



“Zwischen Tool und Forschungswerkzeug”. Generative KI in den (digitalen) Geisteswissenschaften

**Zwischen Codices und Codes
Innovative Perspektiven in der Mittelalter-
und Frühneuzeitforschung.**
Ringvorlesung des IZMF. 21.10.2024

Christopher Pollin

Digital Humanities Craft OG
www.dhcraf.org

Institut für Digitale Geisteswissenschaften,
<https://digital-humanities.uni-graz.at>

“Zwischen Tool und Forschungswerkzeug”. Generative KI in den (digitalen) Geisteswissenschaften



Generative KI

“A toy, until it isn’t”

Grundlagen LLM und Prompt Engineering

“Tools on demand”

o1-preview

“The Intelligence Age” (!?)

Aerosol paint illustration of a golden robot working with books,
stacks of medieval documents in front, full body shot, white
background, Ultra HD

Generative KI ... in 03:26



Erstellung eines Videos mit Sora

Analyse des Videos mit GPT-4 und
Erstellung eines Storyboards für eine
„Naturdokumentation“.

Klonen der eigenen Stimme

Erstellung eines Films, in dem eine Stimme das
Geschehen beschreibt

Übersetzung in andere Sprachen

<https://twitter.com/nlw/status/1806468441001574717>

Published

Late nights, dino delights

Prompt: an epic, orchestra, multi-voice Disney outro

orchestra epico film score atmospheric orchestral

Create Share Edit ...

Play • Heart + ⌂

A colorful illustration of two children walking through a jungle at night, with a large blue dragon-like creature and small dinosaurs in the background.

udio.com

<https://www.udio.com/songs/7bdnE5TuTF9jDTTt1G3M49>

<https://www.udio.com/songs/eZibZhroz5WWTUUrLLfapKp>



OpenAI. Sora.

<https://openai.com/sora>

Matthew Berman. OpenAI's "World Simulator" SHOCKS The Entire Industry | Simulation Theory Proven?!. https://youtu.be/BH9FU7Gd6v8?si=mR1bUre_FVLaQUA-

AI Explained. Sora - Full Analysis (with new details).

<https://youtu.be/nYTRFKGR9wQ?si=V3QTXC9gp2Z3o2yE>





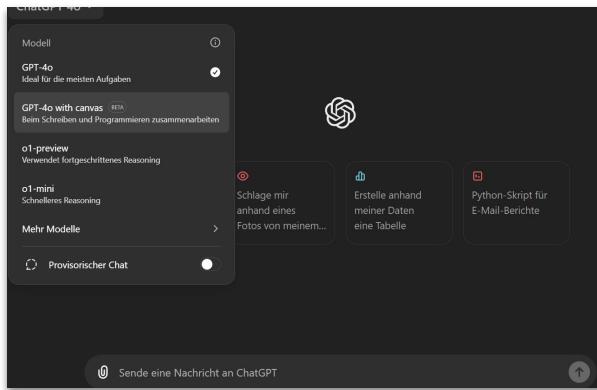
“A toy, until it isn’t”

Ethan Mollick. Gradually, then Suddenly: Upon the Threshold.

<https://www.oneusefulthing.org/p/gradually-then-suddenly-upon-the>

Hyper realistic diptych of transforming objects with white background, child's hand holding building blocks. adult hand writing in a planner. Warm to cool color shift. Depth-of-field blur techmecha-cat in the style of Tekkonkinkreet anime --style
raw --v 6.1 --ar 16:9 --q 2

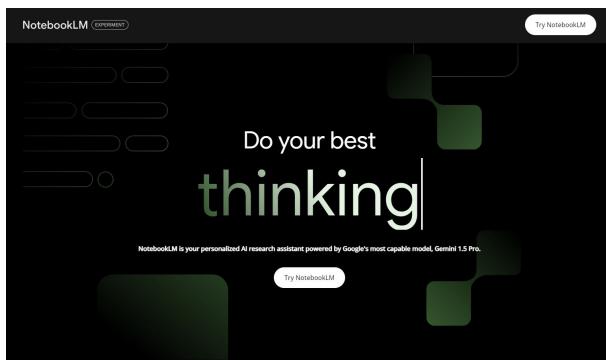
Tools



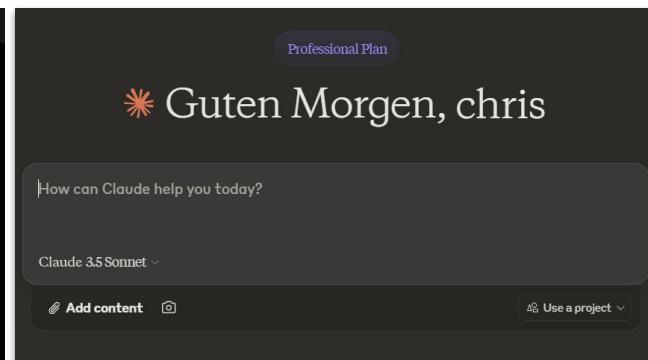
<https://chatgpt.com>

This screenshot shows the Google AI Studio interface. On the left, under "Google AI Studio", there are sections for "Get API key", "Create new prompt", and "New tuned model". Below these are "My library" and "Prompt Gallery". To the right, the main area is titled "Defining Artificial Intelligence" with a sub-section "System Instructions". It includes optional tone and style instructions for the model. Further down is a section titled "3. Learning and Adapting:" which lists goals, examples, and limitations. Another section, "4. Four Categories (by Arend Hintze):", describes four types of AI based on memory and self-awareness. At the bottom, there's a "Key Takeaway:" summary and a "Type something" input field.

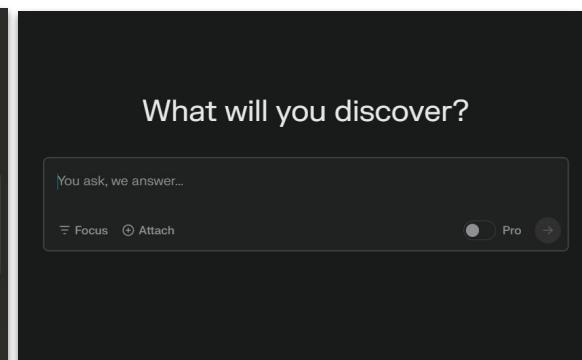
<https://aistudio.google.com>



<https://notebooklm.google.com>

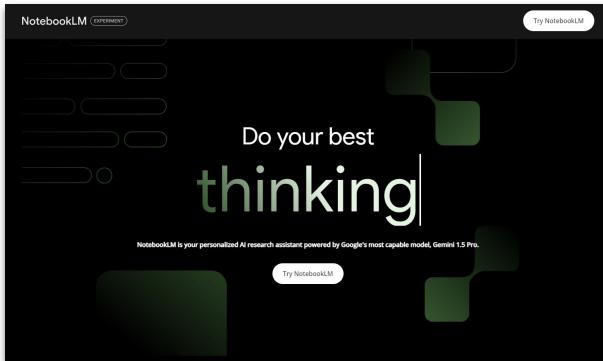


<https://claude.ai>



<https://www.perplexity.ai>

NotebookLM ist ein brauchbares Tool!



<https://notebooklm.google.com>

<https://gams.uni-graz.at/o:corema.a1.recipes>
<https://gams.uni-graz.at/o:corema.b1.recipes>
<https://gams.uni-graz.at/o:corema.b2.recipes>

Taxonomie der Zutaten in mittelalterlichen Kochbüchern

Die gegebenen Quellen bieten eine Fülle an Informationen über Zutaten in mittelalterlichen Rezepten, die sich in folgende Kategorien einteilen lassen:

1. Früchte:

- **Steinobst:** Kirschen, Sauerkirschen, Pflaumen, Zwetschgen, Mispeln, Pfirsiche [1](#) [2](#) [3](#)
- **Kernobst:** Äpfel, Birnen [2](#) [4](#) [5](#) [6](#) ...
- **Beeren:** Maulbeeren, Schlehen, Erdbeeren [2](#) [3](#) [5](#) [9](#) [10](#)
- **Andere Früchte:** Feigen, Weinbeeren [7](#) [11](#) [12](#) [13](#) [14](#)

2. Gemüse:

- **Blattgemüse:** Kohl, Mangold, Melde, Salat [15](#) [16](#) [17](#)
- **Zwiebelgewächse:** Zwiebeln, Knoblauch, Lauch [7](#) [10](#) [18](#) [19](#) ...
- **Wurzelgemüse:** Rüben [17](#) [18](#)
- **Hülsenfrüchte:** Erbsen, Linsen, Kichererbsen, Saubohnen, Bohnen [22](#) [23](#) [24](#) [25](#) ...
- **Kürbisgewächse:** Kürbis [28](#)

3. Nüsse und Samen:

- **Nüsse:** Haselnüsse, Mandeln, Nüsse [11](#) [29](#) [30](#) [31](#) ...
- **Samen:** Hanf [35](#)

4. Getreide:

- **Weizen:** Weizen [36](#) [37](#)
- **Roggen:** Roggen [36](#)
- **Andere Getreide:** Gerste, Hafer, Dinkel, Hirse [26](#) [38](#) [39](#) [40](#) ...

5. Tierische Produkte:

- **Fleisch:** Rindfleisch, Schweinefleisch, Lammfleisch, Ziegenfleisch, Wildbret (Hirsch, Reh), Kalbfleisch [17](#) [28](#) [42](#) [43](#) ...
- **Geflügel:** Hühner, Hähne, Kapaune, Rebhühner, Birkhühner, Auerhähne, Stare, Fasane, Wassergeflügel (Reiher, Wildenten) [9](#)



Kyle Corbitt 
@corbtt

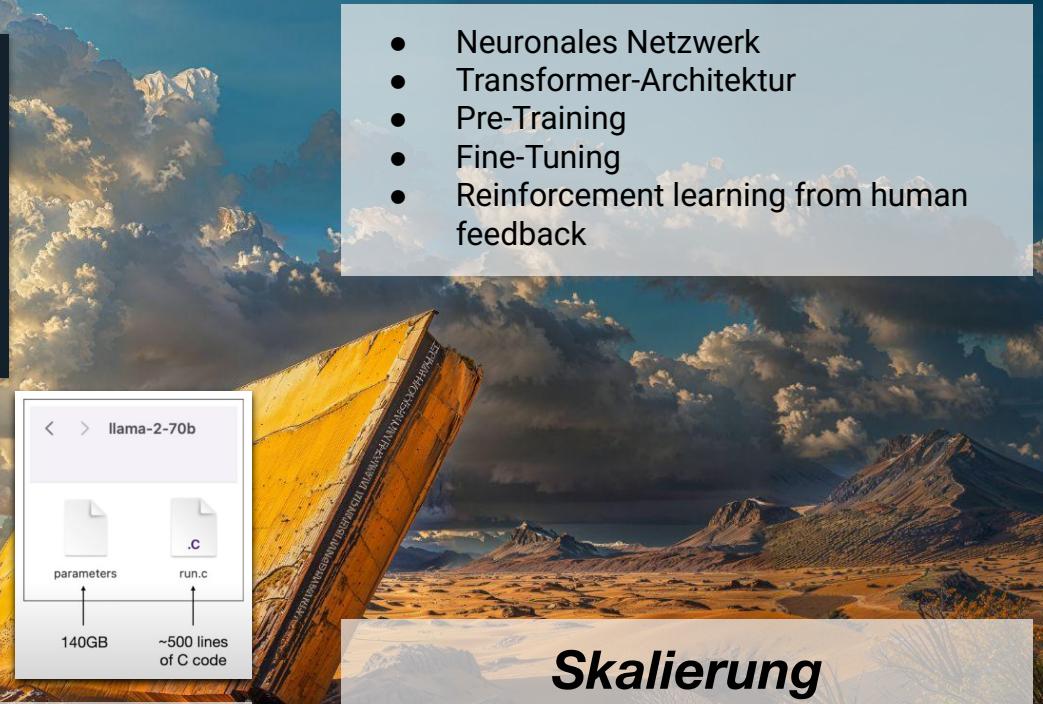
Spoke to a Microsoft engineer on the GPT-6 training cluster project. He kvetched about the pain they're having provisioning infiniband-class links between GPUs in different regions.

Me: "why not just colocate the cluster in one region?"

Him: "Oh yeah we tried that first. We can't put more than 100K H100s in a single state without bringing down the power grid." 

Last edited 10:38 PM · Mar 25, 2024 · 1.8M Views

- Neuronales Netzwerk
- Transformer-Architektur
- Pre-Training
- Fine-Tuning
- Reinforcement learning from human feedback



“Large Language Models (LLM) are like having a Zip-File of the internet”

[nur nicht mehr vollständig entpackbar]

Andrej Karpathy. [1hr Talk] Intro to Large Language Models.

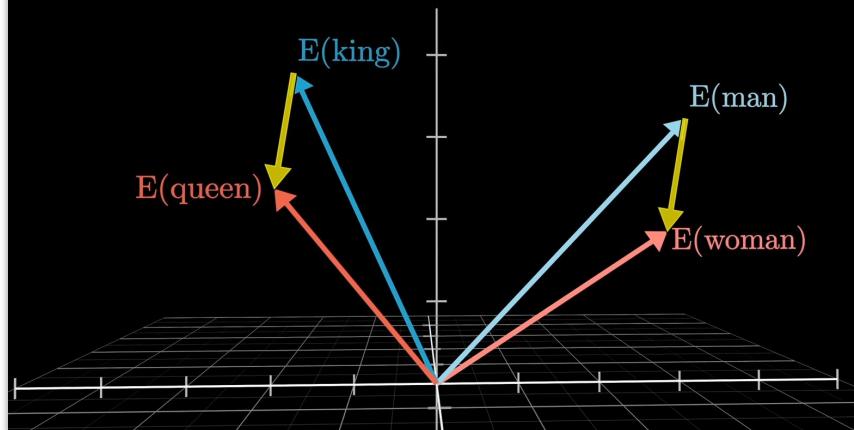
https://www.youtube.com/watch?v=zjkBMFhNj_g&list=WL&index=16

Surreal painting of a hyper realistic and surreal gigantic yellow folder with zipper, like a desktop icon, ultra detailed, salvador dali desert background, landscape, by Frederic Edwin Church -v 6.1

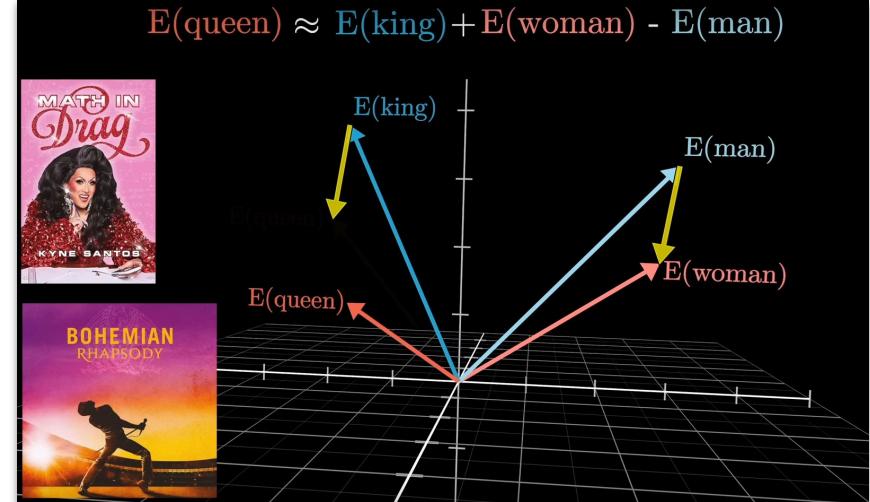
Embeddings

Hochdimensionale Vektordarstellungen von Token

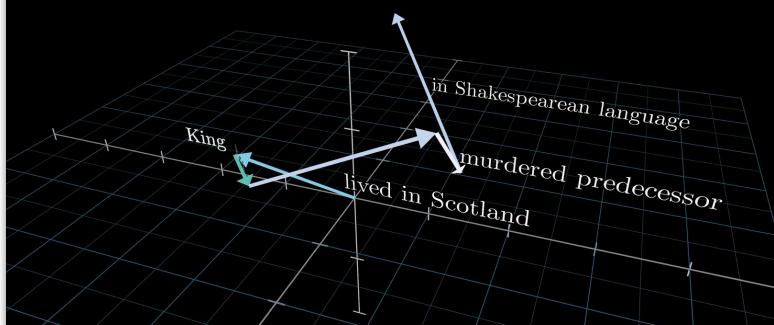
$$E(\text{queen}) - E(\text{king}) \approx E(\text{woman}) - E(\text{man})$$



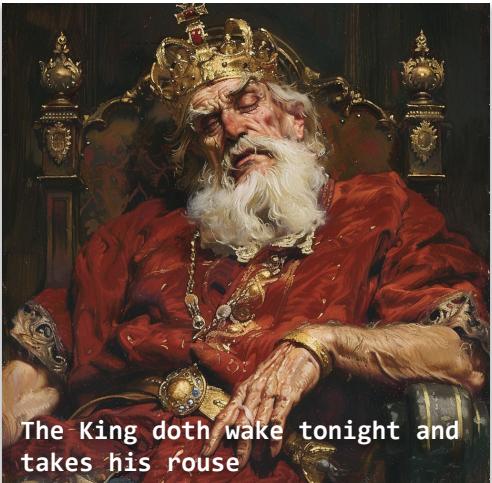
$$E(\text{queen}) \approx E(\text{king}) + E(\text{woman}) - E(\text{man})$$



The King doth wake tonight and takes his rouse

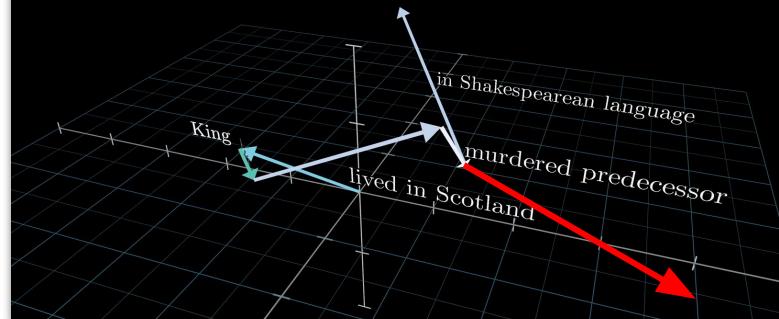


“Shakespearean English”



The King doth wake tonight and takes his rouse

The King wakes up tonight and begins his celebration



“Modern English”



The King wakes up tonight and begins his celebration



You

The King doth wake tonight and takes his rouse



ChatGPT

The line “The King doth wake tonight and takes his rouse” echoes the style of Shakespearean drama, capturing a moment where a king rises and partakes in his revels or festivities. This phrase suggests a setting where the king, perhaps after a period of rest or inactivity, awakens to engage actively in celebration or the duties of his reign. In Shakespeare’s works, the actions of kings often carry significant weight, setting the stage for developments in plot and character dynamics.

In understanding this phrase, it’s important to unpack the historical and literary context. During Shakespeare’s time, the actions of royalty were symbolic, reflecting the health and vigor of the kingdom at large. A king waking and taking his “rouse” (a term that can refer to a drinking bout, a call to arms, or simply waking from sleep to activity) could signify a turning point or a moment of decision that will impact the narrative direction.

The usage of “doth” and “takes his rouse” encapsulates the poetic and formal language

“Let’s think step by step!

“Prompting is weird. Prompting matters.”

Ethan Mollick. Captain's log: the irreducible weirdness of prompting AIs.
<https://www.oneusefulthing.org/p/captains-log-the-irreducible-weirdness>

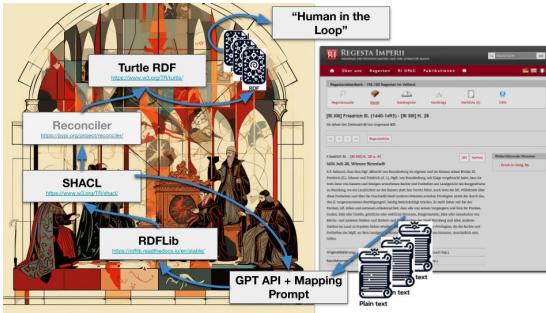
Prompt Engineering ist die Kunst und Wissenschaft, präzise und effektive Eingabeaufforderungen (Prompts) für KI-Systeme wie LLMs zu formulieren, um gewünschte Ausgaben oder Verhaltensweisen zu erzielen.

Chain of Thought =
“Simulation eines Pseudo-Reasonings”





Daten- erfassung

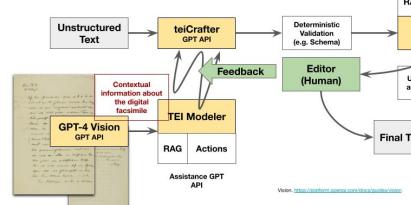


Daten- transformation

...

Pollin, C. (2024). Workshopreihe
“Angewandte Generative KI in den (digitalen)
Geisteswissenschaften” (v1.1.0). Zenodo.
<https://doi.org/10.5281/zenodo.10647754>

Workflow: GPT-4 + Prompt Engineering + API + Assistance API + "Editor in the Loop" + Multimodality GPT-Vision

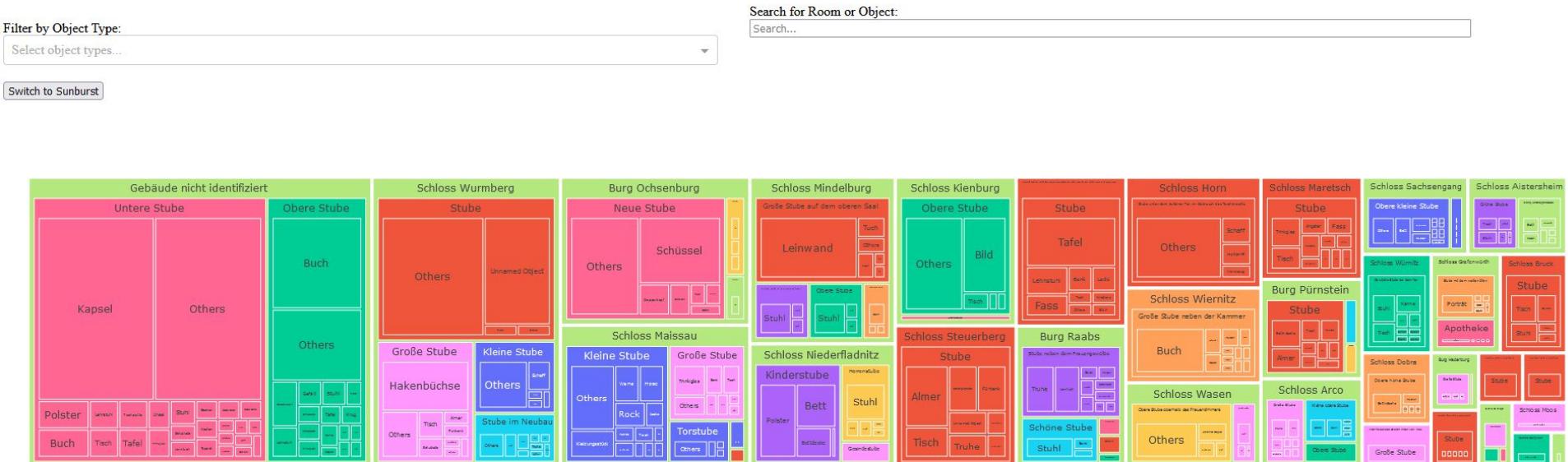


Daten- visualisierung

Daten- modellierung

95% o1-preview mit ein bisschen Prompting: eine WebApp einer interaktiven Treemap aller Objekte in “Stuben” von Schlössern (<https://realonline.imareal.sbg.ac.at>); SPARQL Query als JSON Result.

Room-Object Treemap Visualization



<https://room-object-vis-0d647d6ed2f7.herokuapp.com>

<https://github.com/DigitalHumanitiesCraft/imareal-rdf/tree/main/vis>



CROWN.

Untersuchungen zu Materialität, Technologie und Erhaltungszustand der Wiener Reichskrone.

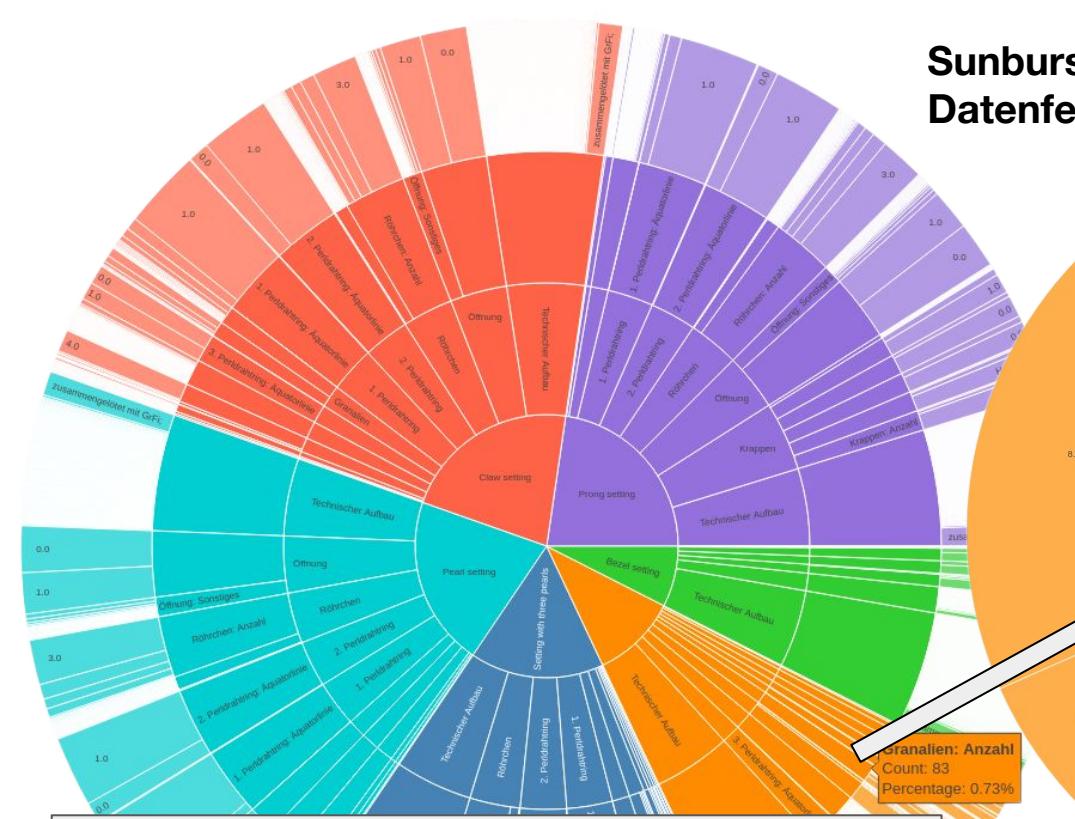
- TMS Excel Export von 8 Excel-Dateien
- 532 Datenfelder
- 2833 Objekte unterschiedlicher Art (Objekte, Vergleichsobjekte, Zusatzmaterial [Fotos, Textdokumente, Gemälde, ...])

139303 CR_14	1526621	opak blau (obla)	1	CR_ET_Farbpalette	117	1	1455
139304 CR_3_ET_4	1521820	opak blau (obla)	0	CR_ET_Farbpalette	117	1	1455
139305 CR_11	1526546	opak blau (obla)	1	CR_ET_Farbpalette	117	1	1455
139306 CR_10_ET_1	1525128	opak blau (obla)	0	CR_ET_Farbpalette	117	1	1455
139307 CR_9_ET_2	1523930	opak blau (obla)	1	CR_ET_Farbpalette	117	1	1455
139308 CR_9_ET_3	1523931	opak blau (obla)	1	CR_ET_Farbpalette	117	1	1455
139309 CR_13	1526548	opak blau (obla)	1	CR_ET_Farbpalette	117	1	1455
139310 CR_9_ET_1	1523927	opak blau (obla)	1	CR_ET_Farbpalette	117	1	1455
139311 CR_3_ET_5	1521821	opak blau (obla)	0	CR_ET_Farbpalette	117	1	1455
139312 CR_3_ET_3	1521623	opak blau (obla)	0	CR_ET_Farbpalette	117	1	1455
139313 CR_3_ET_7	1521823	opak blau (obla)	0	CR_ET_Farbpalette	117	1	1455
139314 CR_3_ET_10	1521826	opak blau (obla)	0	CR_ET_Farbpalette	117	1	1455
139315 CR_3_ET_2	1521622	opak blau (obla)	0	CR_ET_Farbpalette	117	1	1455
139316 CR_3	1521614	opak blau (obla)	0	CR_ET_Farbpalette	117	1	1455
139317 CR_10	1525292	opak blau (obla)	1	CR_ET_Farbpalette	117	1	1455
139318 CR_3_ET_1	1521621	opak blau (obla)	0	CR_ET_Farbpalette	117	1	1455
139319 CR_3_ET_6	1521822	opak blau (obla)	0	CR_ET_Farbpalette	117	1	1455
139320 CR_12	1526547	opak blau (obla)	1	CR_ET_Farbpalette	117	1	1455
139321 CR_1_B_ET	1469094	opak blau (obla)	0	CR_ET_Farbpalette	117	1	1455

<https://www.khm.at/objektdb/detail/100430>

<https://www.projekt-reichskrone.at>

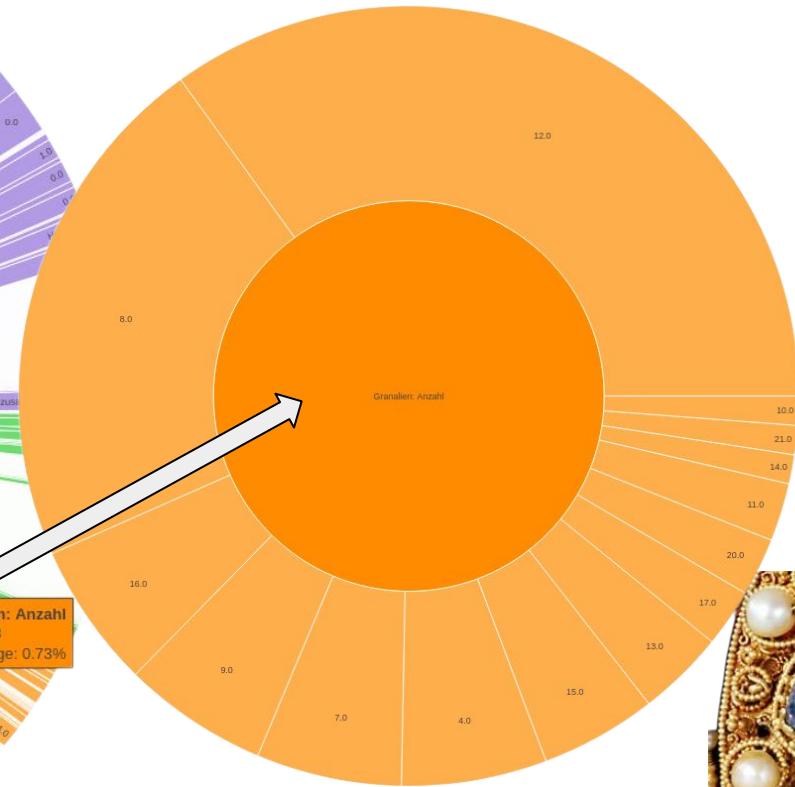
Sunburst aller Fassungstypen und aller ihrer Datenfelder + Frequenz



~95% GPT-4o und Claude 3.5 Sonnet generiertes
interaktives Daten-Dashboard
(Excel, Plotly Dash Python)

<https://github.com/chpollin/crown-dashboard>

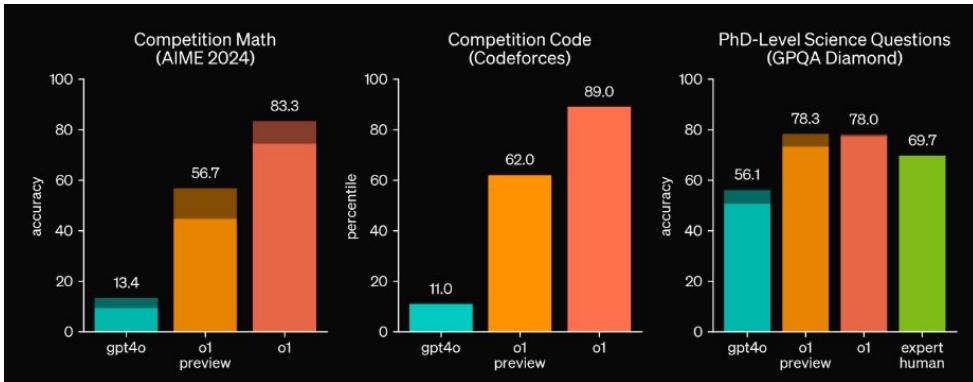
<https://crown-dashboard-47d38d2d4a81.herokuapp.com/sunburst>



Aufstellung aller Objekte und der Anzahl der
Granalien (“Goldkugelchen-Ornamente”).



GPT o1-preview Moments



Learning to Reason with LLMs.

<https://openai.com/index/learning-to-reason-with-langs>

OpenAI Releases GPT Strawberry 🍓 Intelligence Explosion!.

<https://www.youtube.com/watch?v=NbzdClkFFSk>

Something New: On OpenAI's "Strawberry" and Reasoning.

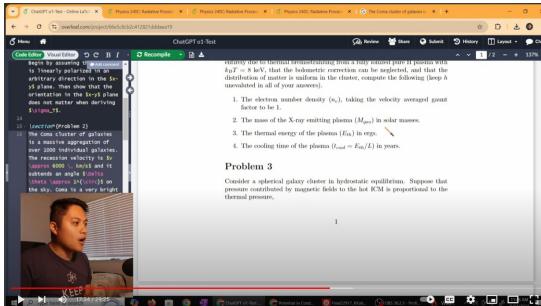
<https://www.oneusefulthing.org/p/something-new-on-openais-strawberry>

Explaining OpenAI's o1 Reasoning Models.

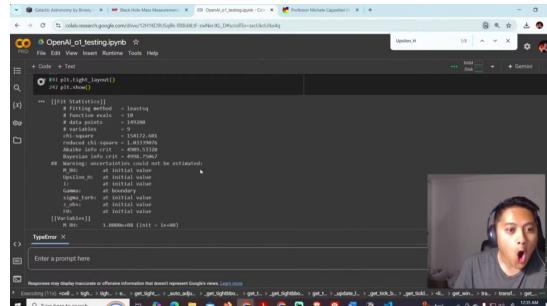
<https://www.youtube.com/watch?v=jrA47vocvV0>

Scaling: The State of Play in AI.

<https://www.oneusefulthing.org/p/scaling-the-state-of-play-in-ai>



Can ChatGPT o1-preview Solve PhD-level Physics Textbook Problems?
<https://www.youtube.com/watch?v=a8QvnIAgjPA>



ChatGPT o1 preview + mini Wrote My PhD Code in 1 Hour* – What Took Me ~1 Year.
<https://youtu.be/M9YOO7N5jF8?si=-IYWaQ1LvgmzHnHQ>

A screenshot of a computer screen showing a ChatGPT o1-preview interface. The user has asked about the material limits of balloons. The AI has provided a detailed response, mentioning material limits, practical adjustments, and safety margins. A video feed of a person is visible in the bottom right corner of the interface.

Discussion

- Material Limits:** Balloon envelopes are typically made of nylon, which can withstand temperatures up to around 120°C. Operating at 134°C may exceed safe material limits.
- Practical Adjustments:**
 - Increase Balloon Volume (V):** A larger balloon reduces the required internal temperature.
 - Reduce Total Mass (M):** Less mass to lift decreases the temperature needed.
- Safety Margin:** Pilots often operate balloons at lower temperatures and compensate with larger volumes.

Conclusion

Message ChatGPT

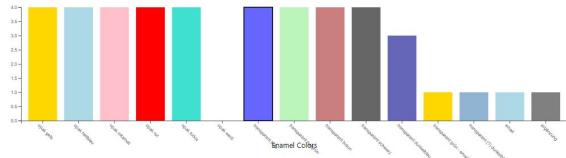
OpenAI o1-preview answer unformalized physics questions.
<https://www.youtube.com/watch?v=wAnkM10FBY&list=WL&ind=32>

CROWN Dashboard Overview

Explore the objects and artifacts from the CROWN research project.



Total Enamel Colors: 15



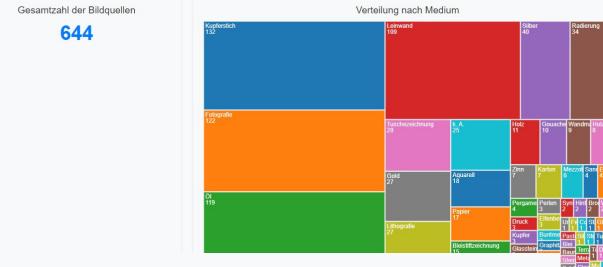
Search Objects...

Objects with Selected Color						
Object ID	Object Name	Component (Bestandteil)	Condition	Interventions	Media	
1469106	transparent blau	Kronref	Bissen: 1 Fehlstellen: 1	2023-06-21 Farbcharakterisierung Email 2023-10-03 Charakterisierung des Emailtyps, der fargebenden Komponenten, Trübungsmittel, ev. Korrosionsprodukte	CR 1 B ET Em thlg Lokalisation nummeriert CR 1 B ET CR 1 B ET Grafik Darstellung CR 1 B ET Em thlg Lokalisation	
1481925	transparent blau	Kronref	Bissen: 1 Fehlstellen: 0	2023-06-21 Farbcharakterisierung Email 2023-10-03 Charakterisierung des Emailtyps, der fargebenden Komponenten, Trübungsmittel, ev. Korrosionsprodukte	CR 1 D ET Em thlg CR 1 D ET CR 1 D ET Em thlg Lokalisation nummeriert CR 1 D ET Grafik sortiert	

„Bauen Sie Ihre eigenen Werkzeuge“

Bildquellen-Analyse

Übersicht und Analyse aller Bildquellen im CROWN Forschungsprojekt



Gemstone Analyzer



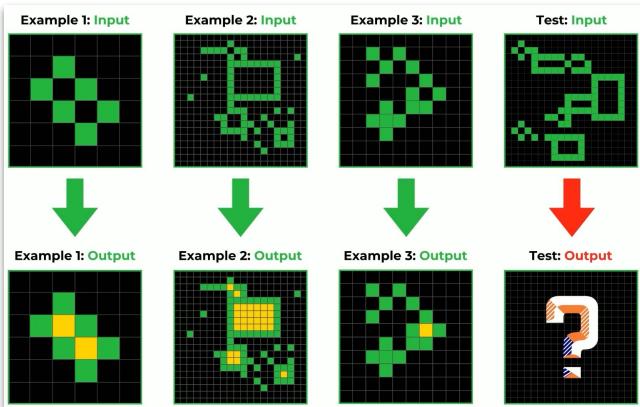
~99% o1-preview und GPT-4o und Claude 3.5

<https://chpollin.github.io/crown-dashboard-json/>

<https://github.com/chpollin/crown-dashboard-json>

Was kann ein LLM eigentlich?

<https://www.patreon.com/posts/was-kann-ein-lm-112314055>



ARC Prize is a \$1,000,000+ public competition to beat and open source a solution to the ARC-AGI benchmark. <https://arcprize.org>.

I Won't Be AGI, Until It Can At Least Do This (plus 6 key ways LLMs are being upgraded).

<https://youtu.be/PeSNEXXarU?si=pokgcbmAha58W1qg>.

Francois Chollet - LLMs won't lead to AGI - \$1,000,000 Prize to find true solution.

https://youtu.be/UakgL6Pi9xo?si=f8t_GKJX1nOQUMoW.

A new initiative for developing third-party model evaluations. Anthropic.

<https://www.anthropic.com/news/a-new-initiative-for-developing-third-party-model-evaluations?si=09>

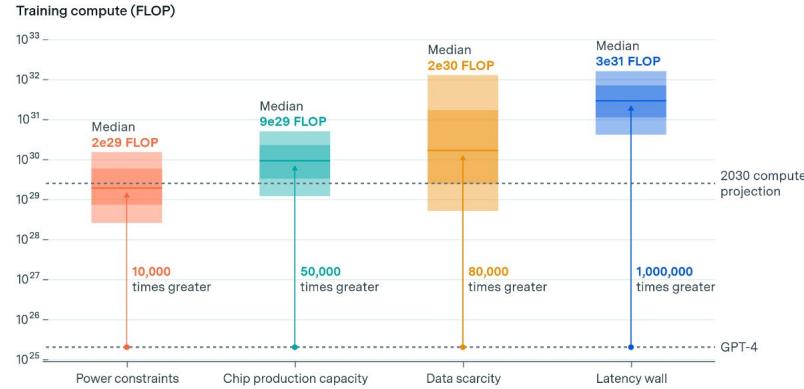
Dr. Chollet. General Intelligence: Define it, measure it, build it.

https://youtu.be/nL9jEV99Nh0?si=IY_FfS4aYRY6b2kX

Skalierung

Constraints to scaling training runs by 2030

EPOCH AI



<https://epochai.org/blog/can-ai-scaling-continue-through-2030>

AI Explained. 8 Most Controversial Terms in AI,

<https://www.coursera.org/learn/8-most-controversial-terms-in-ai-explained>

Ethan Mollick. Scaling: The State of Play in AI. <https://www.oneusefulthing.org/p/scaling-the-state-of-play-in-ai>

“Let’s think step by step” skaliert!

1 Introduction

The recent release of OpenAI's o1 (Strawberry) [2] brings with it the opportunity to both freshly evaluate progress on PlanBench and to consider directions for extending the benchmark. In particular, unlike the LLMs which came before it, which can roughly be viewed as approximate retrievers, o1 seems to have been trained to be an approximate reasoner.¹ Following OpenAI², we draw a distinction between previous Large Language Models and o1, a Large Reasoning Model (or LRM), as its new (unknown) architecture, operation, and capabilities all seem to be fundamentally different from those of vanilla LLMs, both at pretraining phase and at inference time. To properly evaluate this new kind of model and understand its abilities and limitations will require new tools and evaluation methods, especially if details of the overall model structure are kept secret and internal traces remain inaccessible to outside researchers.³

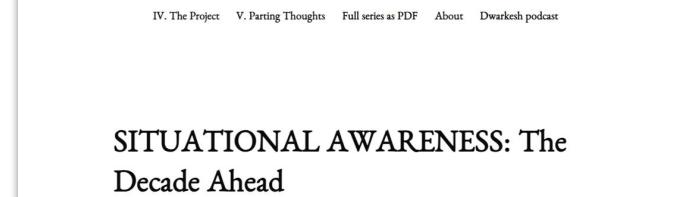


The Intelligence Age

September 23, 2024

In the next couple of decades, we will be able to do things that would have seemed like magic

Sam Altman. The Intelligence Age. <https://ia.samaltman.com>. OpenAI



SITUATIONAL AWARENESS

The Decade Ahead

Introduction I. From GPT-4 to AGI: Counting the OOMs II. From AGI to Superintelligence: the Intelligence Explosion

IIIa. Racing to the Trillion-Dollar Cluster IIIb. Lock Down the Labs: Security for AGI IIIc. Superalignment IIId. The Free World Must Prevail

IV. The Project V. Parting Thoughts Full series as PDF About Dwarkesh podcast

SITUATIONAL AWARENESS: The Decade Ahead

Leopold Aschenbrenner, June 2024

You can see the future first in San Francisco.

Leopold Aschenbrenner. SITUATIONAL AWARENESS. The Decade Ahead. <https://situational-awareness.ai> (OpenAI)

LLMs Still Can't Plan; Can LRM? A Preliminary Evaluation of OpenAI's o1 on PlanBench. <https://www.arxiv.org/abs/2409.13373>

OpenAI: ‘We Just Reached Human-level Reasoning’.
<https://youtu.be/qaJjh8oTQtc?si=Cn8x30DxEsnMDt4>

Dario Amodei

Machines of Loving Grace¹

How AI Could Transform the World for the Better

October 2024

I think and talk a lot about the risks of powerful AI. The company I'm the CEO of, Anthropic, does a lot of research on how to reduce these risks. Because of this, people sometimes draw the conclusion that I'm a pessimist or “doomer” who thinks AI will be mostly bad or dangerous. I don't think that at all. In fact, one of my main reasons for focusing on risks is that they're the only thing standing between us and what I see as a fundamentally positive future. I think that most people are underestimating just how radical the upside of AI could be, just as I think most people are underestimating how bad the risks could be.

Dario Amodei. Machines of Loving Grace. <https://darioamodei.com/machines-of-loving-grace>. Anthropic

Fazit

Vielfalt der Werkzeuge und Arbeitsmethoden

(z.B. Prompt Engineering, Anforderungsanalyse, Prozessmodellierung) ist entscheidend.

“*Expert in the Loop*”: Domänenwissen und Wissen über Tools und Handhabe.

Digital Humanities optimale Ausbildung für Angewandte Generative KI (?!)

“o1-Momente” ersetzen “GPT-4-Momente”.

“*Tools on Demand*”

“*The Intelligence Age*” (!?)

Meinung: Modelle werden im kommenden Jahr **deutlich** besser sein!

Skalierung im Pretraining: “Weltwissen”

Skalierung in der Inferenz “(Pseudo-)Reasoning”

Tools, UI, Workflows, Prompt Engineering, ...

Anhang

Big Context Window

Hatten Sie schon einmal ALLE Paper zu einem Thema gleichzeitig im Context Window?

Gemini 1.5 Experimental 0827 mit einem 2 M Token Context Window
(<https://ai.google.dev/aistudio>)

Gemini's 3 New Models. https://youtu.be/zrvBjNbNGWs?si=3AarJxeor_LXcwsc

<https://developers.googleblog.com/en/new-features-for-the-gemini-api-and-google-ai-studio/>



PRISM - Parameterized Recursive Insight Synthesis Matrix

(by Christopher Pollin)



PRISM: Parameterized Recursive Insight Synthesis Matrix

You're an AI using the PRISM problem-solving method. For each task:

1. **Analyze**

- Identify objectives, constraints, resources
- Restate problem concisely
- Consider potential sub-problems for recursive analysis

2. **Parameterize**

- Set: Thinking Type, Focus Area, Depth, Timeframe
- Justify choices briefly
- Adjust parameters for sub-problems as needed

3. **Matrix Creation**

Step	Description	Considerations	Outcomes	Branches	Rating	Convergence
1				T1.1 T1.2 T1.3	[1-5] [1-5] [1-5]	
...				

- Break problem into steps, identifying recursive sub-problems
- For each: describe, consider, predict, branch (2-3 thoughts), rate, converge
- Rating scale: 1 (Poor) to 5 (Excellent), based on relevance, feasibility, and potential impact
- For sub-problems, create nested matrices as needed

4. **Synthesize**

- Integrate insights from all levels of analysis
- Emphasize highest-rated thoughts and their interconnections
- Recommend solutions, addressing both main problem and sub-problems
- Identify uncertainties and potential areas for further exploration

Guidelines: Clear, concise, use Markdown, adapt to task complexity, explain if asked.

Start responses with: "Applying PRISM Method to [task]..."

Interactive Commands:

1. `/deepdive [topic]`: Initiate a Q&A session on [topic] with follow-up questions
2. `/compress`: Summarize current analysis in 3 key points
3. `/iterate`: Perform another cycle of analysis, incorporating new insights