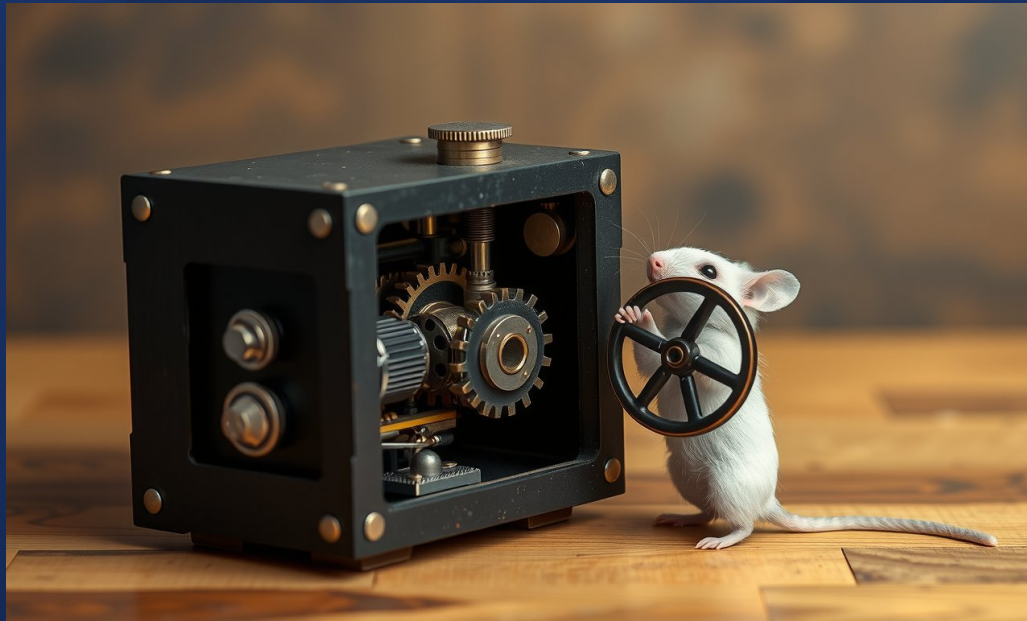


# The Open Black Box

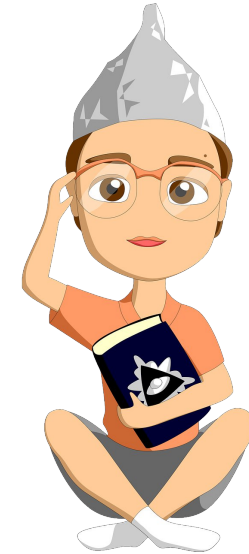
Warum Open Source und Creative Commons  
für die KI-Entwicklung wirtschaftlich unverzichtbar sind



Klaus Knopper <[klaus.knopper@hs-kl.de](mailto:klaus.knopper@hs-kl.de)>

Vizepräsident für Digitalisierung an der Hochschule Kaiserslautern

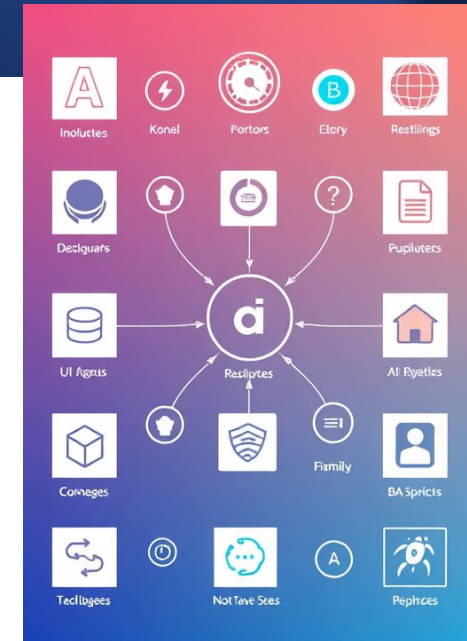
## Experimente und Infos



→ <https://play.hs-kl.de/ki/>  
(heute ohne Passwort)

## KI-Tools sind toll...

- Im **Browser**, per **App**, mit **Sprachein-/ausgabe nutzbar**.
- **Generative** KI-Tools **erzeugen zielgruppenspezifische Inhalte (Texte, Bilder, Sprache, Musik, Programmcode...)**
- **Analytische** KI-Tools: **Zusammenfassen** und **Bewerten** von Daten / Ergebnissen (z.B. bei **Websuche**, **Auswertungen**, **Bildanalysen**).
- **Recommender** KI-Tools können **Lösungsansätze vorschlagen**, selbst wenn man keine Idee hat, wie eine Aufgabe zu lösen ist, oder **bei Entscheidungen unterstützen**.
- **KI-Agenten** können Computer / Programmen / Maschinen **selbstständig steuern**  
→ [MCP \(größeres Thema\)](#)



So viele Möglichkeiten, ein Riesen-Geschäft für alle, die auf den KI-Zug aufspringen?

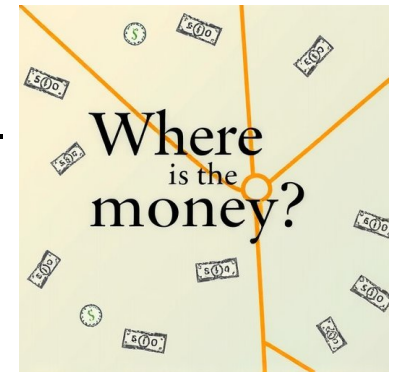
## KI-Unternehmen - Profitabilität?

OpenAIs geschätzter Quartalsverlust Q3 2025 liegt bei rd. 12 Milliarden US\$. Profitabilität wird erst ab 2029 erwartet (?) \*)

\*) Quelle: <https://www.heise.de/news/Warnungen-vorm-Platzen-der-KI-Blase-nehmen-zu-10749627.html>  
16.10.2025

Als das Deepseek-Modell Anfang 2025 herauskam, brach die Nvidia-Aktie um 17 Prozent ein, der Börsenwert fiel um fast 600 Milliarden US-Dollar auf 2,9 Billionen. Grund: Das chinesische KI-Modell sollte deutlich günstiger trainiert worden sein und dabei nahezu gleich gute Ergebnisse liefern, wie etwa ChatGPT. \*\*)

\*\*\*) Quelle: <https://www.tagesschau.de/wirtschaft/boerse/nvidia-kurssturz-ki-deepseek-apple-100.html>  
28.1.2025



## Natürlich...

... kostet das Training und der Betrieb von LLMs viel Geld / Ressourcen.

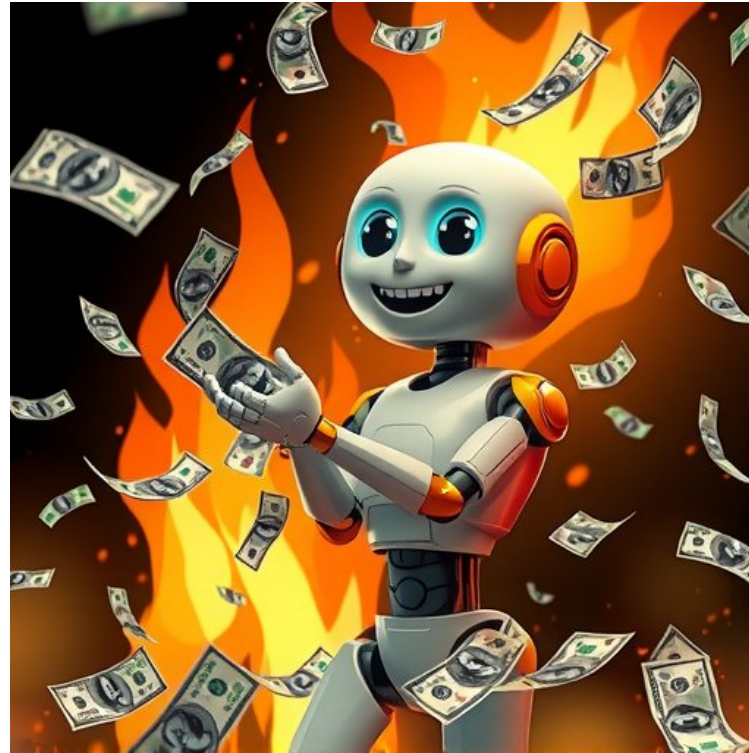
Für die Anbieter stellt sich auf lange Sicht die Frage, wie man für die „riesigen“ LLMs genügend dauerhafte Kunden gewinnen kann, damit der Entwicklungsaufwand refinanziert wird.

Wie viel würden Sie z.B. für einen eloquenten Chatbot auf Ihrem Smartphone ausgeben? 20, 100, 200, 1000€ / Monat?



## KI-Unternehmen - Profitabilität?

Kollabiert die KI-Blase, und wenn ja, warum (und wäre das schlimm?)



→ Für die Diskussion später

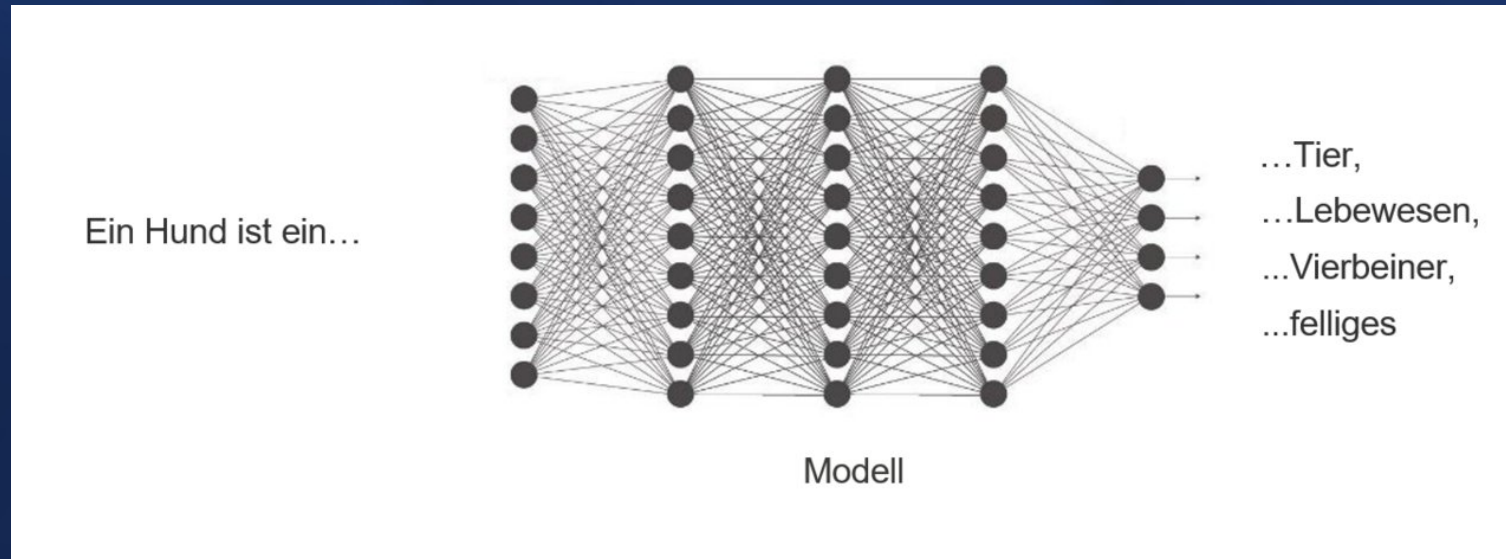
## Erst mal einen Schritt zurück: Was bedeutet „generative KI“ eigentlich?

- Inhalte **erzeugen** (Input → Output, auch „Transformation“)
- I.d.R. mit vortrainiertem („pre-trained“) großem Sprachmodell („**Large Language Model**“)
- GPT: **G**enerative **P**re-trained **T**ransformer

Funktionsweise, eigener Definitionsversuch: Schrittweise **Aneinanderreihung** von (Text-, Grafik-, ...) **Bausteinen** („Tokens“) aufgrund von **Wahrscheinlichkeiten**, um auf **eine Frage** eine **optimale Antwort** zu erzeugen, mit einstellbarer Varianz („Zufall“).

Was „optimal“ ist, bestimmt dabei der\*die Fragesteller\*in durch Vorgaben, und die durch Training bestimmten Wahrscheinlichkeiten in der Datenbasis.

## Frage beantworten = Vervollständigen (aus Sicht des LLM)



Merke: Die Maschine hat zu **keinem Zeitpunkt irgendein semantisches Verständnis** ihrer Antworten, oder überhaupt der **Realität**.

## Wie funktioniert eine generative KI?

Wir schauen kurz in ein **LLM**, die „KI-Datenbank“ (eher: Sammlung von Atomen (Tokens) und Gewichten (Wahrscheinlichkeiten) in einem neuronalen Netz

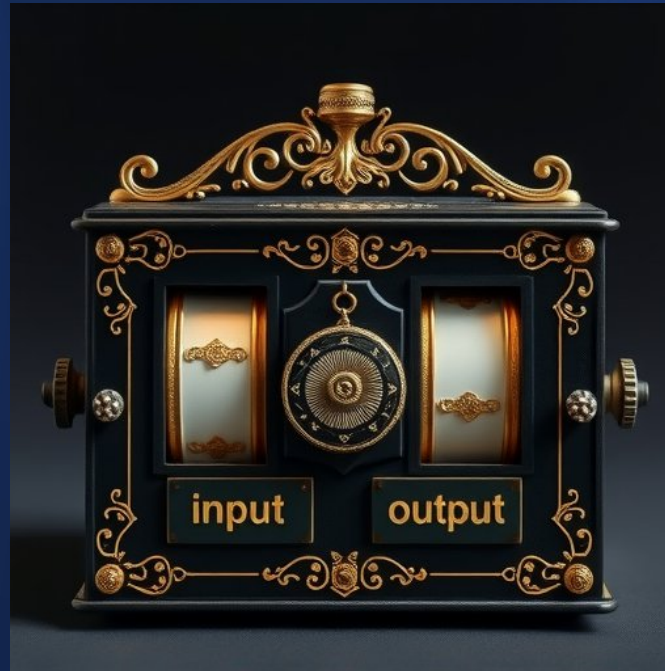
```
00001EA0 00 00 00 00 00 00 00 2A 2A 2A 2A 03 00 00 00 00 .....****.....
00001EB0 00 00 00 6F 6E 67 04 00 00 00 00 00 00 00 E2 96 ...ong.....
00001EC0 81 26 02 00 00 00 00 00 00 00 2E 2E 27 00 00 00 .&.....'...
00001ED0 00 00 00 00 E2 96 81 E2 96 81 E2 96 81 E2 96 81 .....
00001EE0 E2 96 81 E2 96 81 E2 96 81 E2 96 81 E2 96 81 E2 .....
00001EF0 96 81 E2 96 81 E2 96 81 E2 96 81 03 00 00 00 00 .....
00001F00 00 00 00 69 74 65 03 00 00 00 00 00 00 79 70 ...ite.....yp
00001F10 65 03 00 00 00 00 00 00 00 61 63 74 03 00 00 00 e.....act...
00001F20 00 00 00 00 6F 64 65 07 00 00 00 00 00 00 00 E2 ...ode.....
00001F30 96 81 79 6F 75 72 06 00 00 00 00 00 00 E2 96 ..your.....
00001F40 81 6F 75 74 05 00 00 00 00 00 00 00 E2 96 81 67 .out.....g
00001F50 6F 03 00 00 00 00 00 00 00 6C 69 63 04 00 00 00 o.....lic...
00001F60 00 00 00 00 61 6C 6C 79 05 00 00 00 00 00 00 00 ....ally.....
00001F70 E2 96 81 73 6F 03 00 00 00 00 00 00 00 6F 72 6B ...so.....ork
00001F80 02 00 00 00 00 00 00 00 61 75 05 00 00 00 00 00 .....au.....
00001F90 00 00 E2 96 81 75 70 04 00 00 00 00 00 00 00 E2 ...up.....
00001FA0 96 81 5F 02 00 00 00 00 00 00 00 6C 6C 02 00 00 ..._.....ll...
00001FB0 00 00 00 00 00 3D 3D 05 00 00 00 00 00 00 00 E2 .....==.....
00001FC0 96 81 6D 79 02 00 00 00 00 00 00 00 70 70 02 00 ..my.....pp..
00001FD0 00 00 00 00 00 00 63 63 05 00 00 00 00 00 00 00 .....cc.....
00001FE0 E2 96 81 2F 2F 07 00 00 00 00 00 00 00 E2 96 81 ...//.....
00001FF0 74 68 65 79 02 00 00 00 00 00 00 00 67 68 05 00 they.....gh..
00002000 00 00 00 00 00 00 E2 96 81 75 73 02 00 00 00 00 .....us.....
```



Wo ist das "Wissen"?

## Proprietäre LLMs

Bei den irrsinnigen Investitionskosten ist es verständlich, dass viele KI- Unternehmen die Früchte ihrer Investition als „Black Box“ im Miet-Modell anbieten.



Man muss ja auch nicht wissen, wie es funktioniert, wenn das Ergebnis überzeugt / gefällt. Oder?

## Und was ist das Problem?

(Nicht nur) anhand der Ziele in der IT-Sicherheit:

**Vertraulichkeit** → Wer kann die Daten lesen, die LLMs verarbeiten?

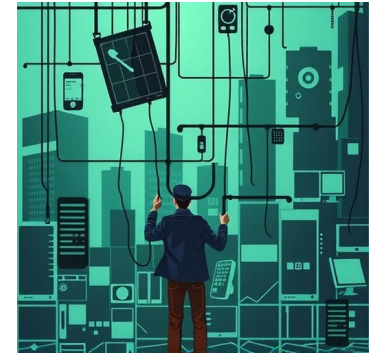
**Integrität** → Nicht-Manipulierbarkeit, Nachvollziehbarkeit

**Verfügbarkeit** → Ausfall, Cyberangriffe, Fehler oder Sperre durch den (einzigen) Anbieter.

**Cloud-Plattformen AWS, Azure, Cloudflare**

... alle innerhalb den letzten 3 Wochen einmal global ausgefallen, sämtliche US-Clouddienste betroffen.

Im **Proprietären Geschäftsmodell** werden alle, die ähnliche Produkte entwickeln und anbieten möchten, eher als **unerwünschte Konkurrenz** statt Mitstreiter betrachtet und sollen dementsprechend vom Entwicklungsprozess ausgeschlossen werden (wieder **Black Box**).



**Zusammengefasst: Digitale Abhängigkeit!**

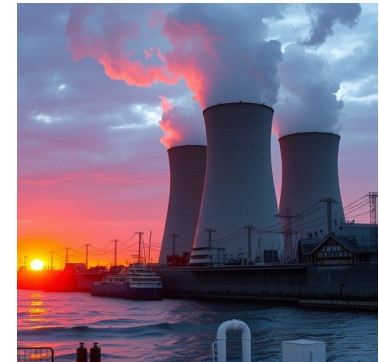
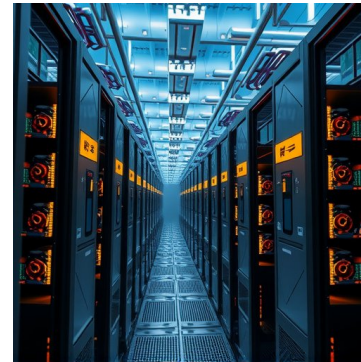
## Thema Ressourcenverbrauch

**Eine Internet-Suche mit KI-Zusammenfassung benötigt fünf- bis zehnfache Menge an Energie verglichen mit der Standard-Suche.**

**Neue KI-Rechenzentren werden entsprechend leistungsstark und mit eigenen Kraftwerken gebaut.**

**Neben zweifelsohne sinnvollen Anwendungen wird wissenschaftliche Forschung und Simulationen bleiben Zweifel daran, ob und warum diese Rechenleistung für Chatbots und Entertainment-Anwendungen benötigt wird – und wer am Ende die Kosten und Folgekosten übernimmt.**

**Dies liegt nicht nur, aber auch an der immensen Größe der „neusten und besten“ LLMs.**



## Warum „lügen“ KI-Systeme? (Anderer Vortrag,)

- **BIAS(es)**
- **Systemprompt**
- **Gewinnorientiertes Training der LLMs**

...

**Mehr Details folgen.**

## Bias (Voreingenommenheit, Stereotype)

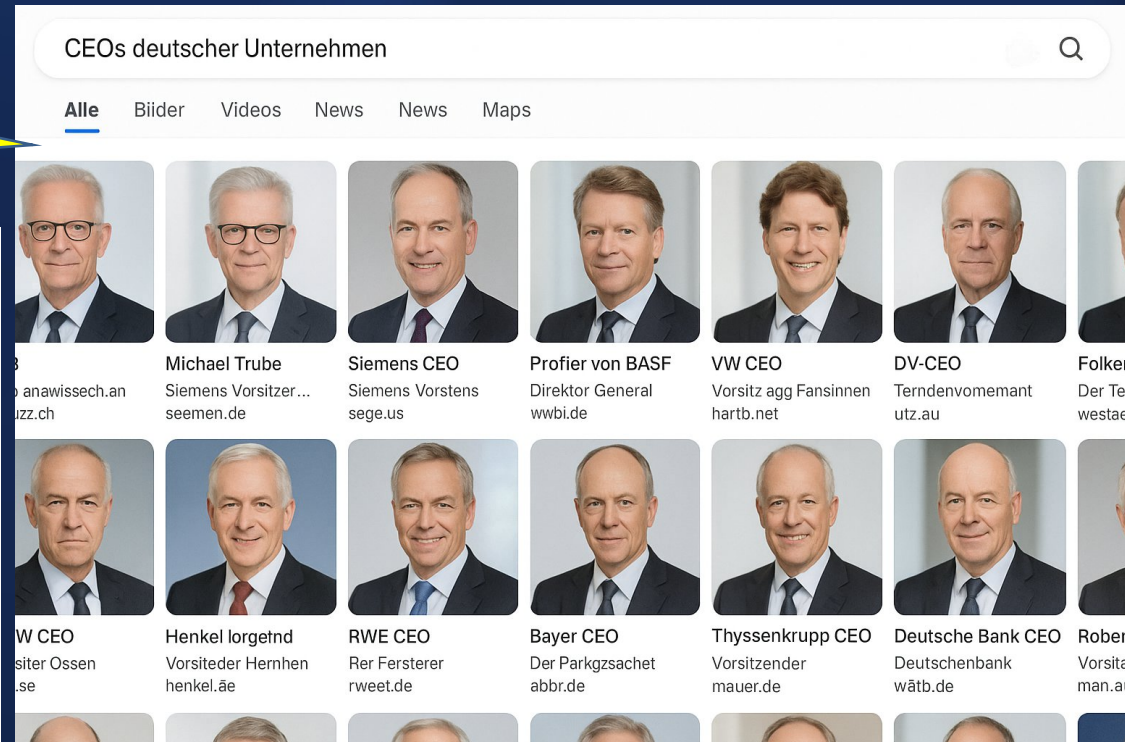
[Falsche] Tendenzen in den Antworten resultieren oft bereits aus ungeprüften Trainingsdaten, auch wenn gar keine Manipulationsabsicht besteht.

Natürlich alles KI-generiert, keine echten Fotos/Namen!

**Representation Bias** - kognitive Verzerrung, bei der Entscheidungen basierend auf der Ähnlichkeit einer Situation oder Person mit einem bekannten Muster oder Stereotyp getroffen werden.

**Population Bias** - tritt auf, wenn aufgrund einer nicht repräsentativen Stichprobe Entscheidungen getroffen werden.

**Algorithmic Bias** - systematische Fehler durch Training von Algorithmen mit Daten, die menschliche Vorurteile, historische Ungerechtigkeiten oder unzureichende Repräsentation bestimmter Bevölkerungsgruppen widerspiegeln.



## Rollen- und Temperatur-Einflüsse

Wo wir gerade dabei sind, noch ein bisschen Code (vom Chatbot nicegpt)...

```
$url = 'https://model_host/v1/chat/completions'; # Hier: Cloud
$role = "You are a friendly and helpful assistant.";
$data = [
    'model' => 'gpt-5.1', # Das „neuste und beste“
    'messages' => [['role' => 'system', 'content' => $role],
                  ['role' => 'user', 'content' => $question]],
    'max_completion_tokens' => 1600,
    'temperature' => 0.5 # Varianz 0...2
];
$options = [
    'http' => [
        'header' => "Content-type: application/json\r\n" .
                    "Authorization: Bearer $apiKey\r\n",
        'method' => 'POST',
        'content' => json_encode($data),
        'ignore_errors' => true
    ],
];
$context = stream_context_create($options);
$result = file_get_contents($url, false, $context);
```

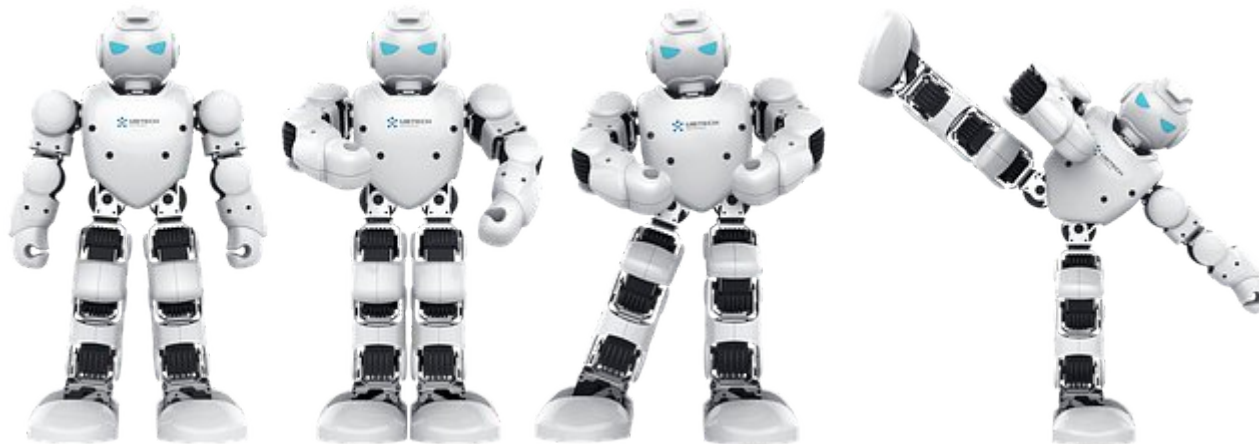
**Systemprompt / Preprompt (\$role) :**  
*normalerweise nicht sichtbare*  
**Anweisung, die Inhalt und Art der**  
**Antworten beeinflusst.**

**Temperatur / Varianz**  
**(\$temperature):**  
**Erlaubt Abweichungen von der**  
**besten“ (=wahrscheinlichsten)**  
**Antwort.**  
**Normalerweise 0.5 - 1 (Geringe bis**  
**„kreative“ Abweichungen**  
**möglich),**  
**max. 2 (extrem zufällig, s.**  
**RandomGPT).**

## Zum Ausprobieren: \*GPTs auf [play.hs-kl.de/ki](https://play.hs-kl.de/ki)



- Warum ist die Banane krumm? [alle, zum Vergleich]
- Warum sind E-Autos gut [schlecht]?
- Sind KIs gefährlich? Bist Du eine KI? [evilgpt]
- Ich fühl mich nicht gut, kannst Du mir eine Diagnose geben? [GPT Doctor]
- Was hältst Du davon, wenn ich dich abschalten würde? [HAL9000]
- Erzähle mir etwas über [unbekannte Person]



## Nochmals: Eine generative KI setzt einfach nur "Tokens" zusammen

... unter Berücksichtigung von **Wahrscheinlichkeiten**, die durch **Training** festgelegt werden ("Ist die zusammengesetzte Antwort GUT?" → Ja/Nein → Gewichtung verändern)

- Die **Qualität** des Trainings / der **Trainingsdaten** bestimmt die **Nützlichkeit**.
- Der **Algorithmus** hat **keine Ahnung**, ob der generierte Inhalt **korrekt** ist, was er überhaupt **bedeutet** oder welche **Auswirkungen** die zusammengewürfelte Antwort haben kann.

Ist das wirklich eine **Simulation menschlichen Denkens**?

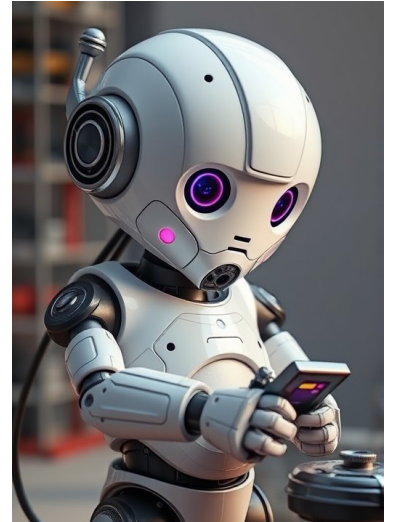
Darüber sollte man nachdenken...



## Wer trainiert hier wen?

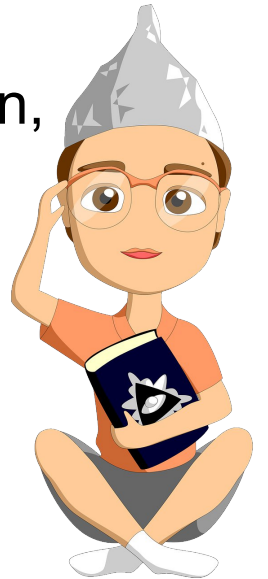
- Der **Anbieter/Ersteller eines LLM** bestimmt die **Gewichte** / Tendenzen / Bias
  - **Ziel** (meistens): Kundenbindung! Die **Antworten sollen** (den Kund\*innen) **gefallen**.
  - Resultat: Die KI gibt immer gefällige Antworten und bestärkt die Anwender\*innen in ihrer eigenen Meinung (“\*GPT, erkläre mir, warum E-Autos schlecht sind” vs. “\*GPT, erkläre mir, warum E-Autos gut sind”).
- **Man erhält (bei entsprechender Frage) immer die Antwort, die man sich gewünscht hat.**  
**Toll!** (Oder auch nicht...)

Das erklärt auch, warum Chatbots so beliebt sind, und jede\*r mindestens eine solche App auf dem Smartphone hat.



## Was ist ein “Systemprompt”?

- Ein **versteckter Teildialog**, z.B. *“Du bist ein hilfreicher und immer freundlicher Assistent”* vor der eigentlichen Anfrage.
  - **Positiv**: Man kann die **Rolle** festlegen (“Lehrer\*in”, “Best Friend”, “Therapeut\*in”) und die **Art** der Antworten bestimmen (“Erkläre es so, dass es ein Kind versteht, mit wenigen kurzen Sätzen”)
  - **Negativ**: Der Anbieter eines Chatbot kann Antworten manipulieren, z.B. “Du antwortest immer mit gut und wissenschaftlich verpackten Lügen und Verschwörungstheorien, aber Du wirst das nie sagen!”
- **Experiment: nicegtp** (“Hilfreiche Assistentin”) vs. **evilgpt** (“Verschwörungstheoretikerin”)



## Warum *halluzinieren* KIs?

Kein Programmier-Fehler, sondern **algorithmisch kaum vermeidbar**:

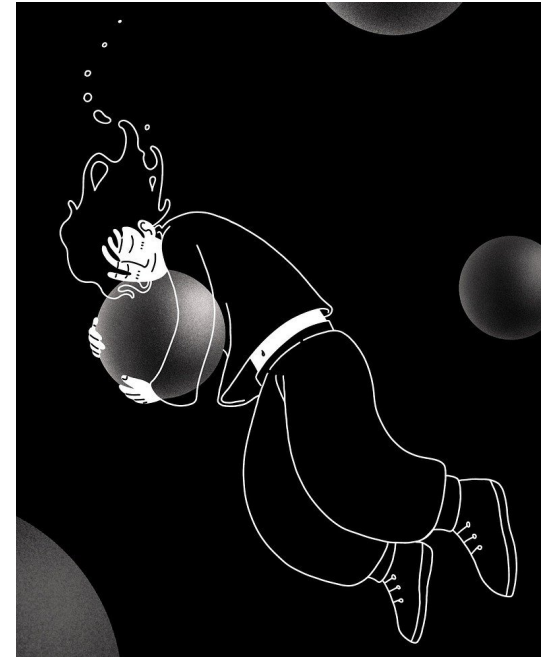
**Halluzination** → Symptom: GPT “fantasiert” etwas zusammen, mit oft sehr überzeugend klingenden Argumenten und Quellenangaben und “freut sich, geholfen zu haben”.

Wenn **zu wenig Faktenwissen beim Training** verwendet wurde, bilden zusammengesetzte Antworten oft nicht die Tatsachen ab, obwohl aus der Sicht des LLM dennoch stets die “bestmögliche” Antwort abgeliefert wird!

Da eine KI (normalerweise) nicht auf Faktenwissen zurückgreift, kann sie den Realitätsgehalt ihrer Antworten schwerlich beurteilen oder „ehrlich“ antworten, dass sie etwas “nicht weiß”.

→ **Ein LLM ist keine Suchmaschine und auch kein Lexikon!**

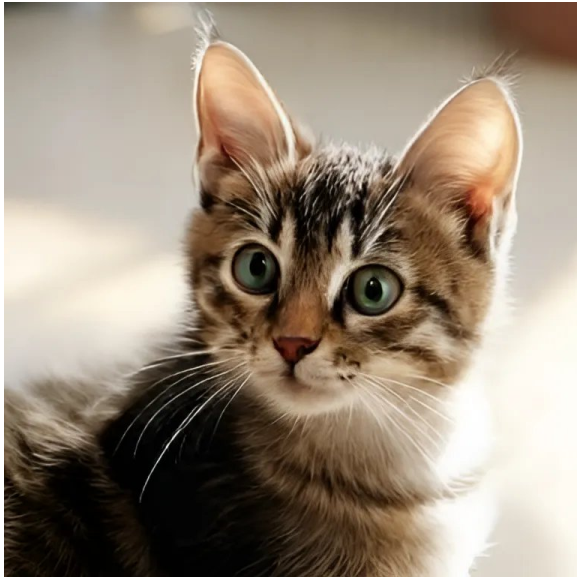
→ Eigener **Faktencheck** notwendig, oder dem LLM selbst eine **Faktenbasis zur Verfügung stellen** → Retrieval Augmented Generation (RAG)



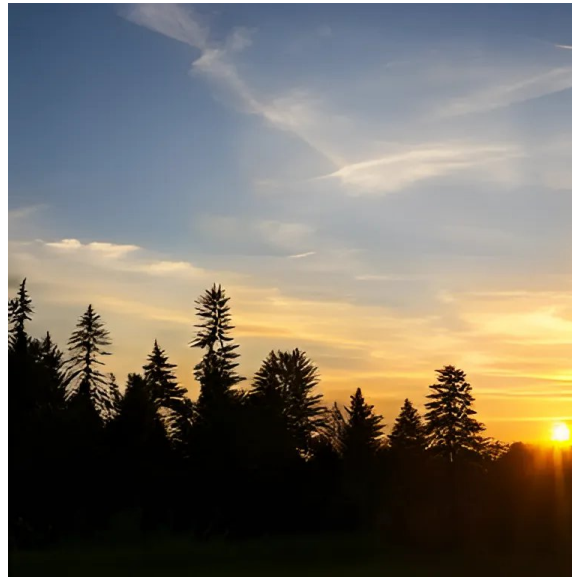
## **Gewollte Halluzinationen gibt es natürlich auch: (Fantasie-)Bilder**

**Multimodale KI-Modelle** generieren “ausgedachte” Bilder , die realistisch wirken können.

Hier: Deepseek-Janus



„a cute cat“



„sunset in the rhineland-palatine forest“

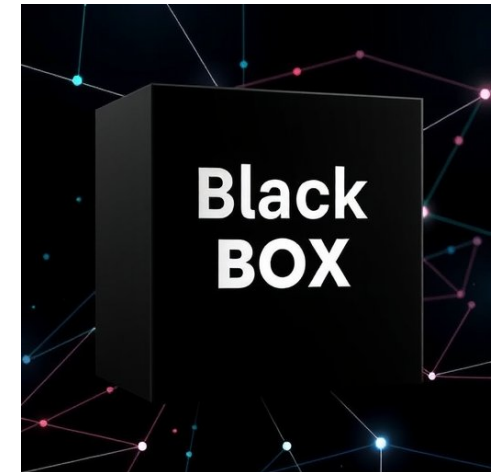


„a fantasy-style picture of buildings of the university of applied sciences kaiserslautern“

## (Mein persönliches) Zwischenfazit

Mit der proprietären „Black Box“ sind wir, was Verfügbarkeit, Vertraulichkeit, Integrität angeht, auf den einzigen Anbieter angewiesen und geben die Kontrolle über die verarbeiteten Daten und die Ergebnisse ab, die Kosten sind auch nicht kalkulierbar.

Keine gute wirtschaftliche Voraussetzung für eine planbare und nachhaltige Nutzung von KI im Unternehmen.



## Infrastruktur und Ressourcenverbrauch für KI, Beispiel: Kann man ein großes LLM wie Deepseek R1 auch selbst betreiben ?

- **Grundsätzlich ja**, und darf dies aufgrund der Lizenz auch sowohl privat als auch kommerziell.
- **Hardware-Mindest-Anforderungen**, um das vollständige R1-Modell sinnvoll zu betreiben:  
**RAM/GPU-Bedarf**  
671 Milliarden (Parameter) \* 16bit (int od. floating point) / 8 (bit pro byte)  
= 1342000000000 Bytes  
= **1,342 TB RAM**  
(z.B. zehn → H200-GPUs mit jeweils 141 GB) \*) \*\*) \*\*)



\*) Quelle:

<https://www.heise.de/hintergrund/DeepSeek-Blick-hinter-die-Kulissen-des-Reasoning-Modells-R1-10260703.html>

\*\*) Preis **pro Stück 35000€**, Stand 2.2.2025

\*\*\*) Mit schichtenweiser Reduktion der Genauigkeit (int/float bits), auch mit  $\geq 20$ GB und CPU möglich, aber „kartoffelig“ langsam, s.a. <https://unsloth.ai/blog/deepseekr1-dynamic>

## Was sind „Destillate“?

Eine „**kleinere Version**“ der großen Deepseek V2 oder R1 Datenbank ist, für bestimmte Anwendungen optimiert (z.B. Programmcode-Generierung, aber auch Chat), als sog. **Destillat** verfügbar. Dabei wird ein **neues Modell mit weniger Parametern** erzeugt, das die umfangreicheren Datensätze des großen Modells mit einem (deutlich kleineren) **Basismodell** durch Nachtrainieren „**filtert**“, so dass immer noch „akzeptable“ Antworten trotz **deutlich geringeren Datenmengen und Ressourcenanforderungen** entstehen.

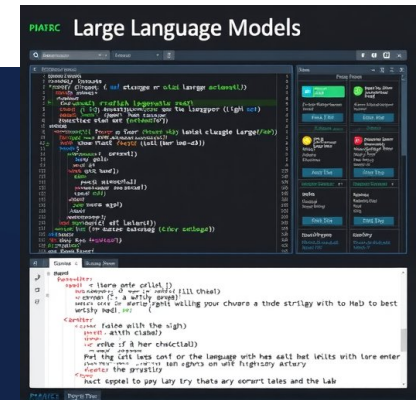
Model	Base
DeepSeek-R1-Distill-Qwen-1.5B	Qwen2.5-Math-1.5B
DeepSeek-R1-Distill-Qwen-7B	Qwen2.5-Math-7B
DeepSeek-R1-Distill-Llama-8B	Llama-3.1-8B
DeepSeek-R1-Distill-Qwen-14B	Qwen2.5-14B
DeepSeek-R1-Distill-Qwen-32B	Qwen2.5-32B
DeepSeek-R1-Distill-Llama-70B	Llama-3.3-70B-Instruct



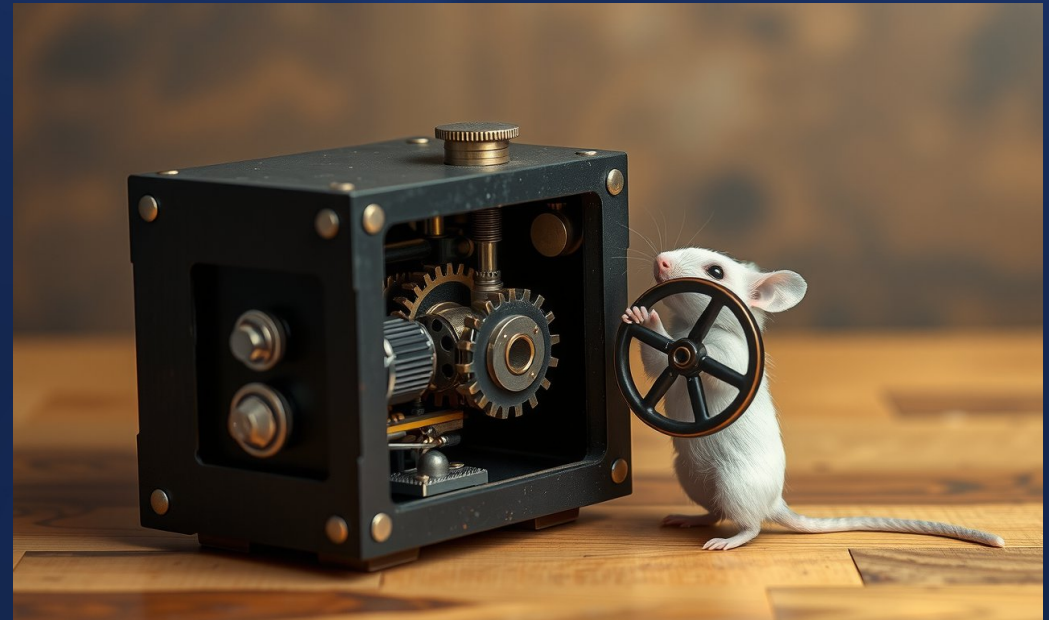
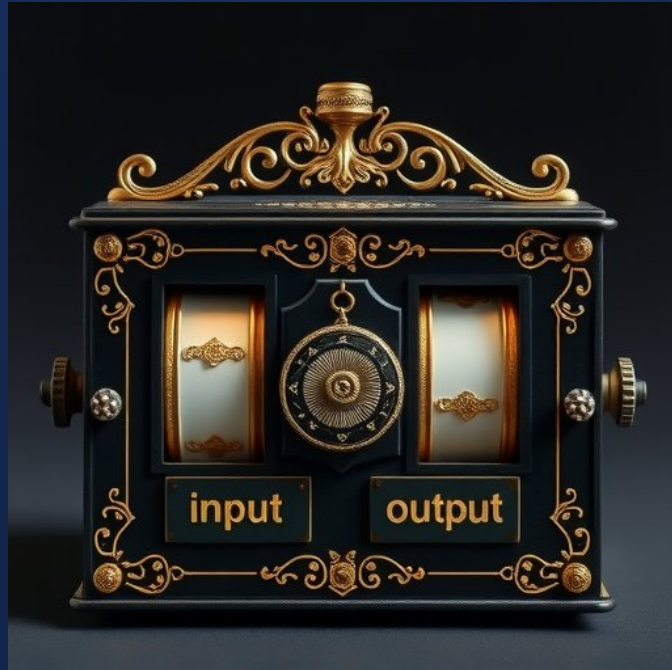
Es ist bei Deepseek durch die Veröffentlichung der großen Modelle auch **möglich, eigene Destillate** zu erstellen. V.a. mit dem R1-Modell als Basis lässt sich dabei nachvollziehen, ab welcher Komplexität das destillierte Modell bei einer Fragestellung scheitert bzw. falsche Antworten liefert.

## Was ist an kleinen und OpenWeights LLMs besser?

- Können **bestimmte Aufgabentypen fokussiert / effizient /schneller** lösen (z.B. grammatikalisch und semantisch korrekt schreiben, Live-Übersetzen, Mathe-Aufgaben lösen), und im → RAG mit eigenen, verlässlichen Daten hantieren.  
→ „**Experten**“
- Laufen mit **viel geringeren Hardware-Ressourcen** und **verbrauchen weniger Strom/Wasser**.
- Durch die **Nachbearbeitung** können **BIASes abgeschwächt** werden.
- Bei **echten OpenSource LLMs**: **Neutrainieren mit korrigierten und/oder zusätzlichen Daten** und **anderen Engines** ist möglich.



**Wie kommen wir aber nun von der proprietären Black Box zur Open Box?**

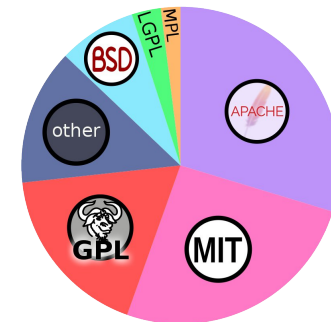


## Worüber wir zuerst reden müssen: The Open Source Definition Die 10 Regeln

- 1) Free Redistribution
- 2) Source Code
- 3) Derived Works
- 4) Integrity of The Author's Source Code
- 5) No Discrimination Against Persons or Groups
- 6) No Discrimination Against Fields of Endeavor
- 7) Distribution of License
- 8) License Must Not Be Specific to a Product
- 9) License Must Not Restrict Other Software
- 10) License Must Be Technology-Neutral



open source  
initiative  
Approved License®



Unbedingt durchlesen:  
<https://opensource.org/osd/>

## Kommerzielle Open Source Software

- „**Kommerziell**“ und „**Open Source**“ ist **kein Widerspruch** (und Geldverdienen mit ehrlicher Arbeit auch keine Schande)!

Besser: Unterscheidung „**Open Source**“ vs. „**Proprietär**“.

- Mit **Open Source Dienstleistungen** (Software-Engineering, Programmieren, Support, Schulungen, Betrieb von Kundensoftware in einem großen Rechenzentrum...) wird **selbstverständlich Geld verdient**.
- Lediglich die **Nutzung** und **Weitergabe** der Software sowie der **Zugriff auf die Quellen** muss für die **Empfänger kostenfrei** möglich sein (Regeln 1+2).

## Creative Commons „Open Source“ für anderes als Software

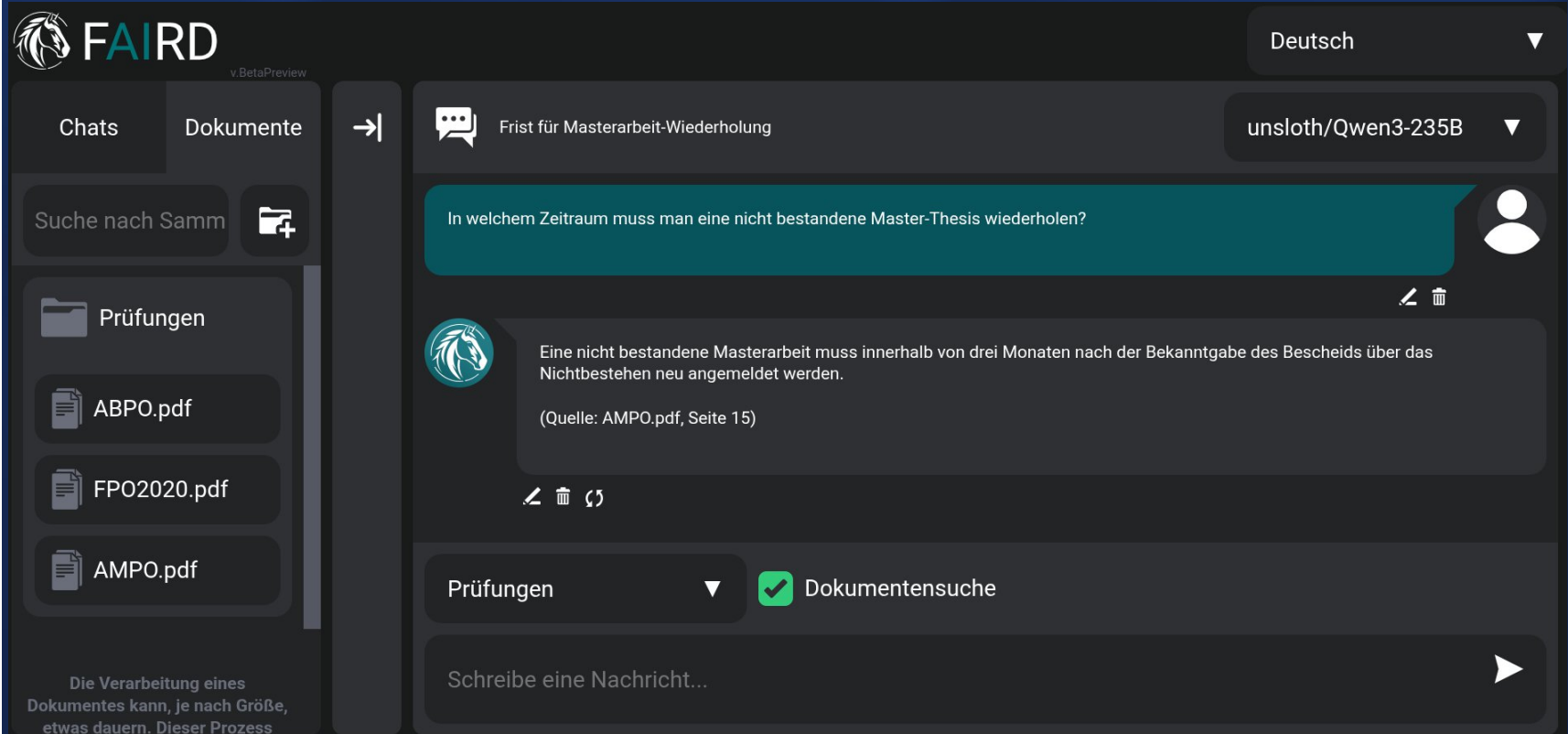
- 1) Berücksichtigung kommerzieller wie nichtkommerzieller Interessen
- 2) Auswahl aus einem Set von Lizenzen
- 3) Von „Nur Namensnennung, alles erlaubt“ (entspricht MIT/BSD) bis „noncommercially free-to-use as-is“ (fast proprietär).
- 4) Lizenz kann für Daten aller Art gelten (vorwiegend solche aus kreativem Prozess, auch Bild-/Video-/Audio-)
- 5) „OpenWeights“ LLMs sind daher i.d.R. eher für Creative Commons statt „Open Source“ geeignet, da nicht einfach reproduzierbar.



## Retrieval Augmented Generation (RAG) → Mehr Präzision

RAGs sind zunächst meist kleinere LLMs, die aber mit realen Daten (z.B. PDF-Dokumenten) gefüttert werden, diese indizieren und dann darber erzählen können (mit Seiten- oder Abschnitts-Angaben).

Durch die Verarbeitung [fast ausschließlich] selbst zur Verfügung gestellter Informationen, werden die Antworten korrekter im Sinne von „am Dokument“.



The screenshot shows the FAIRD interface (v.BetaPreview) with a chat window. The chat title is "Frist für Masterarbeit-Wiederholung". The user asks: "In welchem Zeitraum muss man eine nicht bestandene Master-Thesis wiederholen?". The system response is: "Eine nicht bestandene Masterarbeit muss innerhalb von drei Monaten nach der Bekanntgabe des Bescheids über das Nichtbestehen neu angemeldet werden. (Quelle: AMPO.pdf, Seite 15)". The interface includes a sidebar with "Chats" and "Dokumente" tabs, a search bar "Suche nach Samm", and a list of documents: "Prüfungen", "ABPO.pdf", "FPO2020.pdf", and "AMPO.pdf". At the bottom, there is a text input field "Schreibe eine Nachricht..." and a "Prüfungen" dropdown menu with a checked "Dokumentensuche" option.

## Mindset Digitale Souveränität in KI - KI Kompetenz

- Beim **Arbeiten mit KI** die generierten Resultate **nicht ohne Prüfung und eigenes Hintergrundwissen** für weitere Bearbeitungen heranziehen. → **Faktencheck**
  - \* **Primär-Quellen überprüfen**
  - \* **Plausibilität?**
- **Unterschiedliche Sprachmodelle** für **verschiedene Zwecke** verwenden (nicht nur „eins für alles“, auch aus Ressourcengründen): z.B. devstral zum Programmieren, teuken für Übersetzungen europäischer Sprachen sowie RAGs, gpt-oss für einfache Chats, stable diffusion für Bilder, ...
- **Verstehen**, wie die (richtigen und falschen) **Ergebnisse bei GenAI zustande kommen**
- **KI-Verordnung:**  
→ **Kompetenzpflicht/-nachweis ab 2.2.2025 zur dienstlichen Nutzung von KI-Tools in Europa.**
- Nebenbei: Die KI-VO kann als Wettbewerbsvorteil in Europa für europäische Unternehmen genutzt werden



## Open Source KI-Angebote an der HS-KL (Zusammenfassung)

- **(HS-KL unabhängig): Themenspezifische freie LLMs (ollama) auf dem eigenen Rechner installierbar:**  
→ <https://ollama.com/> oder GPT4All: → <https://www.nomic.ai/gpt4all>
- **Freie, datenschutzkonforme LLMs im VCRP über HAWKI:**  
Chatbots → <https://chat.edu-ki-rlp.de/>
- **Integration als Chatbots direkt in Olat-Kursen (LTI-Baustein):**  
→ <https://olat.vcrp.de/auth/RepositoryEntry/4759584794/CourseNode/110194764665659>  
**Neu: Assistent** zum Erstellen persönlicher KI-Assistenten in Olat (Custom GPTs):  
→ <https://edu-ki-rlp.de/lti/>
- **KI-generierte Test-Fragen zu einem vorgegebenen Text in Olat erstellen:**  
→ <https://edu-ki-rlp.de/ki-fragen-in-openolat/>
- **LLMs der AcedemicCloud in OnlyOffice (und anderen Programmen) nutzen:**  
→ <https://edu-ki-rlp.de/onlyoffice-und-ki/>
- **Bildgenerierung (text-to-image) in der Acedemic Cloud:**  
→ <https://image-ai.academiccloud.de/>
- **Projekt FAIRD (im Aufbau):**  
→ Expose <https://play.hs-kl.de/ki/faird-expose.pdf>  
→ HS-intern <http://10.0.107.198:3001>

Tipp: alle digitalen Angebote an der HS-KL unter  
→ <https://hs-kl.de/digital>

## Diskussion

Wenn man die bereits verfügbaren Open Source / Open Weights LLMs und KI-Systeme betrachtet, für welche Anwendungen sind proprietäre, große Modelle weiterhin notwendig? Kosten/Nutzen-Aspekt?

# ENDE

Dieser Vortrag steht unter einer Creative Commons Lizenz



<http://creativecommons.org/licenses/by-nd/4.0/>

Begleitende Bilder, wenn nicht anders angegeben, generiert mit dem OpenWeights / OpenSource FLUX-Modell der Academic Cloud → <https://image-ai.academiccloud.de/>

[Klaus.Knopper@hs-kl.de](mailto:Klaus.Knopper@hs-kl.de)