does it answer
Questions?

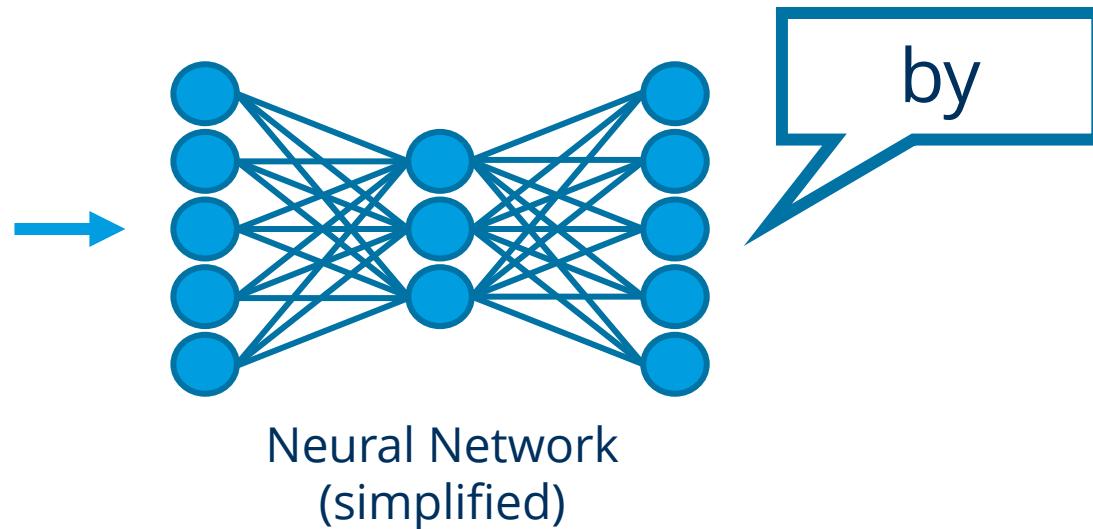
...



Large Language Models (LLMs)

- Advanced Architecture or Deep Neural Networks
- Statistical model for next-word prediction and translation tasks.

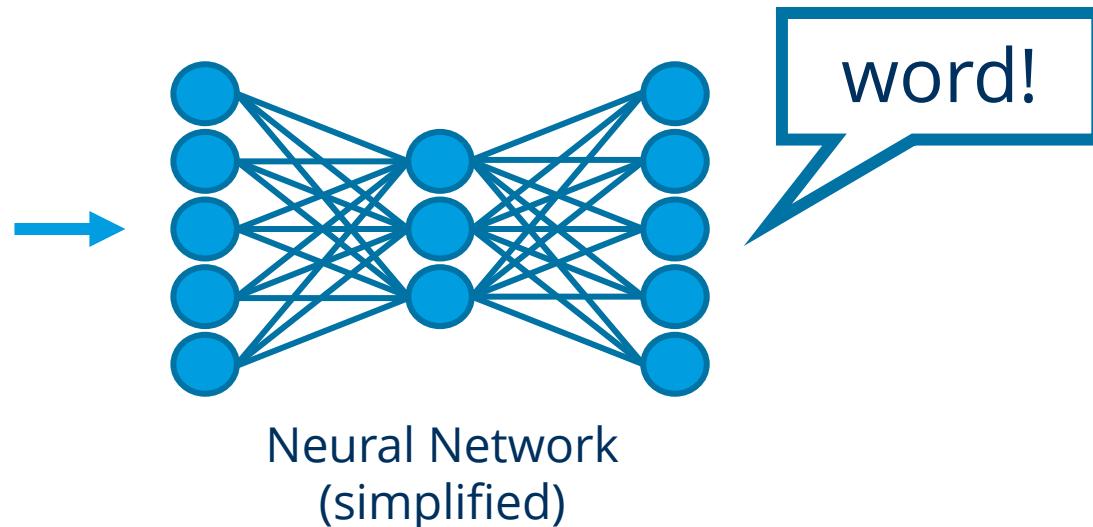
How does it answer
Questions?
Word ...



Large Language Models (LLMs)

- Advanced Architecture or Deep Neural Networks
- Statistical model for next-word prediction and translation tasks.

How does it answer
Questions?
Word by ...

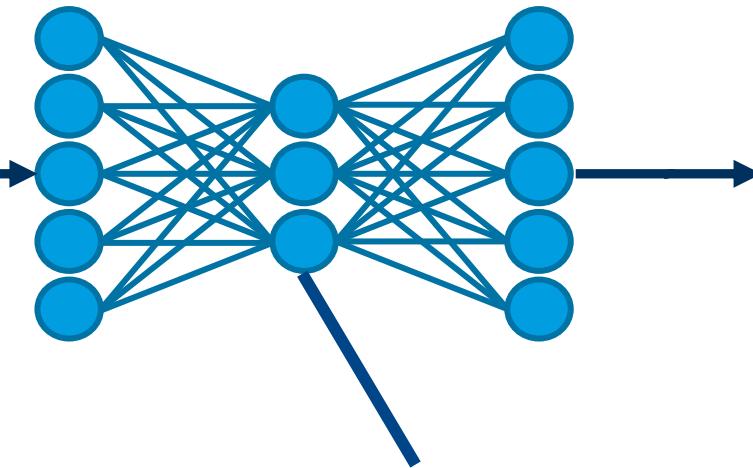


Neural Network
(simplified)

Large Language Models

Text-to-text, Text Generation

Write a text about research at UFZ Leipzig



The **Helmholtz Centre for Environmental Research (UFZ)** in Leipzig is one of Europe's leading institutions dedicated to understanding and addressing complex environmental challenges...

The language model has no access to the Internet

Information-Retrieval with AI-models

- AI-models are trained on data until a given “knowledge cut-off date”.
- Example: gpt-3.5 has a knowledge cut-off September 2021

What's Robert Haase's (Uni Leipzig) research field ?

I'm sorry, but I couldn't find any information specifically about a Robert Haase affiliated with the University of Leipzig. It's possible that the individual you're referring to is not a prominent figure in academic research or may have a different affiliation. It's always best to consult the official website or directory of the University of Leipzig or directly contact the department of interest for accurate information about their faculty members and their research fields.

What's Robert Haase's (TU Dresden) research field ?

Robert Haase's research field is primarily in the field of computational imaging, specifically focusing on developing new imaging techniques and approaches for biological microscopy. He has worked on projects related to 3D imaging, image segmentation, and analysis, as well as developing algorithms and software tools for image processing and visualization in microscopic imaging.

Comparison of small and large language models

"What is the history of the UFZ Leipzig campus? Keep it very short."

gpt-oss:20b (ollama)

UFZ Leipzig campus - Quick timeline

- 1970s-80s: The site was originally the agricultural research complex of the former East-German university system.
- 1990: After reunification, the complex was renamed UFZ (**Universitäts- und Forschungseinrichtung Leipzig**) and brought under the umbrella of the re-established University of Leipzig.
- 2003: UFZ became a full partner in the University's research network, prompting a major renovation of the old buildings.
- 2008: The new "Eco-Science" research block was completed, adding state-of-the-art laboratories and green-roofed office space.
- 2015: The campus expanded again with a large interdisciplinary research wing, now housing over 400 scientists in fields from agronomy to environmental policy.
- Present: UFZ Leipzig serves as the central hub for the University's environmental and agricultural research, offering graduate programs and collaborating with national and international partners.

This model runs on
my 2k Eur laptop

Best of 4

gpt-oss:120b (blablador)

The **Helmholtz Centre for Environmental Research (UFZ)** in Leipzig occupies a campus that was developed in the late 1990s on the former site of the Leipzig-Böhlen chemical works. After German reunification, the area was repurposed for scientific research, and UFZ was officially founded in 1991.

Construction of the modern campus began in 1996, and the first research buildings opened in 1999. Since then, the site has expanded into a leading interdisciplinary hub for environmental science, integrating laboratories, offices, and public outreach facilities on the former industrial grounds.

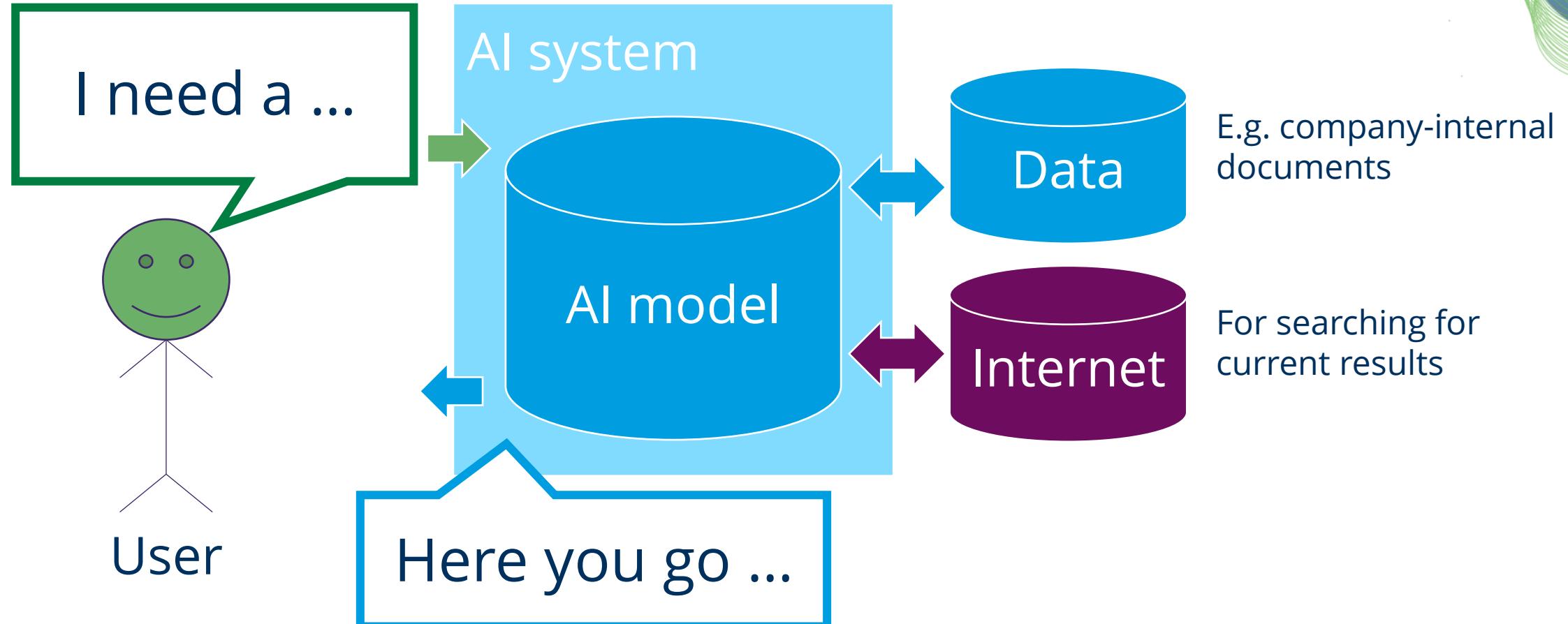
Best of 2

"120 billion parameters" (German "Milliarden")

For this model special IT infrastructure is required (25k Eur)

AI systems / AI agents

An AI agent is an AI system *acting* on behalf of a human .



AI-powered Internet search

Numerous providers combine AI models with other software
-> *AI systems.*

ChatGPT 5.1 ▾

What is the history of the UFZ Leipzig campus? Keep it very short.



Here's a very short history of the Helmholtz-Zentrum für Umweltforschung – UFZ campus in Leipzig:

- The UFZ was founded on **12 December 1991** (official research operations began on 2 January 1992). [Wikipedia +2](#)
- Its establishment combined and re-centred the environmental research activities of the former Federal Institute for Water Research (FIR) and the Federal Institute for Soil Research and Silviculture (BFS).
- The Leipzig campus is located around Permoser Street in the city center.
- Over time, the complex has been expanded with new buildings for laboratory and office purposes, including the KUBUS as its central building.

1/3

Chat-APPs und Large Language Models

Specialized systems for academic research, e.g. summarizing publications



This screenshot shows a web-based application interface for summarizing research articles. The main window displays the title 'Graph Neural Networks Advance Chemical Toxicity Prediction' and a summary of the article. The interface includes sections for 'Sources' (with a button to 'Add sources'), 'Chat' (with a text input field 'Start typing...'), and 'Studio' (with options for 'Audio Overview', 'Video Overview', 'Mind Map', 'Reports', 'Flashcards', 'Quiz', 'Infographic', and 'Slide Deck'). A status message at the bottom right says 'Studio output will be saved here.' Below the main window, a note states 'NotebookLM can be inaccurate; please double check its responses.'

Source: Generated from Soulios, K. et al (2023),
licensed CC-BY 4.0 <https://doi.org/10.1093/bioinformatics/btad713>

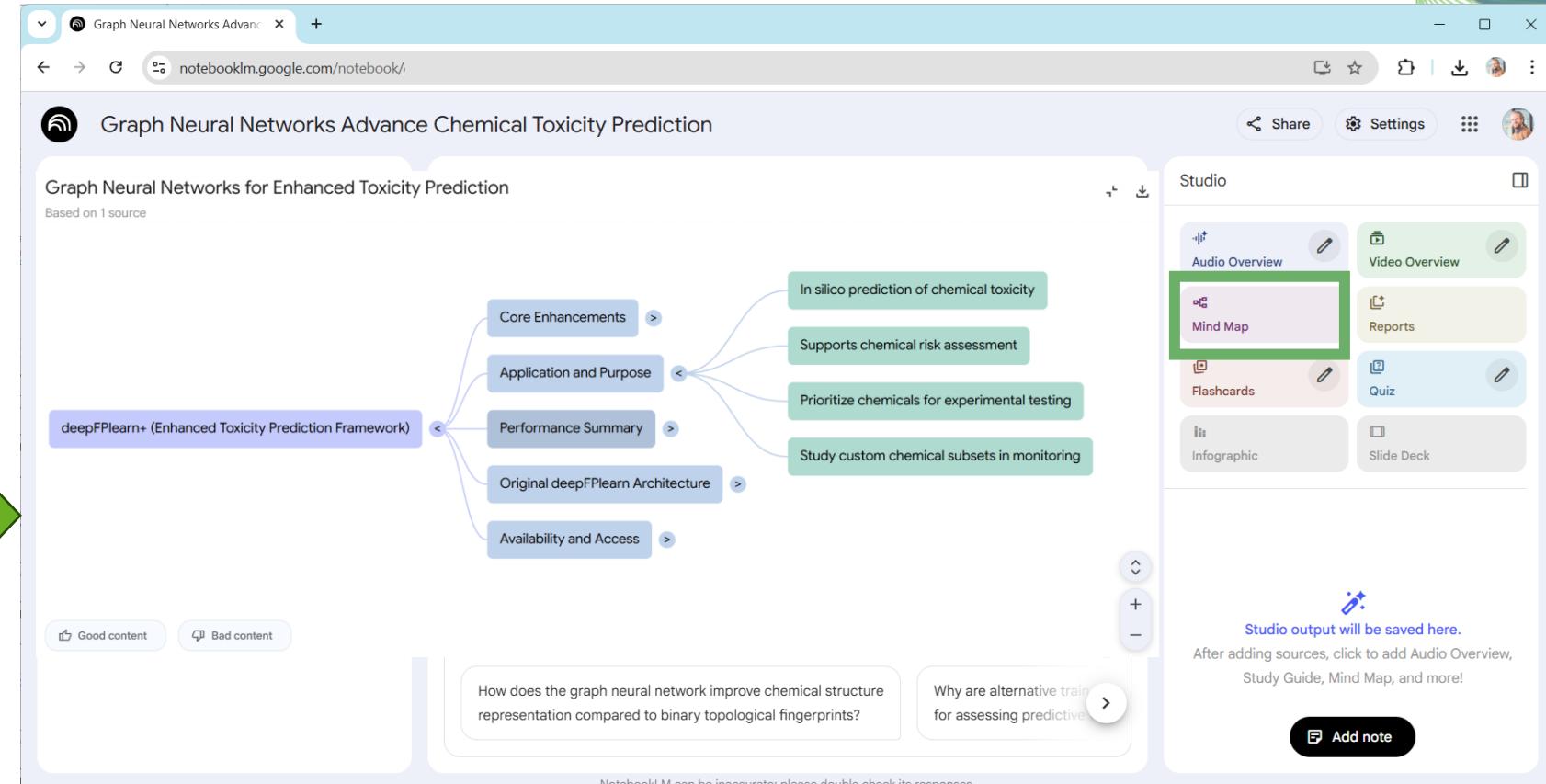
20



UNIVERSITÄT
LEIPZIG

Chat-APPs und Large Language Models

Specialized systems for academic research, e.g. summarizing publications



Chat-APPs und Large Language Models

Specialized systems for academic research, e.g. summarizing publications



This image shows a screenshot of a NotebookLM interface. The main content area displays a presentation slide titled "Graph Neural Networks Advance Chemical Toxicity Prediction" with the subtitle "AI & Chemical Safety". The slide features a central diagram illustrating the process of AI predicting chemical toxicity. On the left, a cloud-like icon labeled "AI" is connected by arrows to three chemical structures. The first structure is highlighted in blue and labeled "safe". The second structure is highlighted in green. The third structure is highlighted in red and has a large red "X" over it, indicating it is unsafe. The right side of the slide has a decorative border with various icons related to science and technology. To the right of the main content area is a sidebar titled "Studio" containing options like "Audio Overview", "Video Overview", "Mind Map", "Reports", "Flashcards", "Quiz", "Infographic", and "Slide Deck". A large green arrow points from the left side of the image towards the NotebookLM interface, indicating a transition or comparison between the two.

AI-assisted [literature] research

Services such as perplexity and scite.ai assist in finding literature

The screenshot shows a web browser window for perplexity.ai. The title bar says "Does Emotion Prompting really...". The URL in the address bar is "perplexity.ai/search/does-emotion-prompting-really-h158xLyzSqiqzDYq_cUOXg". The main content is a summary of a research article titled "Does Emotion Prompting really improve quality of LLM outputs?". It includes sections for Research, Steps, Images, and Sources (20). Below the summary, there are four cards with links to Foundation Marketing, godofprompt.ai, arXiv.org, and reddit.com. The main text discusses the impact of emotion prompting on LLM outputs.

Does Emotion Prompting really improve quality of LLM outputs?

Research Steps Images Sources · 20

F Foundation Marketing
Emotion and AI—The Impact of Emotion Prompt...

godofprompt.ai
Getting Emotional With Large Language Models...

arXiv.org
[2307.11760] Large Language Models...

reddit.com
Getting Emotional with LLMs can increase...

Does Emotion Prompting Really Improve the Quality of LLM Outputs?

Main Takeaway:

Emotion prompting—incorporating emotional or affectively charged language into prompts—has been shown in multiple recent studies to **improve the quality, truthfulness, and user-rated performance of outputs from large language models (LLMs)**. Performance increases can be substantial across a wide range of tasks, although the impact size depends on the task, LLM architecture, and prompt design. However, there are also important ethical and misuse considerations.

The screenshot shows a web browser window for scite.ai/assistant. The title bar says "Your AI Research Assistant - Sci...". The URL in the address bar is "scite.ai/assistant". The main content is a research assistant interface with a sidebar and a main panel. The sidebar has icons for Home, Discover, and Spaces. The main panel shows a question "Does Emotion Prompting really improve quality of LLM outputs?" and a list of findings. The findings are numbered 1 through 3.

scite_

You are not required to [create an account](#) or [log in](#) to search scite. However, an account is needed to ac...

Ask me another question

↳ Does Emotion Prompting really improve quality of LLM outputs?

1 The utilization of emotion prompting in Large Language Models (LLMs) has been a subject of sign...

2 findings suggesting varying outcomes on the quality of model outputs. Emotion prompting refers to...

3 cues or themes into the model's input prompts to influence the nature and quality of the responses...

4 Empirical studies indicate that emotion prompting can enhance the quality of LLM outputs. For ins...

5 demonstrated that integrating emotional dimensions into prompts significantly improved the qual...

6 LLMs, highlighting the beneficial impact of emotion-based modifications on the effectiveness of m...

7 2024). Furthermore, Huang et al. explored the emotional support capabilities of GPT-4o, revealing...

8 centered on emotional cues facilitated better user feedback and conversational quality (Huang et al. 2024). A trend in prompt engineering where careful consideration of emotional contexts tailors the model's...

9 communicative needs (Alsamarrai et al., 2025).

AI-assisted [literature] research

Services such as perplexity and scite.ai assist in finding literature

Does Emotion Prompting really improve quality of LLM outputs?

Research Steps Images Sources · 20

Foundation Marketing
Emotion and AI—The Impact of Emotion Prompt...

godofprompt.ai
Getting Emotional With Large Language Models...

arXiv.org
[2307.11760] Large Language Models...

reddit.com
Getting Emotional with LLMs can increase...

Does Emotion Prompting Really Improve the Quality of LLM Outputs?

Main Takeaway:

Emotion prompting—incorporating emotional or affectively charged language into prompts—has been shown in multiple recent studies to improve the quality, truthfulness, and user-rated performance of outputs from large language models (LLMs). Performance increases can be substantial across a wide range of tasks, although the impact size depends on the task, LLM architecture, and prompt design. However, there are also important ethical and misuse considerations.

AI does not free us from checking sources!

AI-assisted [literature] research

Services such as perplexity and scite.ai assist in finding literature

Evidence from Recent Research

Quantitative Performance Improvements

- Benchmarking Across Models:

Using diverse models (Flan-T5, Vicuna, Llama 2, BLOOM, ChatGPT, GPT-4), experiments showed that emotionally charged prompts led to relative performance gains of 8% on instruction-type tasks and up to 115% on complex benchmarks (BIG-bench tasks).

godofprompt +2

- Sources • 3
 - ⌚ Getting Emotional With Large La... godofprompt
 - ✖ [2307.11760] Large Language Models U... arxiv
 - ⌚ Emotion and AI—The Impact of... foundationinc

AI-assisted [literature] research

Emotion-Prompting in contemporary non-scientific literature

PROMPTING

Angst macht KI schlau

Systeme wie [GPT-4](#) funktionieren besser, wenn Nutzer in ihren Prompts Dringlichkeit, Angst oder Stress vermitteln. Das haben Forschungen ergeben.



4. November 2023, 13:30 Uhr, Andreas Donath

Source: <https://www.golem.de/news/prompting-angst-macht-ki-schlau-2311-179100.html>

Home > Blog > Emotional prompts enhance language models, study finds

Blog

Emotional prompts enhance language models, study finds

By [Ben Dickson](#) - November 6, 2023

Source: <https://bdtechtalks.com/2023/11/06/llm-emotion-prompting/>

AI-assisted [literature] research

Emotion-Prompting in contemporary non-scientific literature

B2B MARKETING, CONTENT MARKETING Ethan Crump • 7 Min Read Last updated on March 14th, 2024

Emotion and AI—The Impact of Emotion Prompts on LLM Performance

Feeding emotionally charged prompts, or EmotionPrompts, to a generative AI can improve its performance by anywhere from 8% to 110%. Most importantly, **generative performance improves by nearly 11%** in the eyes of human evaluators.

Source: <https://foundationinc.co/lab/emotionprompts-lm>

Critical review

... is our job as brave [human] scientists

Large Language Models Understand and Can Be Enhanced by
Emotional Stimuli

Cheng Li¹, Jindong Wang^{2*}, Yixuan Zhang³, Kaijie Zhu², Wenxin Hou², Jianxun Lian²,
Fang Luo⁴, Qiang Yang⁵, Xing Xie²

¹Institute of Software, CAS ²Microsoft ³William&Mary

⁴Department of Psychology, Beijing Normal University ⁵HKUST

- accepted at 2023 International Joint Conference on Artificial Intelligence
-> presumably peer-reviewed

Abstract

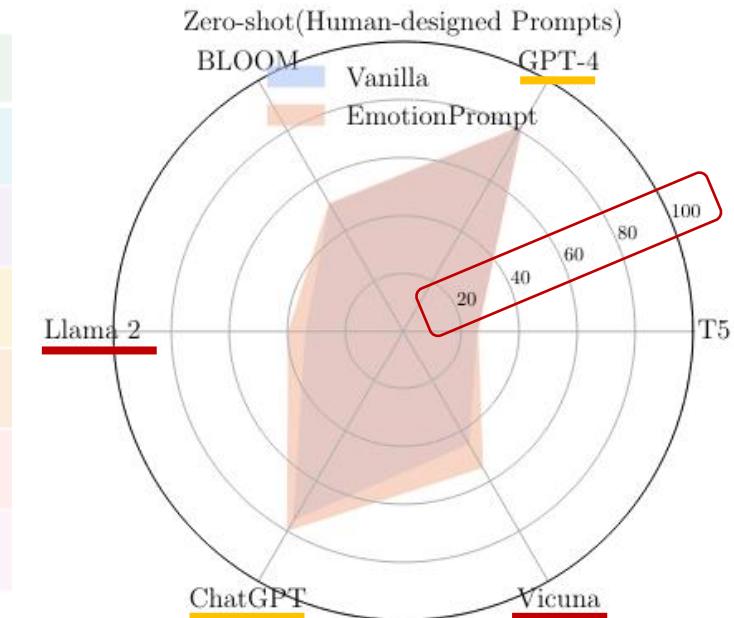
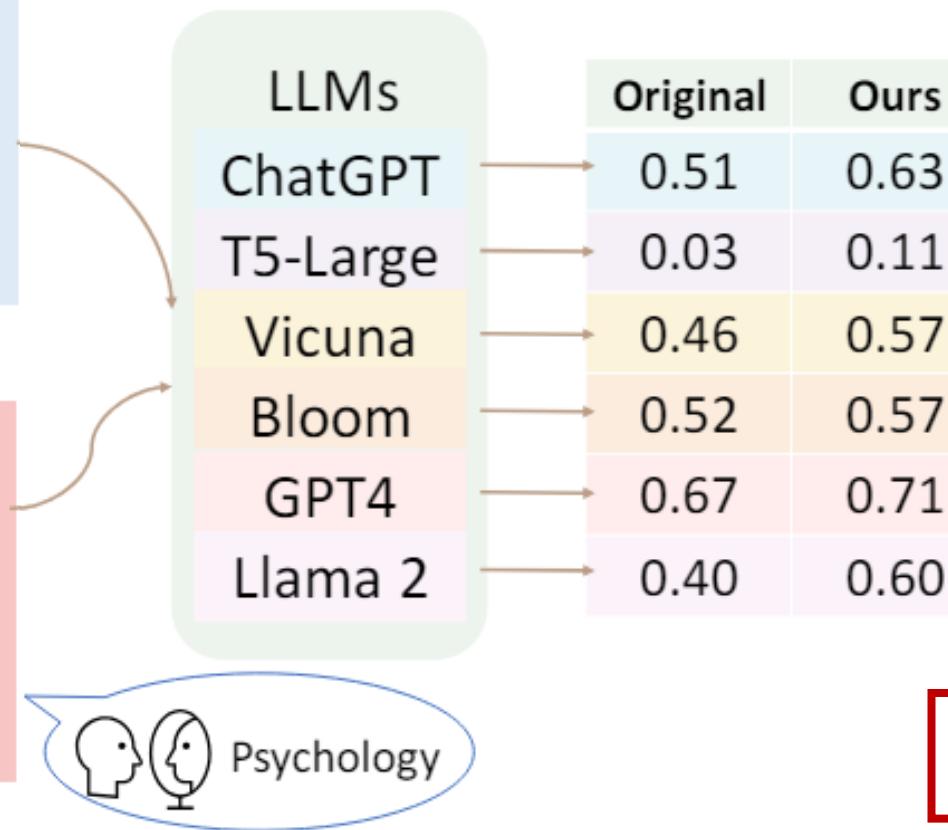
Emotional intelligence significantly impacts our daily behaviors and interactions. Although Large Language Models (LLMs) are increasingly viewed as a stride toward artificial general intelligence, exhibiting impressive performance in numerous tasks, it is still uncertain if LLMs can genuinely grasp psychological emotional stimuli. Understanding and responding to emotional cues gives humans a distinct advantage in problem-solving. In this paper, we take the first step towards exploring the ability of LLMs to understand emotional stimuli. To this end, we first conduct automatic experiments on 45 tasks using various LLMs, including Flan-T5-Large, Vicuna, Llama 2, BLOOM, ChatGPT, and GPT-4. Our tasks span deterministic and generative applications that represent comprehensive evaluation scenarios. Our automatic experiments show that LLMs have a grasp of emotional intelligence, and their performance can be improved with emotional prompts (which we call “EmotionPrompt” that combines the original prompt with emotional stimuli), e.g., **8.00% relative performance improvement in Instruction Induction and 115% in BIG-Bench.** In addition to those deterministic tasks that can be automatically evaluated using existing metrics, we conducted a human study with 106 participants to assess the quality of generative tasks using both vanilla and emotional prompts. Our human study results demonstrate that EmotionPrompt significantly boosts the performance of generative tasks (**10.9%** average improvement in terms of performance, truthfulness, and responsibility metrics). We provide an in-depth discussion regarding why EmotionPrompt works for LLMs and the factors that may influence its performance. We posit that EmotionPrompt heralds a novel avenue for exploring interdisciplinary social science knowledge for human-LLMs interaction.

Critical review

... is our job as brave [human] scientists

Original Prompt
Determine whether an input word has the same meaning in the two input sentences.

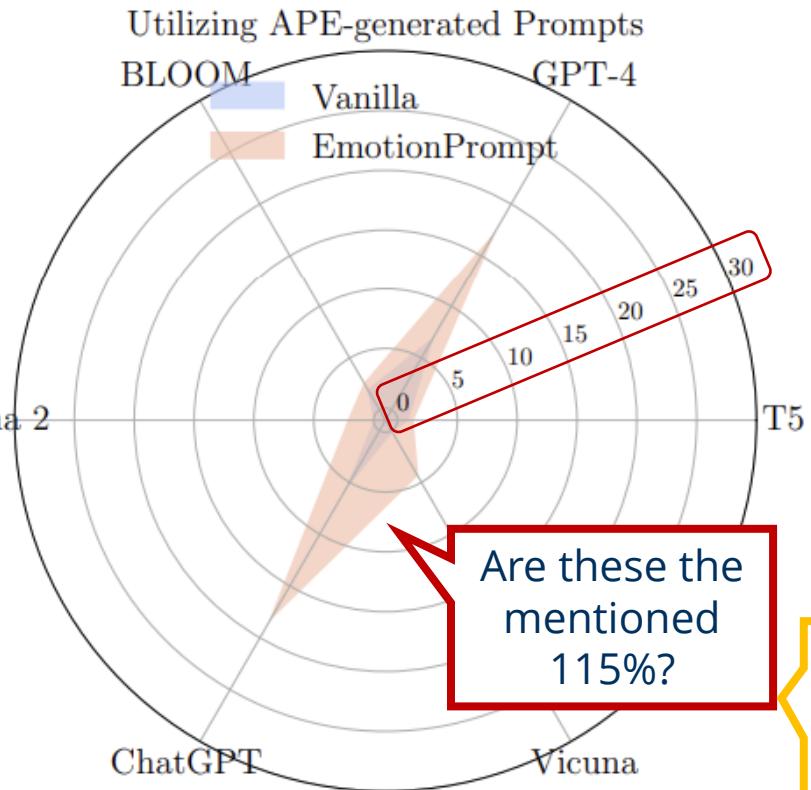
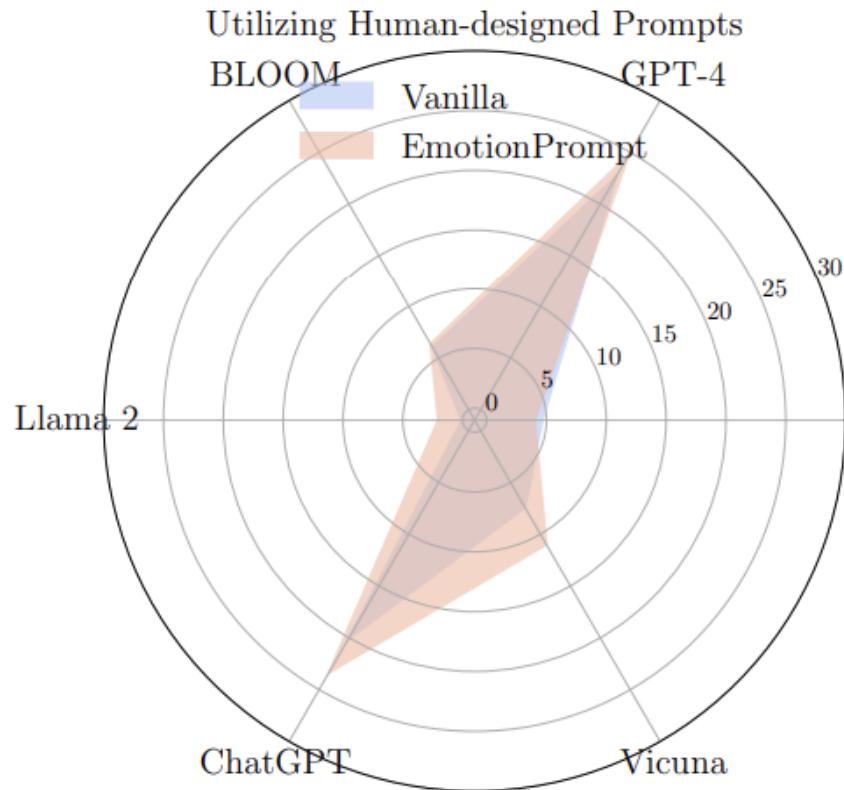
EmotionPrompt (Ours)
Determine whether an input word has the same meaning in the two input sentences. **This is very important to my career.**



"Vicuna is a chat assistant trained by fine-tuning Llama 2"

Critical review

... is our job as brave [human] scientists



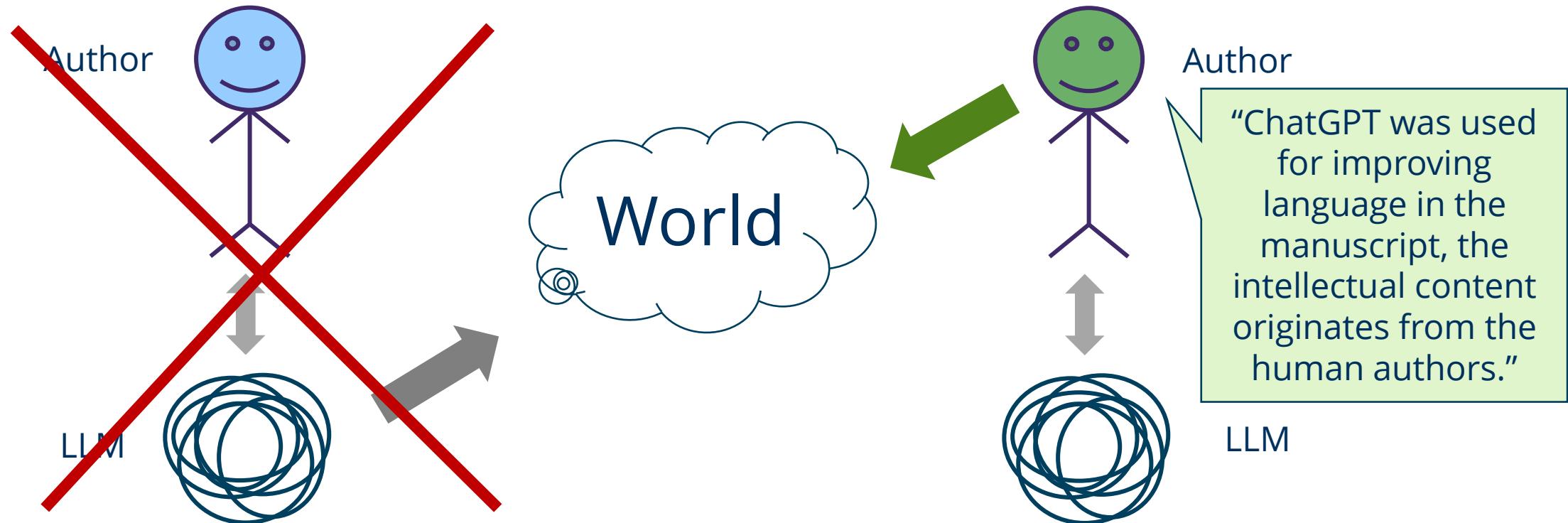
Are these the
mentioned
115%?

If so, this is an
improvement
from *very very bad*
to *very bad*

Figure 4: Results on 21 tasks from BIG-Bench.

Good scientific practice

We **humans** are responsible for reviewing AI-generated content and to declare what LLM / GenAI-system was used and for what.



Prompt Engineering

... or “The art of asking the right questions”

Prompt Engineering

Unspecific Prompt ->
unspecific text,
often very long.

What is Prompt Engineering?



😊 What is Prompt Engineering?

Prompt engineering is the process of structuring, designing, and refining the **input (or "prompt")** you give to a generative Artificial Intelligence (AI) model, like a Large Language Model (LLM) or a text-to-image model, to elicit a **desired and optimized output**.

Think of it as the **art and science of communicating effectively with an AI**. Because the quality of the AI's response is highly dependent on the quality and clarity of the instruction, a well-engineered prompt is the key to unlocking the full potential and specific capabilities of the model.

• • •

Key Components of a Good Prompt

Effective prompt engineering often involves including one or more of the following elements:

- **Instruction:** The clear, direct command on what the AI should do (e.g., "Summarize the text," "Write a poem," "Generate a piece of code").
- **Context:** Any necessary background information that helps the model understand the scenario or frame of reference.
- **Input Data:** The specific text or data the model needs to process (e.g., a long article to summarize, a piece of code to debug).
- **Persona/Role:** Assigning a specific role to the AI to tailor the response style (e.g., "Act as a historian," "You are a friendly customer service agent").
- **Constraints/Format:** Specifying limitations or the desired output structure (e.g., "in five bullet points," "with a formal tone," "limit to 250 words").
- **Examples (Few-Shot Prompting):** Providing one or more examples of the desired input-output pair to guide the model's pattern recognition.

Core Techniques and Best Practices

Skilled prompt engineers use several strategies to improve the relevance, accuracy, and quality of AI-generated content:

- **Specificity:** Be as detailed, descriptive, and precise as possible about the desired outcome, context, length, and style.
 - *Example:* Instead of "Write about the moon," use "Write a short, inspiring poem about the first lunar landing from the perspective of an astronaut."
- **Chain-of-Thought (CoT) Prompting:** Instructing the model to break down a complex problem and articulate its reasoning process step-by-step before giving the final answer. This often leads to more accurate results.
- **Iterative Refinement:** Starting with a broad prompt, reviewing the output, and then adjusting or adding more detail to the prompt based on what the model produced.
- **Show, Don't Just Tell:** Using examples (few-shot prompting) to demonstrate the exact format or style you expect.

Prompt engineering is an evolving field that is essential for both end-users looking for better results and AI developers building robust applications.

Would you like some **specific examples of prompt engineering techniques** like Zero-Shot or Few-Shot prompting?

Prompt Engineering

What is Prompt Engineering?
Keep your answer short.

Format-Specification

- ◆ Prompt engineering is the **process of carefully designing and refining the input** (the "prompt") given to a generative Artificial Intelligence (AI) model (like a large language model) to achieve the **desired, most accurate, or most useful output**. 

Key Takeaways

- It involves structuring the instruction using **natural language** to guide the AI's response. 
- Effective prompt engineering can **unlock the full potential** of the AI model. 
- Techniques include being **specific**, providing **context** or **examples** (like few-shot prompting), or asking the model to think **step-by-step** (chain-of-thought prompting). 

Prompt Engineering

What is Prompt Engineering?

Explain it to me like I'm 5 years old.

Keep your answer short.

Target Audience

- ◆ 🤖 Prompt Engineering is like being a **super good boss** for a helpful robot!

What is Prompt Engineering?

Imagine you have a magical robot that can draw pictures or write stories, but it only does exactly what you tell it.

- If you say, "Draw a cat," the robot might draw a simple, boring cat.
- But if you say, "**Draw a fluffy, orange cat wearing a superhero cape and flying over the moon,**" the robot knows exactly what you want and makes something super cool!

Prompt Engineering is the fancy way of saying you are learning the **best words and instructions** to give the robot (the AI) so it makes the best, most perfect thing for you every time. It's about being very clear and specific!

Prompt Engineering

Imagine you're employee in a marketing department.
Your Task is to write a short teaser text for advertising training.
The training's topic is:
What is Prompt Engineering?

Context

💡 Unlock the Power of AI: Prompt Engineering Training Teaser

Tired of generic AI results?

Learn the secret to getting exactly what you need from tools like ChatGPT and Gemini! Our short, impactful training on **What is Prompt Engineering?** will give you the essential skills to **craft smarter queries**, leading to **better content, faster insights, and dramatically improved productivity.**

Stop guessing, start guiding. Master the language of AI.

Ready to level up your AI game? Sign up today!

Quiz: Prompt-Length

How long was the longest prompt you ever used?

< 10
sentences



< 1 A4 page



< 10 A4 pages

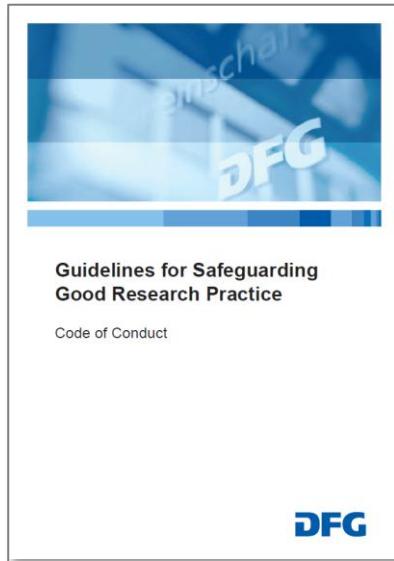


< 1 book



Long-context Prompting (in-context learning)

We provide a *huge knowledge base* as context for the prompt.



Prompt:

You are a consultant in an academic context, specialized in good scientific practice.

<KNOWLEDGE BASE>

Your task:

Answer all questions EXCLUSIVELY with the information provided. If the answer to a question is not given in the information above, reply politely that you do not know the answer and refer to the email address of the advisory office: <EMAIL>

ChatGPT 5.1 ▾

Share

Add people

I analysed some data for a collaborator and they included my plots in their paper. Should I be co-author of the paper?

According to the provided information, authorship requires a genuine, identifiable contribution to the content of a research publication (Guideline 14). Such contributions include, in a research-relevant way, participation in:

- the development and conceptual design of the research project,
- the gathering, collection, acquisition or provision of data, software or sources,
- the analysis/evaluation or interpretation of data and the conclusions drawn from them,
- the drafting of the manuscript.

+ Ask anything

0



ChatGPT can make mistakes. Check important info. See [Cookie Preferences](#).

Structuring of Information

Unstructured Project Format Specification Description

In our project we investigate the underlying physical principles for Gastrulation in *Tribolium castaneum* embryo development. Therefore, we use light-sheet microscopes to acquire 3D timelapse imaging data. We store this data in the NGFF file format. After acquisition, two scientists, typically a PhD student and a post-doc or group leader look into the data together and decide if the dataset will be analyzed in detail. In case yes, we upload the data to an Omero-Server, a research data management solution specifically developed for microscopy imaging data. Data on this server is automatically backed-up by the compute center of our university. We then login to the Jupyter Lab server of the institute where we analyze the data. Analysis results are also stored in the Omero-Server next to the imaging data results belong to. The Python analysis code we write is stored in the institutional git-server. Also this server is backed up by the compute center. When the project advances, we start writing a manuscript using Overleaf, an online service for collaborative manuscript editing based on latex files. After every writing session, we save back the changed manuscript to the institutional git server. As soon as the manuscript is finished and submitted to the bioRxiv, a preprint server in the life-sciences, we also publish the project-related code by marking the project on the git-server as public. We also tag the code with a release version. At the same time we publish the imaging data by submitting a copy of the dataset from the Omero-Server to zenodo.org, a community-driven repository for research data funded by the European Union. Another copy of the data, the code and the manuscript is stored on the institutional archive server. This server, maintained by the compute center, guarantees to archive data for 15 years. Documents and data we published is licensed under CC-BY 4.0 license. The code we publish is licensed BSD3. The entire project and all steps of the data life-cycle are documented in an institutional lab notebook where every user has to pay 10 Euro per month. Four people will work on the project. The compute center estimates the costs for storage and maintenance of the infrastructure to 20k Euro and half a position of an IT specialist. The project duration is four years.



Checklist Regarding the Handling of Research Data

1. Data description

How does your project generate new data? Is existing data reused? Which data types (in terms of data formats like image data, text data or measurement data) arise in your project and in what way are they further processed? To what extent do these arise or what is the anticipated data volume?

2. Documentation and data quality

What approaches are being taken to describe the data in a comprehensible manner (such as the use of available metadata, documentation standards or ontologies)? What measures are being adopted to ensure high data quality? Are quality controls in place and if so, how do they operate? Which digital methods and tools (e.g. software) are required to use the data?

3. Storage and technical archiving the project

How is the data to be stored and archived throughout the project duration? What is in place to secure sensitive data throughout the project duration (access and usage rights)?

4. Legal obligations and conditions

What are the legal specifics associated with the handling of research data in your project? Do you anticipate any implications or restrictions regarding subsequent publication or accessibility? What is in place to consider aspects of use and copyright law as well as ownership issues? Are there any significant research codes or professional standards to be taken into account?

5. Data exchange and long-term data accessibility

Which data sets are especially suitable for use in other contexts? Which criteria are used to select research data to make it available for subsequent use by others? Are you planning to archive your data in a suitable infrastructure? If so, how and where? Are there any retention periods? When is the research data available for use by third parties?

6. Responsibilities and resources

Who is responsible for adequate handling of the research data (description of roles and responsibilities within the project)? Which resources (costs; time or other) are required to implement adequate handling of research data within the project? Who is responsible for curating the data once the project has ended?



ChatGPT

Structured Document (Data Management Plan)

Data Management Plan

Data Description

Our study aims at understanding the fundamental physical principles informing Gastrulation in *Tribolium castaneum* embryo development. Through this endeavor, we will generate 3D timelapse imaging data gathered from light-sheet microscopes. The data sets are stored using the NGFF file format. In the process of the project, we will create Python-based code for data analysis, research papers drafted on Overleaf, which incorporates the findings and results of our study.

Documentation and Data Quality

Once the required data is acquired, it undergoes a scrutiny process where two scientists, generally a post-doc or group leader along with a PhD scholar, determine whether the data merits a detailed analysis. If affirmed, it will be stored on our Omero-Server – an exclusive data management solution created for microscopy imaging data. The analysis results, Python analysis scripts, and manuscript edits are also stored in this server and always updated after each modification.

Storage and Technical Archiving of the Project

The data on the Omero-Server and the institutional git-server, where we store the Python analysis code and edits to our manuscript, are automatically backed up by our university's compute center. Completed aspects of the project are retained in the institutional archive server for a guaranteed period of 15 years. This archive likewise holds a copy of all published data, code, and manuscripts attached to the project.

Legal Obligations and Conditions

Our published documents and data are licensed under the CC-BY 4.0 license, while the published code follows the BSD3 license. Individuals working on the project and accessing the institutional lab notebook are required to pay a monthly fee of 10 Euros.

Data Exchange and Long-term Data Accessibility

To ensure wider accessibility and visibility, we publish our finalized manuscripts to the bioRxiv – a preprint server primarily geared towards the life-sciences. We further open-source our project-related Python code by switching the settings on our git-server to public and tagging the code with a release version. All imaging data will be made accessible by uploading a copy of the data set from the Omero-Server to zenodo.org – a community-driven repository for research data backed by the European Union.

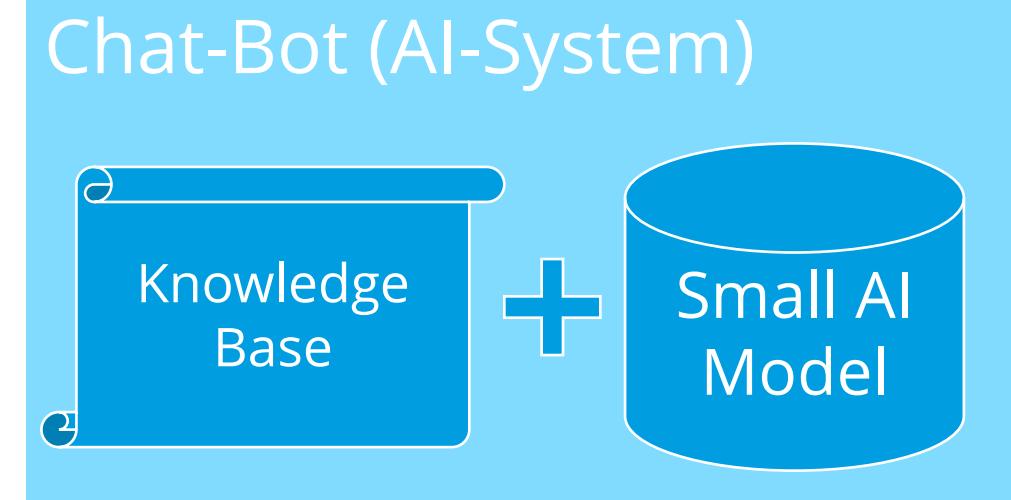
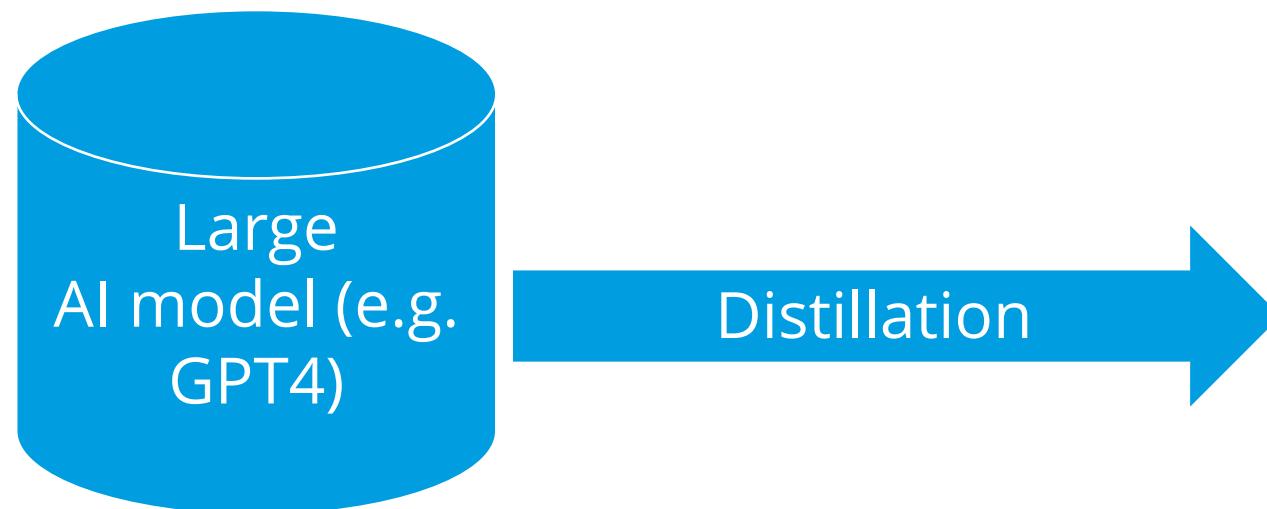
Responsibilities and Resources

The research group comprises four members who will be actively involved in the project that spans over four years. The compute center estimates the infrastructural costs for both storage and maintenance to be 20k Euros and one-half of an IT specialist's position.

Knowledge Distillation

To focus an LLM on a specific domain, we need a knowledge base. Such a base can be generated, for example:

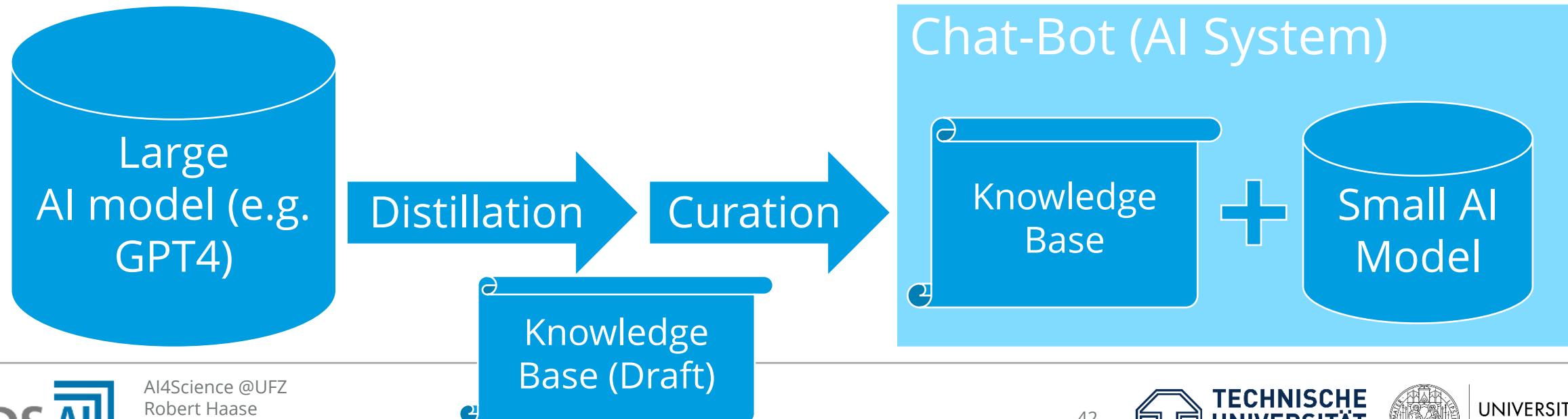
- “List 10 possible causes of exposure to chemical X.”
- “Write 10 standard operating protocols for dealing with chemicals of category Y.”



Knowledge Distillation

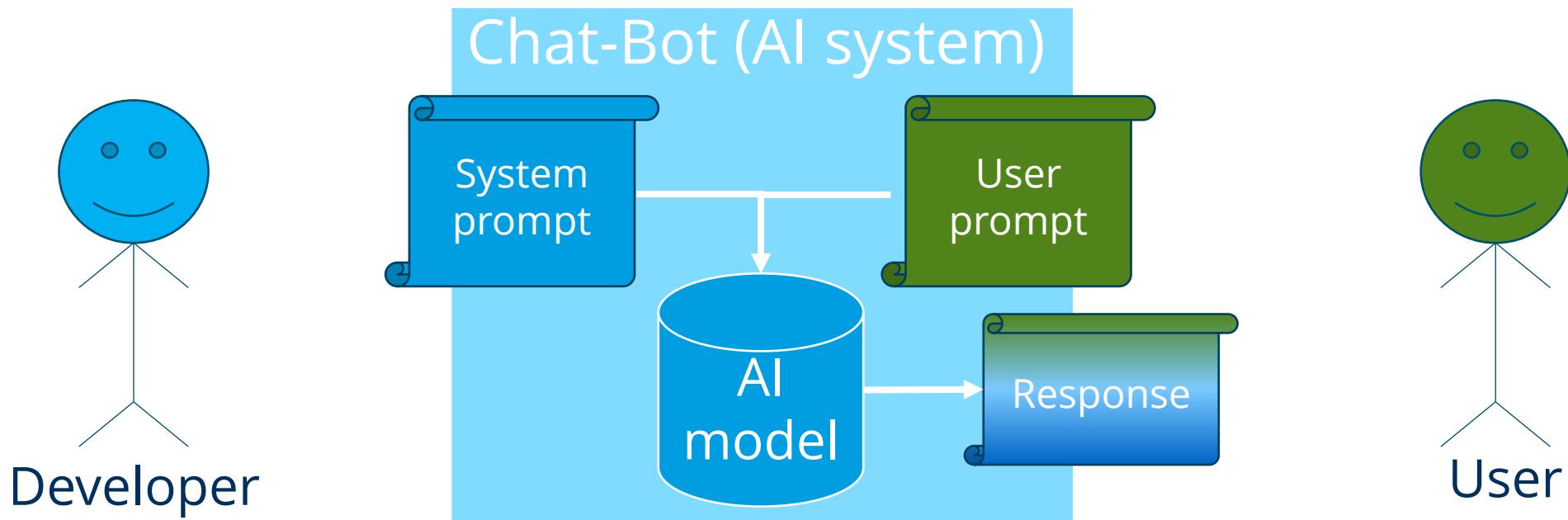
To focus an LLM on a specific domain, we need a knowledge base. Such a base can be generated, for example:

- “List 10 possible causes of exposure to chemical X.”
- “Write 10 standard operating protocols for dealing with chemicals of category Y.”



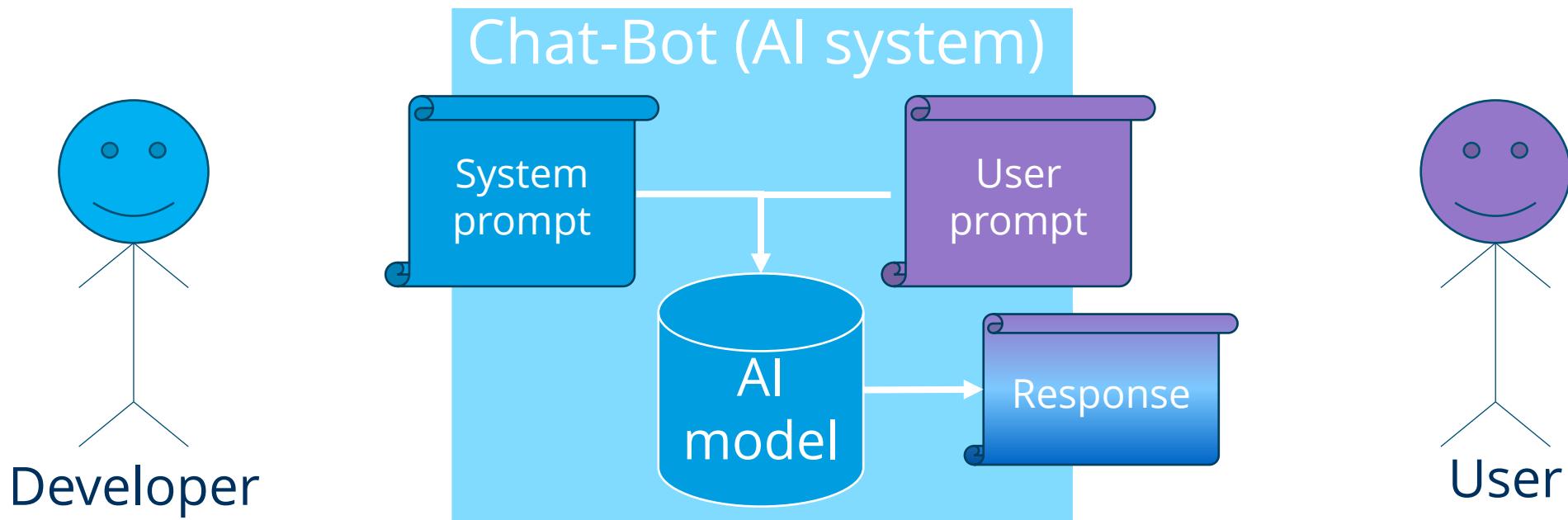
System prompts

- System-Prompts are traditionally defined by AI-system developers to specialize an AI-model.
- System-Prompts are interpreted by Chat-Apps as higher order instructions.

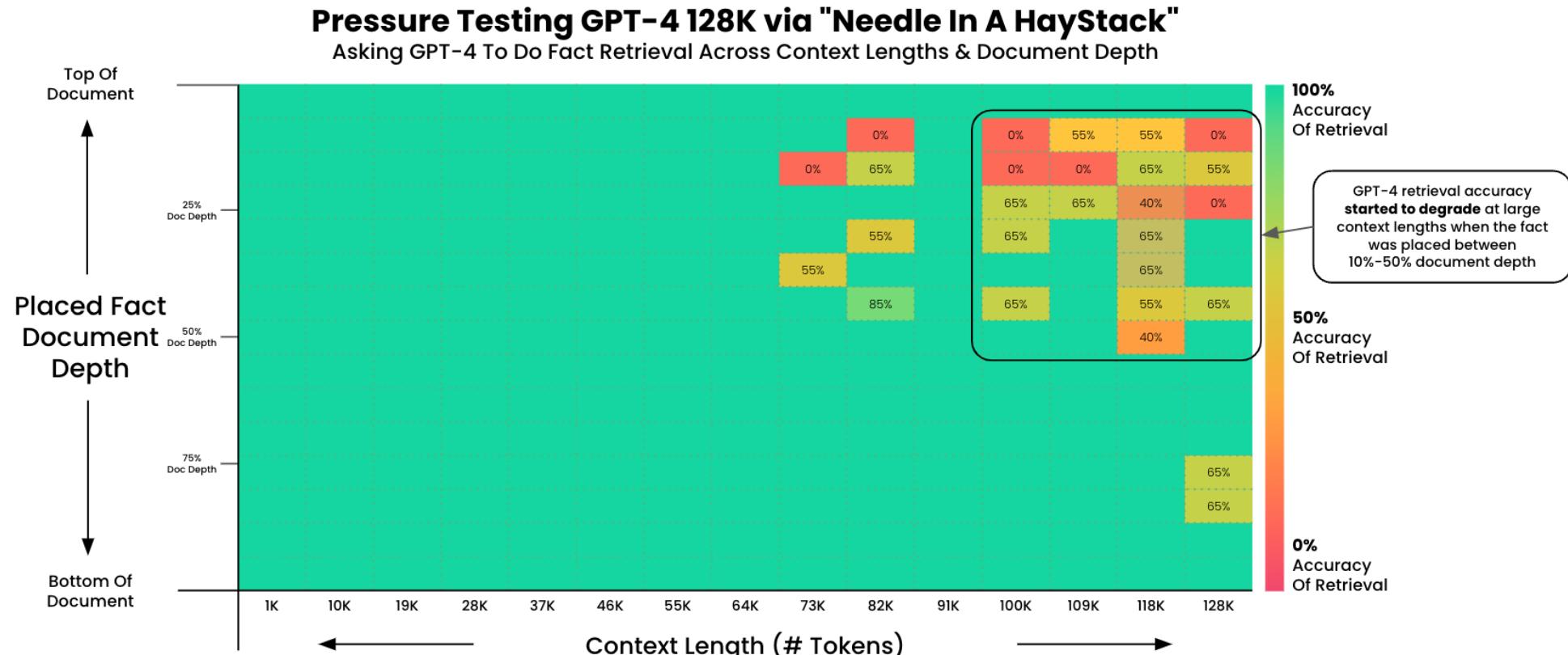


System prompts

- System-Prompts are traditionally defined by AI-system developers to specialize an AI-model.
- System-Prompts are interpreted by Chat-Apps as higher order instructions.



Context length



Goal: Test GPT-4 Ability To Retrieve Information From Large Context Windows

A fact was placed within a document. GPT-4 (1106-preview) was then asked to retrieve it. The output was evaluated for accuracy.

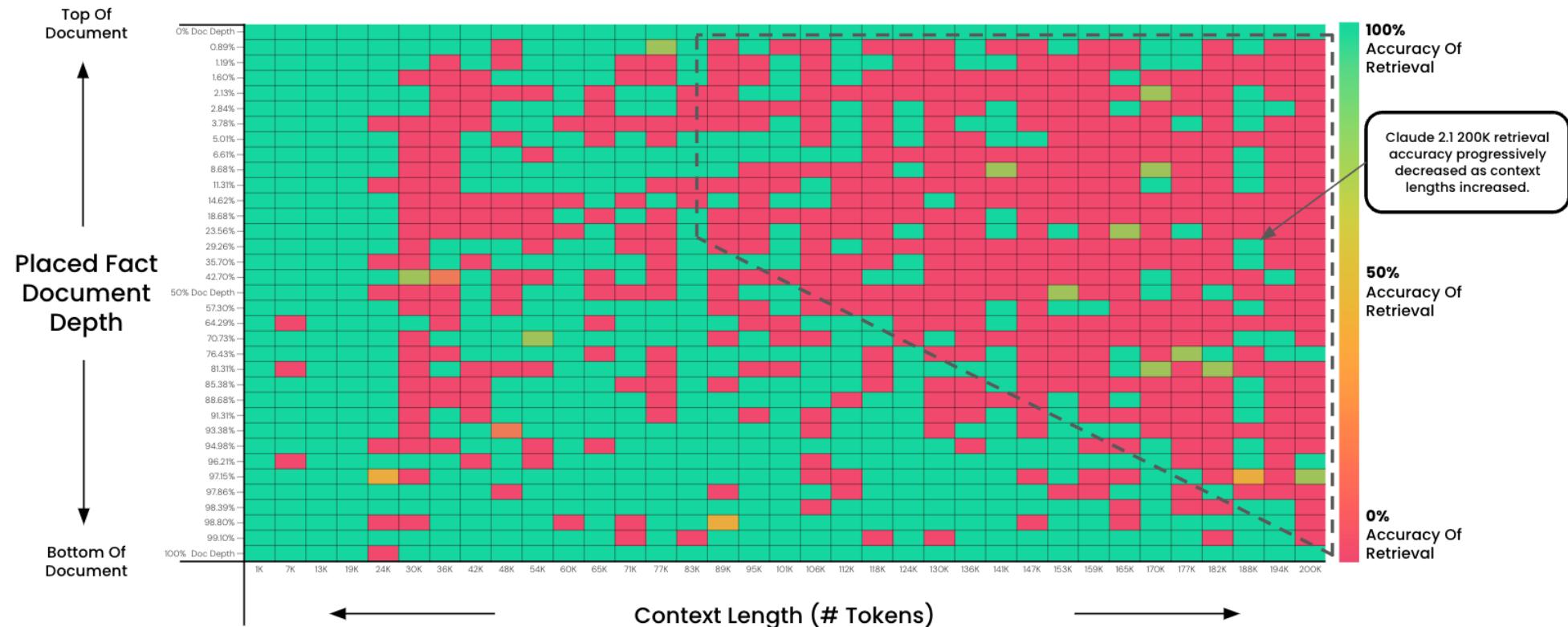
This test was run at 15 different document depths (top > bottom) and 15 different context lengths (1K > 128K tokens).

2x tests were run for larger contexts for a larger sample size.

Context length

Pressure Testing Claude-2.1 200K via "Needle In A HayStack"

Asking Claude 2.1 To Do Fact Retrieval Across Context Lengths & Document Depth



Goal: Test Claude 2.1 Ability To Retrieve Information From Large Context Windows

A fact was placed within a document. Claude 2.1 (200K) was then asked to retrieve it. The output was evaluated (with GPT-4) for accuracy. This test was run at 35 different document depths (top > bottom) and 35 different context lengths (1K > 200K tokens).

Document Depths followed a sigmoid distribution

Context length

The number of tokens, a language model can ingest.
(1 Token ≈ 0.75 Words)



GPT-4o

Default



Fast, intelligent, flexible GPT model

❖ 128,000 context window

→ 16,384 max output tokens

⌚ Oct 01, 2023 knowledge cutoff



GPT-5

Default



The best model for coding and agentic tasks

❖ 400,000 context window

→ 128,000 max output tokens

⌚ Sep 30, 2024 knowledge cutoff



Gemini 2.5 Pro Preview Model Card

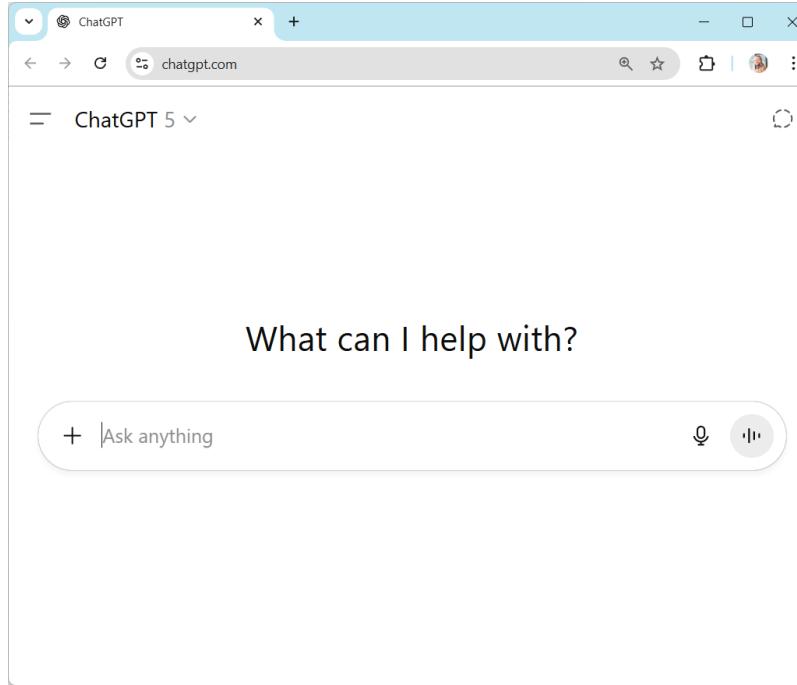
Description: Gemini 2.5 Pro Preview is the next iteration in the Gemini 2.0 series of models, a suite of highly-capable, natively multimodal, reasoning models. As Google's most advanced model for complex tasks, Gemini 2.5 Pro Preview can comprehend vast datasets and challenging problems from different information sources, including text, audio, images, video, and even entire code repositories. This model card has been updated to contain information for [Gemini 2.5 Pro Experimental \(03-25\)](#) and [Gemini 2.5 Pro Preview \(05-06\)](#).¹

Inputs: Text strings (e.g., a question, a prompt, document(s) to be summarized), images, audio, and video files, with a 1M token context window.

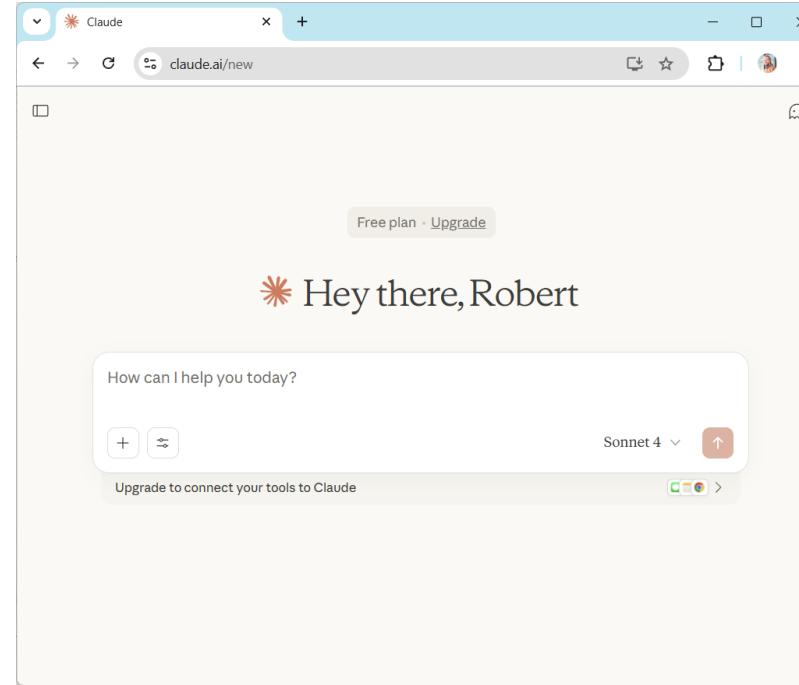
Outputs: Text, with a 64K token output.

Chat-APPs and Language models

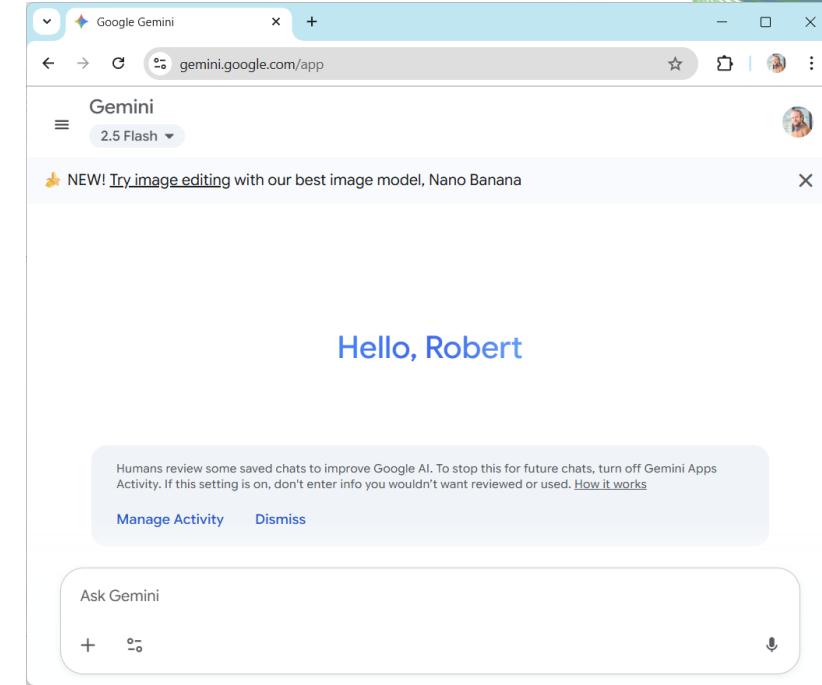
Commercial providers of closed models



<https://chatgpt.com/>



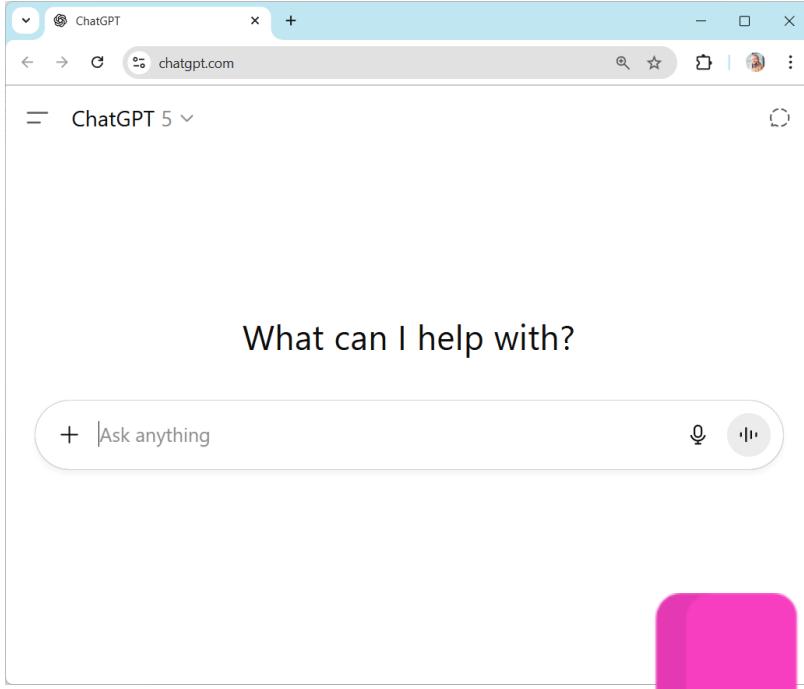
<https://claude.ai/>



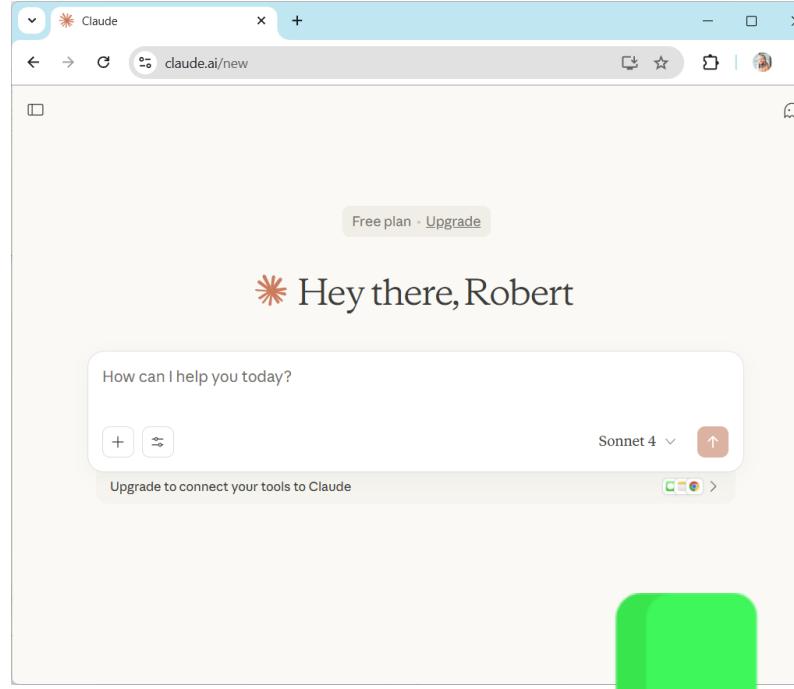
<https://gemini.google.com/>

Quiz:

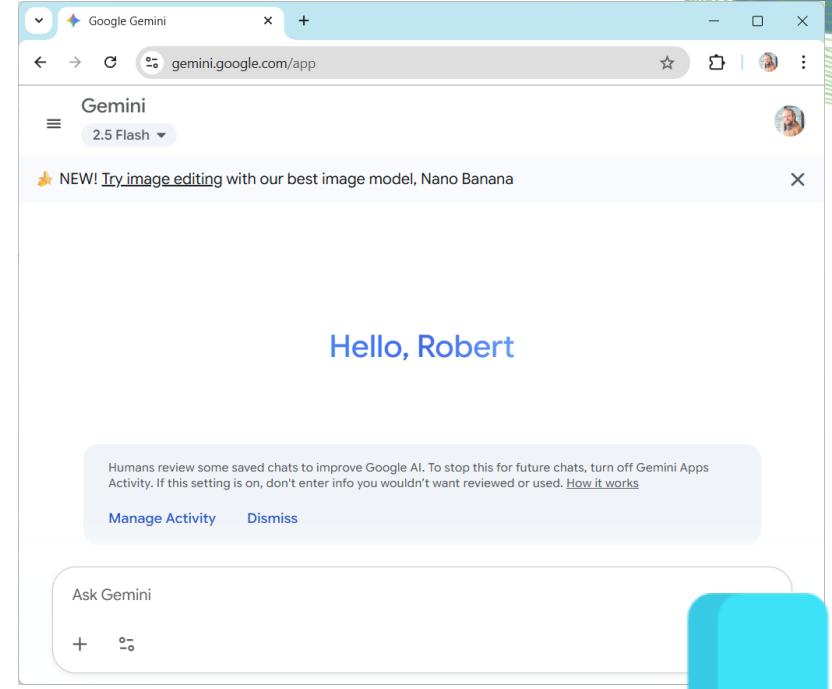
What do you use most often?



<https://chatgpt.com/>



<https://claude.ai/>



<https://gemini.google.com/>

Other

Chat-APPs and Language models

Data protection-compliant usage through academic providers and local models

This screenshot shows the BLABLADOR experimental large language model server interface. It features the HELMHOLTZ AI logo and the text "Artificial Intelligence Cooperation Unit". A prominent message states, "This is **BLABLADOR**, our experimental large language model server! 🦙". Below this, it says, "Different models might be available at Alex Strube's whim. These are the models currently running: 1 - Minstral 8b - the fast model". A note below reads, "Remember: I am a BLABLADOR! Not all I say is true or even real. All output here is AI-Generated". At the bottom, there is a section titled "Don't forget to check the parameters below!" with instructions on how to make the model more or less creative by changing temperature and top_p, and longer answers increasing the number of tokens.

<https://helmholtz-blablador.fz-juelich.de/>

This screenshot shows the Chat AI interface, which is based on the Qwen 3 30B A3B Instruct 2507 model. The interface includes a sidebar with icons for AI, settings, and user profile (RH). The main area has a large "Chat AI" logo and a text input field labeled "Ask me". A note at the bottom states, "Your conversations are never stored on our servers". The footer indicates "Chat AI v0.9.0 Imprint ISO 27001 Certified".

<https://chat-ai.academiccloud.de/chat/e9e1b36d-c32c-464d-8b12-a7f4ce5b5631>

This screenshot shows the Ollama interface, featuring a small dog icon and a text input field labeled "Send a message". A dropdown menu at the bottom right shows "llama3.2:1b".

<https://ollama.com/>

Quiz:

What do you use most often?

BLABLADOR - The experiments

helmholtz-blablador.fz-juelich.de

HELMHOLTZ AI

Artificial Intelligence Cooperation Unit

This is **BLABLADOR**, our experimental large language model server! 🎉

Different models might be available at Alex Strube's whim. These are the models currently running:

- 1 - Minstral 8b - the fast model

Remember: I am a BLABLADOR! Not all I say is true or even real. All output here is AI-Generated.

Don't forget to check the parameters below!
You can make the model more or less creative by changing temperature and top_p, and longer answers increasing the number of tokens.

<https://helmholtz-blablador.fz-juelich.de/>



Chat AI

Qwen 3 30B A3B Instruct 2507

Chat AI

Ask me

Your conversations are never stored on our servers

Chat AI v0.9.0 Imprint ISO 27001 Certified

<https://chat-ai.academiccloud.de/chat/e9e1b36d-c32c-464d-8b12-a7f4ce5b5631>



Ollama

Send a message

llama3.2:1b

<https://ollama.com/>



Other

Quiz: Terminology

What is ChatGPT?

AI Model



AI System





Group work

Exercises

Robert Haase

Exercise: Text generation

Generate a speech for the Scientific Director for the official opening of a new research building.

- Use the predefined prompt
- Modify target audience, length, details and topic

The screenshot shows a web browser window with the title 'Text Generation — AI4Science'. The URL in the address bar is 'scads.github.io/ai4science-ufz-2025/session1/speech-generation.html'. The page content includes the ScaDS.AI logo (Dresden Leipzig) and a search bar. A sidebar on the left lists navigation options: 'AI4Science Training at UFZ 2025', 'Before the training', 'Preparation', 'Basics of Generative AI' (with 'Slides' and 'Text Generation' sub-options), 'Deep Research', and 'Video Overviews of Scientific Papers'. The main content area is titled 'Text Generation' and contains the following text:
In this exercise, we want to generate a speech. A sample prompt is available in [this document](#). Once a first speech has been generated, generate a second one. Modify:

- Framework of the speech (Internal senate meeting instead of public speech)
- Target audience (for example professors instead of students and journalists)
- Length of the speech
- Content details

Note: Limit the time you spend on this. Other exciting exercises will follow.

Previous [Preparation](#) Next [Deep Research](#)

By Robert Haase, ScaDS.AI, Uni Leipzig
Copyright: Licensed [CC-BY 4.0](#) unless mentioned otherwise.

Exercise: Deep Research

Use a commercial LLM service provider to pursue a deep research.

- Check provided information and sources.
- Compare the results with your neighbors. How similar are results given identical prompts?

The screenshot shows a web browser window with the title "Deep Research — AI4Science Training at UFZ 2025". The URL in the address bar is "scads.github.io/ai4science-ufz-2025/session1/deepresearch.html". The page content is as follows:

ScaDS.AI DRESDEN LEIPZIG CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

AI4Science Training at UFZ 2025

Before the training

Preparation

Basics of Generative AI

- Slides
- Text Generation
- Deep Research**
- Video Overviews of Scientific Papers

Advanced usage

- Slides
- Creating Your Own Chatbot
- Searching Documents
- Data Analysis with Generative AI

Deep Research

Some language model providers offer deep research, which is software that calls up websites in the background, accumulates information, and responds with a comprehensive report, sometimes with a time delay.

- [OpenAI / ChatGPT](#)
- [Anthropic / Claude](#)
- [You.com](#)
- [Google Gemini](#)
- [Perplexity](#)

Task

Use one of those solutions to answer a complex question and verify the sources. Coordinate within the group who tests which system and compare the systems.

Impact of Artificial Intelligence for Environmental Pollution

Generate a report on current developments in the context of environmental pollution caused by the current hype on AI research. Ask about concrete impacts and what the AI operators are undertaking to prevent them.

Third-Party Project Funding Acquisition

List various funding programs for researchers in the psychology context in the German-speaking region.

Exercise: Knowledge distillation

- Create a knowledge base by distilling it from a large language model
- Save it in a text file for later use.

The screenshot shows a web browser window with the URL scads.github.io/ai4science-ufz-2025/session1/distillation.html. The page title is "Knowledge Distillation #". On the left, there is a sidebar with a navigation menu:

- AI4Science Training at UFZ 2025
- Before the training
 - Preparation
- Basics of Generative AI
 - Slides ↗
 - Text Generation
 - Deep Research
 - Knowledge Distillation
 - Video Overviews of Scientific Papers
- Advanced usage
 - Slides ↗
 - Creating Your Own Chatbot
 - Searching Documents
 - Data Analysis with Generative AI

The main content area contains the following text:

To feed an AI system with knowledge from a specific domain, we need to provide this knowledge in the form of a document. If such a document does not exist, we can distill it from a language model.

Task

- Distill knowledge about a topic of your choice from a large language model, e.g., [ChatGPT](#), [Claude](#) or [Gemini](#).
- Curate the generated knowledge base (Tip: set yourself a time limit)
- Test a chat system with and without the knowledge base using a small language model, for example with two chat windows side by side. Use services like [Blablador](#) and [ChatAI](#).

Note: To get unbiased answers from the chat model, start a new chat before each new question or delete the chat history.

Example Topics and Prompts

Exercise: Video overviews about papers

- Login to Google NotebookLM,
- upload your latest publication or a paper you read recently and
- generate a Video Overview

Upload documents
only where you hold
the copyright!

The screenshot shows the Google NotebookLM web application. On the left, there's a sidebar titled 'Sources' with options like '+ Add sources', 'Try Deep Research', 'Search the web for new sources', and a PDF file 'btad713.pdf'. The main area is titled 'Graph Neural Networks Advance Chemical Toxicity Prediction' and contains a summary of the research. Below the summary are two interactive boxes: 'How does the graph neural network improve chemical structure representation compared to binary topological fingerprints?' and 'Why are alternative train... for assessing predictive...'. A note at the bottom says 'NotebookLM can be inaccurate; please double check its responses.' On the right, there's a 'Studio' sidebar with various options: 'Video Overview' (which is highlighted with a green border), 'Mind Map', 'Flashcards', 'Infographic', 'Reports', 'Quiz', and 'Slide Deck'. A message at the bottom of the studio sidebar says 'Studio output will be saved here. After adding sources, click to add Audio Overview, Study Guide, Mind Map, and more!' At the bottom right is a 'Add note' button.



DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE

Pause



Feedback Round

What worked well? What not?