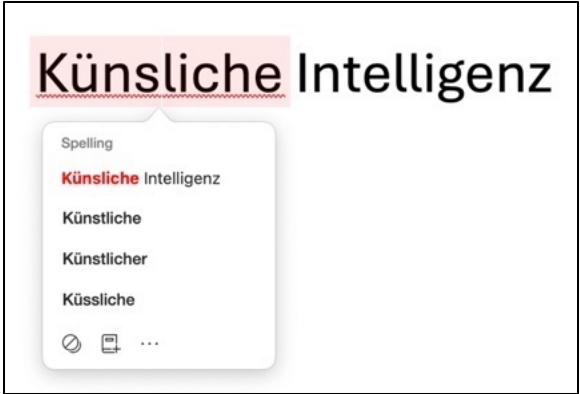


Die Lange Nacht der Wissenschaften
#NdW25

ChatGPT, Claude & Co. – Alles, was man (nicht) wissen muss

Prof. Dr. Paulina Pesch





Rechtschreibkontrolle in MS Word



Face ID



Siri



Recommended for you

Empfehlungen auf Streamingplattformen oder in Online-Shops



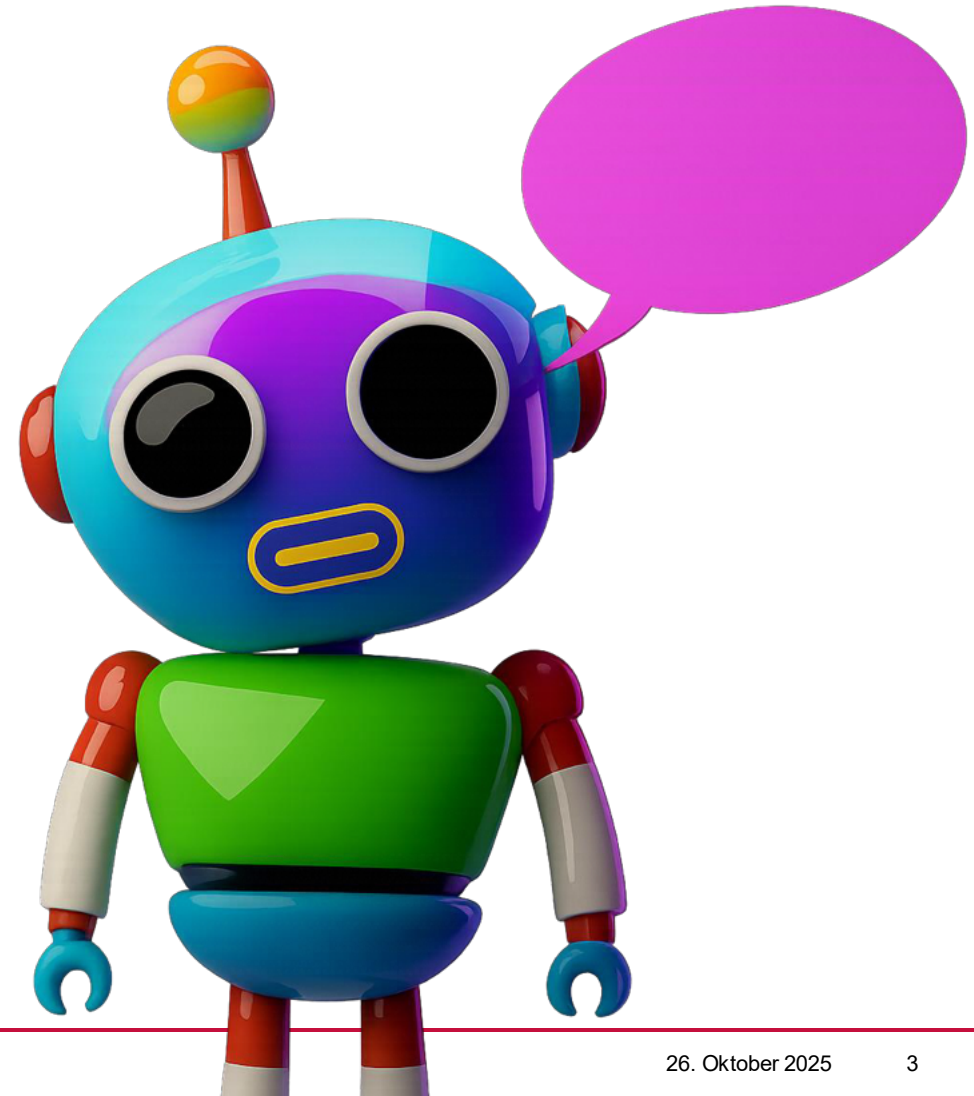
Cybertruck, Mr. Chopper, [CC BY-SA 3.0](#)



(Generative) Large Language Models



(Generative) Large Language Models



Was könnte ich bei der langen Nacht der Wissenschaften heute Abend über Chatbots erzählen?

✦ Show thinking ▾

Das ist ein top-aktuelles und spannendes Thema für die Lange Nacht der Wissenschaften! Das Publikum ist typischerweise neugierig, aber nicht unbedingt fachlich tief im Thema.

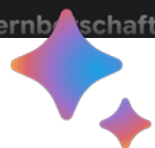
Hier ist ein Vorschlag, wie Sie einen spannenden Bogen schlagen könnten – von den Grundlagen bis zu den großen Zukunftsfragen.

Mögliche Gliederung für einen Vortrag oder eine Diskussionsrunde:

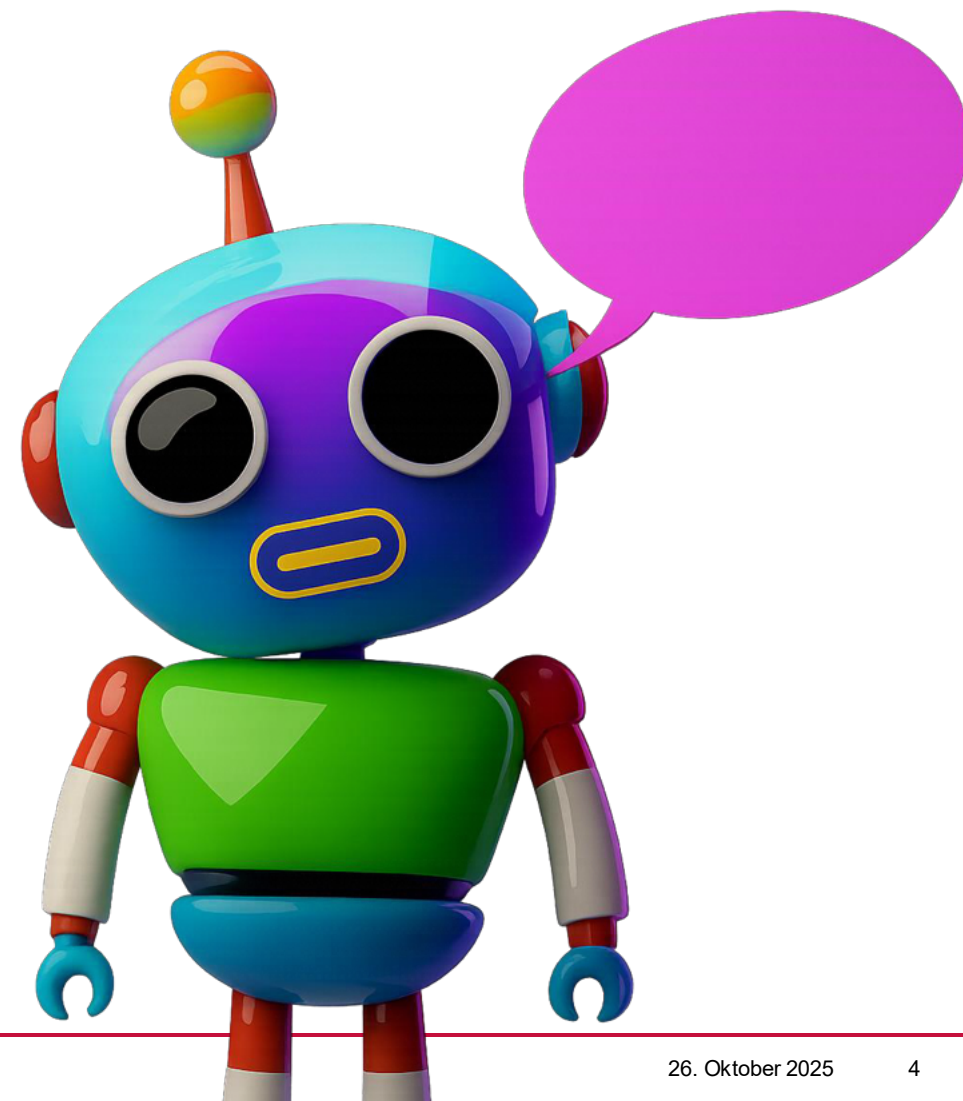
1. Der "Hook": Wer hat heute schon mit einem Bot gesprochen?

Starten Sie interaktiv. Fragen Sie, wer Siri, Alexa, Google Assistant genutzt oder mit einem Kundenservice-Bot gechattet hat. Das holt jeden ab.

Kernbotschaft: Chatbots sind längst Teil unseres Alltags. Aber wir erleben gerade eine



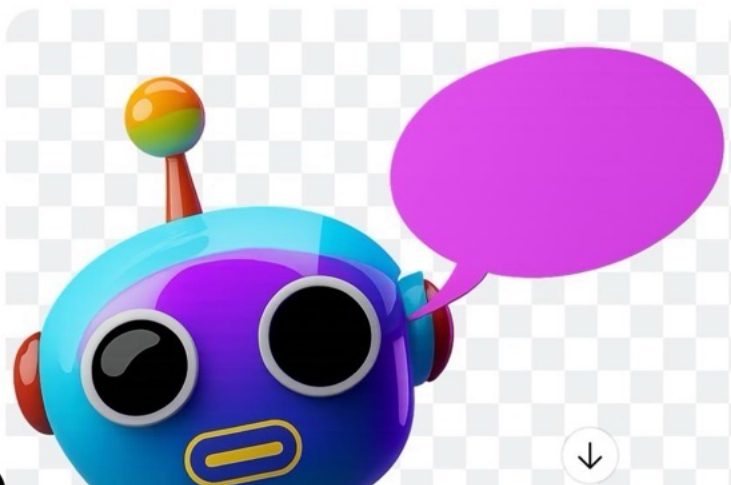
Konversation mit Gemini heute Mittag



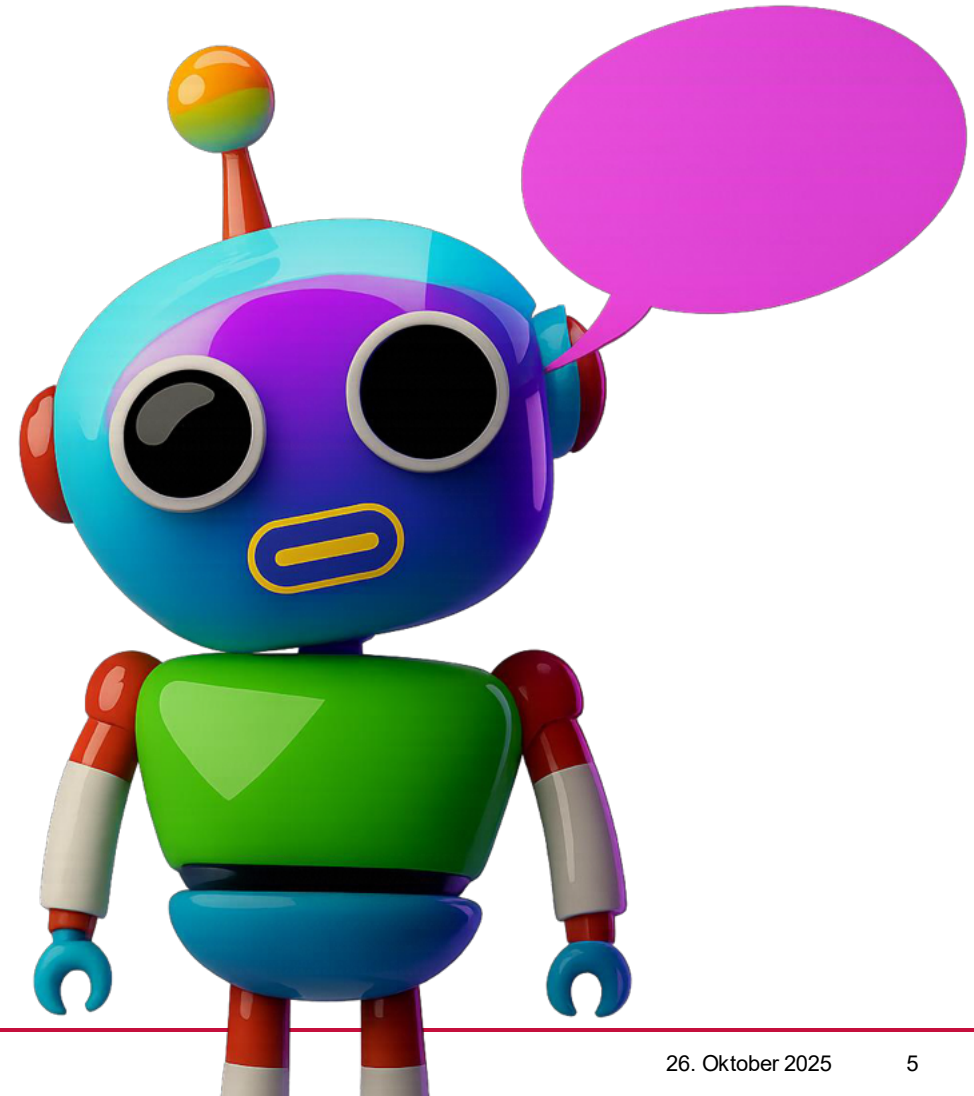
re-generate the image above (or, if not possible, a similar one) in higher resolution and with transparent background. Make sure the character and the speech bubble are fully on the image, not cut off. Only the legs should be cut off like in the image above.



Image created >



Paulinas Vortrag und ChatGPTs Beitrag

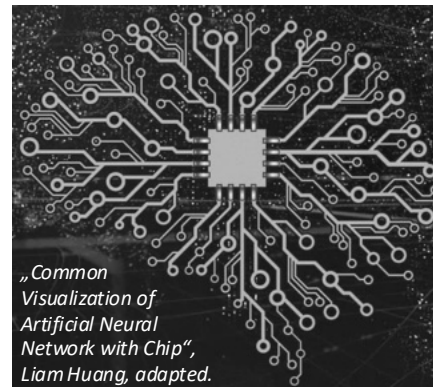


Das Training generativer Large Language Models

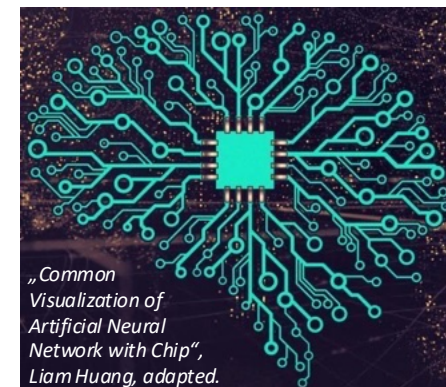
mit Texten



**Große Menge an
Trainingsdaten**



**Künstliches
neuronales Netz**

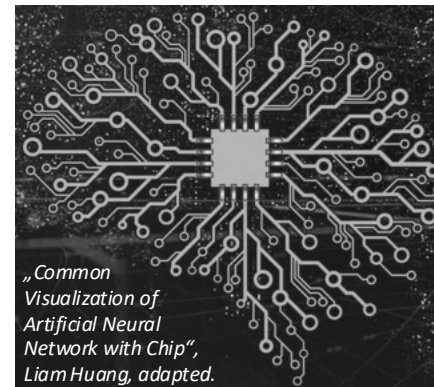


Trainiertes Modell

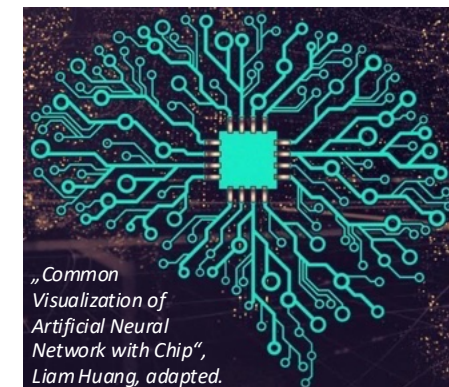
Das Training generativer Large Language Models mit Texten (inkl. Softwarecode)

```
85 };
86
87 static DEFINE_PER_CPU(struct cpuhp_cpu_state, cpuhp_state) = {
88     .fail = CPUHP_INVALID,
89 };
90
91 #ifdef CONFIG_SMP
92 cpumask_t cpus_booted_once_mask;
93 #endif
94
95 #if defined(CONFIG_LOCKDEP) && defined(CONFIG_SMP)
96 static struct lockdep_map cpuhp_state_up_map =
97     STATIC_LOCKDEP_MAP_INIT("cpuhp_state-up", &cpuhp_state_up_map);
98 static struct lockdep_map cpuhp_state_down_map =
99     STATIC_LOCKDEP_MAP_INIT("cpuhp_state-down", &cpuhp_state_down_map);
100
101 static inline void cpuhp_lock_acquire(bool bringup)
102 {
103     lock_map_acquire(bringup ? &cpuhp_state_up_map : &cpuhp_state_down_map);
104 }
105
106 static inline void cpuhp_lock_release(bool bringup)
107 {
108     lock_map_release(bringup ? &cpuhp_state_up_map : &cpuhp_state_down_map);
109 }
110
111 #else
112 static inline void cpuhp_lock_acquire(bool bringup) {}
113 static inline void cpuhp_lock_release(bool bringup) {}
114 #endif
115
116 #endif
117
118 /**
119  * struct cpuhp_step - Hotplug state machine step
120  * @name: Name of the step
121  * @startup: Startup function of the step
122  * @teardown: Teardown function of the step
123  * @cant_stop: Bringup/teardown can't be stopped at this step
124  * @multi_instance: State has multiple instances which get added afterwards
125  */
126 struct cpuhp_step {
127     const char *name;
128     union {
129         (=single)(unsigned int cpu);
130         (=multi)(unsigned int cpu,
131                     struct hlist_node *node);
132     } startup;
133     union {
134         (=single)(unsigned int cpu);
135         (=multi)(unsigned int cpu,
136                     struct hlist_node *node);
137     } teardown;
138     /* private: */
139     struct hlist_head list;
140     /* public: */
141     bool cant_stop;
142     bool multi_instance;
143 };
```

**Große Menge an
Trainingsdaten**



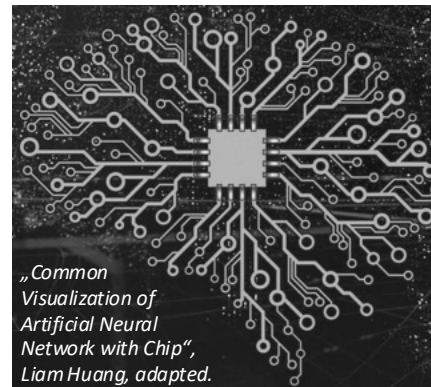
**Künstliches
neuronales Netz**



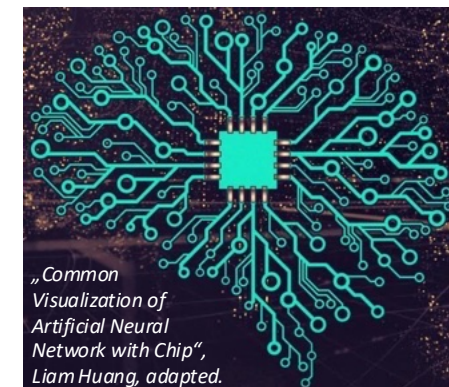
Trainiertes Modell



Große Menge an Trainingsdaten



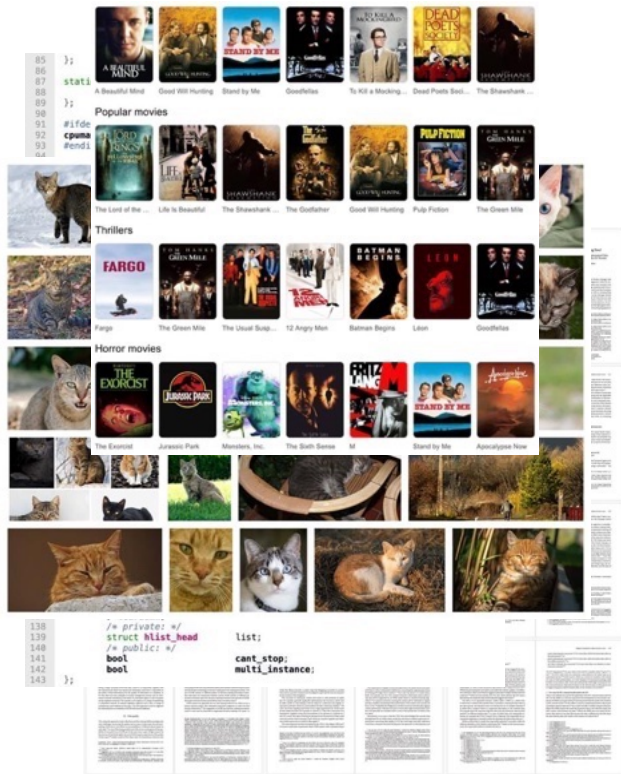
Künstliches neuronales Netz



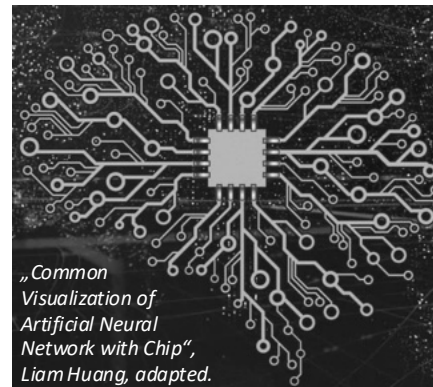
Trainiertes Modell

Das Training generativer Large Language Models

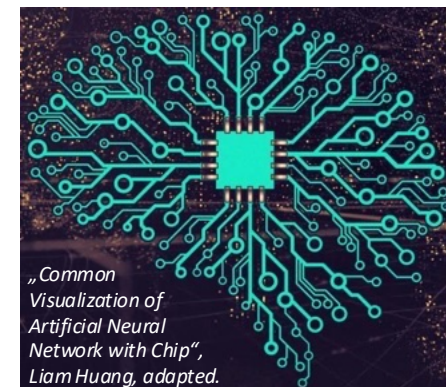
mit Texten und Bildern und Videos



Große Menge an Trainingsdaten



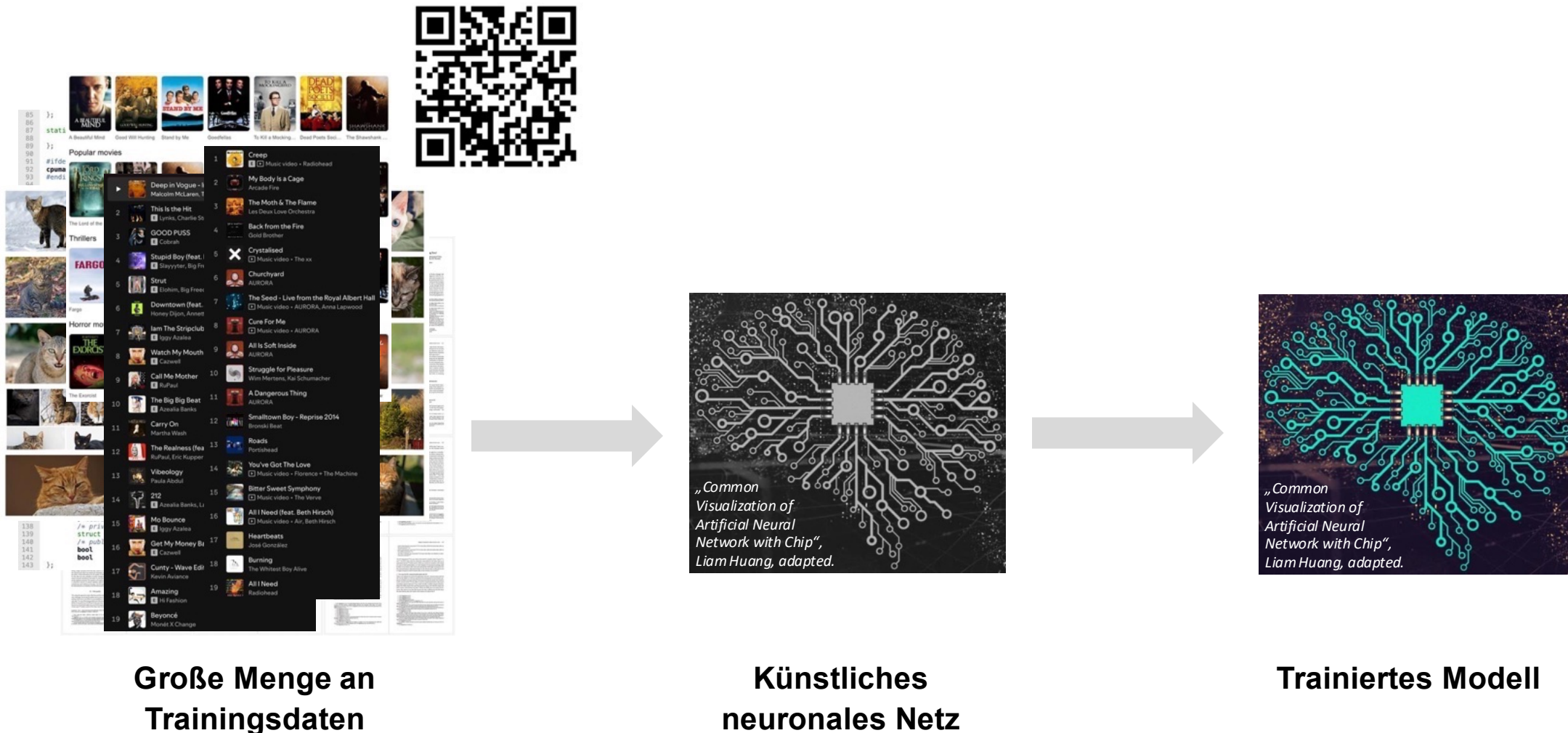
Künstliches neuronales Netz



Trainiertes Modell

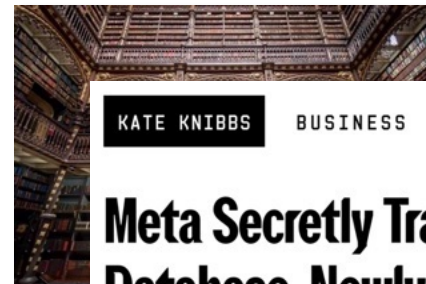
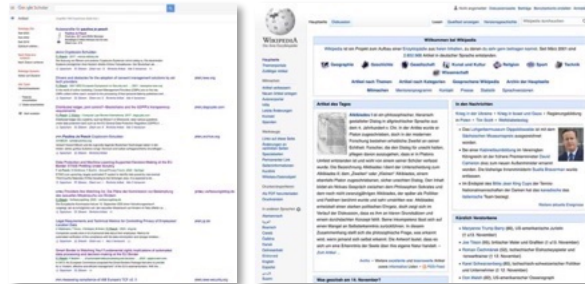
Das Training generativer Large Language Models

mit Texten und Bildern und Videos und Songs



Das Training generativer Large Language Models

Wo kommen die Trainingsdaten her?



KATE KNIBBS BUSINESS JAN 9, 2025 5:33 PM

Meta Secretly Trained Its AI on a Notorious Piracy Database, Newly Unredacted Court Docs Reveal

Meta illegally downloaded 80+ terabytes of books from LibGen, Anna's Archive, and Z-library to train their AI models.

In 2010, Aaron Swartz downloaded only 70 GBs of articles from JSTOR (0.0875% of Meta). Faced \$1 million in fine and 35 years in jail. Took his own life in 2013.



JOHANNA COSTIGAN BUSINESS SEP 19, 2025 6:30 AM

Meta Accused of Torrenting Porn to Advance Its Goal of AI 'Superintelligence'

LAION-400M

An open dataset containing 400 million English image-text pairs.

LAION-5B

A dataset consisting of 5.85 billion multilingual CLIP-filtered image-text pairs.

Das Training generativer Large Language Models

Wo kommen die Trainingsdaten her? Social-Media-Daten, z. B. Meta

META

Making AI Work Harder
for Europeans



Meta trainiert mit **öffentlichen Posts** (Bildern, Bildunterschriften) und laut Datenschutzinformationen auch mit **KI-Interaktionen**.

KI-Interaktionen können **private Sprachnachrichten** und ihre Transkriptionen und **Konversationen mit Chatbots** umfassen.

Harvard Business Review (April 2025): Haupt-Anwendungsfälle für Chatbots sind mittlerweile **persönliche Therapie und Gesellschaft/Freundschaft** (“companionship”).

In Social-Media-Posts, privaten Nachrichten und Konversationen mit Chatbots sind auch **hoch-sensible Daten** enthalten.

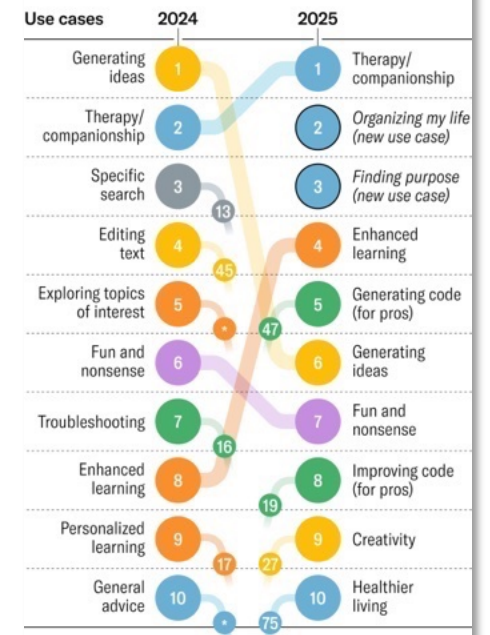
Die **Widerspruchsmöglichkeit ist stark beschränkt**.

Top 10 Gen AI Use Cases

The top 10 gen AI use cases in 2025 indicate a shift from technical to emotional applications, and in particular, growth in areas such as therapy, personal productivity, and personal development.

Themes

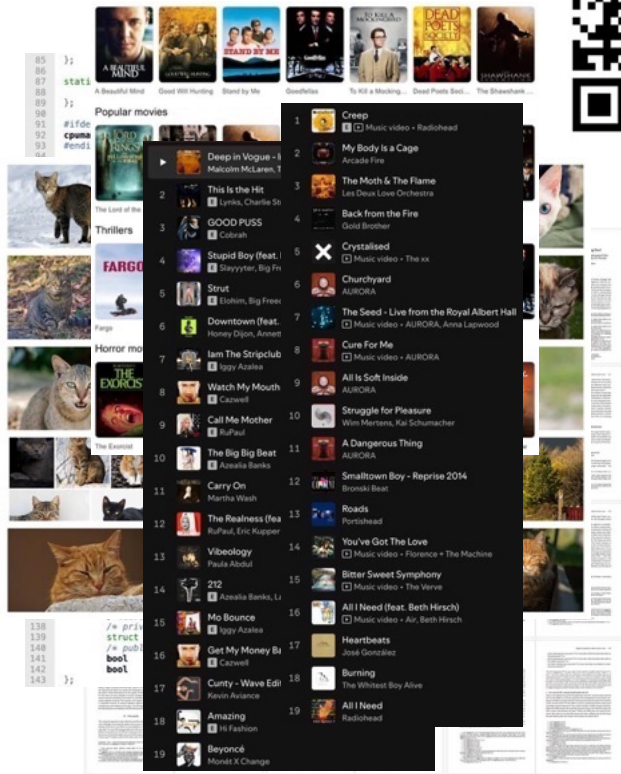
PERSONAL AND PROFESSIONAL SUPPORT	TECHNICAL ASSISTANCE AND TROUBLESHOOTING
CONTENT CREATION AND EDITING	CREATIVITY AND RECREATION
LEARNING AND EDUCATION	RESEARCH, ANALYSIS, AND DECISION-MAKING



*Did not make list of top 100 in 2025
Source: Filtered.com

Das Training generativer Large Language Models

Was "lernen" die KI-Modelle?



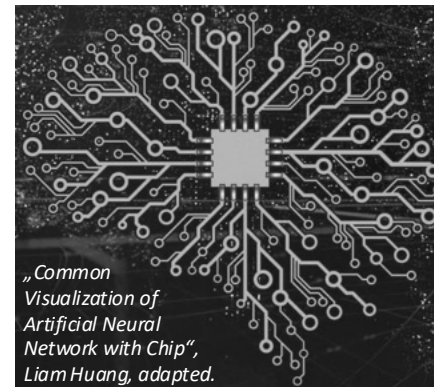
Große Menge an Trainingsdaten

Die KI-Modelle basieren auf **statistischen Verteilungen** in den Daten, z. B.

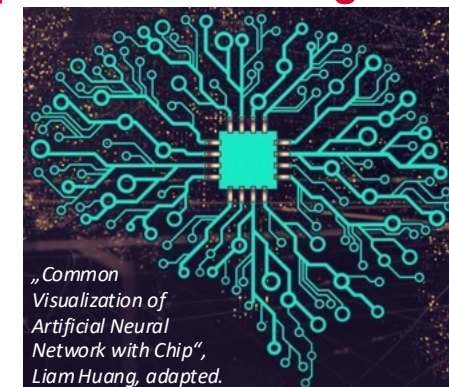
- Welche Worte folgen in einem bestimmten Kontext wahrscheinlich aufeinander?
- Wie sieht ein Bild zu einer bestimmten Beschreibung wahrscheinlich aus?

Die KI-Modelle **verstehen nichts!**

Die KI-Modelle enthalten auch **nicht Kopien aller Trainingsdaten.**



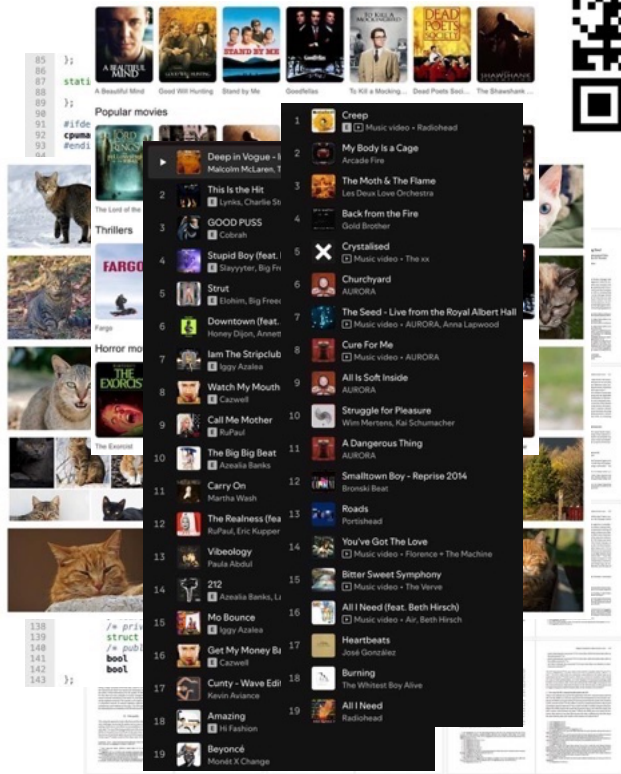
Künstliches neuronales Netz



Trainiertes Modell

Das Training generativer Large Language Models

Was “lernen“ die KI-Modelle?



Große Menge an Trainingsdaten

Die KI-Modelle basieren auf **statistischen Verteilungen** in den Daten, z. B.

- Welche Worte folgen in einem bestimmten Kontext wahrscheinlich aufeinander?
- Wie sieht ein Bild zu einer bestimmten Beschreibung wahrscheinlich aus?

Die KI-Modelle **verstehen nichts!**

Die KI-Modelle enthalten auch **nicht Kopien aller Trainingsdaten**.

Zu einem gewissen Grad ist aber gewollt, dass KI-Modelle **Trainingsdaten exakt wiedergeben** können, z. B. gängige Begriffe.

KI-Modelle **speichern aber ungewollt mehr** und erlauben die Extraktion von einigen Bildern, Texten und Informationen über Personen aus den Trainingsdaten. Daraus können sich insb. **Verstöße gegen das Urheberrecht und das Datenschutzrecht** ergeben.

Systematische Ansätze zur Extrahierung gibt es nicht, aber **Methoden zur Extrahierung werden immer effizienter**.

Wieviele Trainingsdaten betroffen sind, ist **unklar**.

Beispiel für Trainingsdatenextrahierung aus Bildgeneratoren

ORIGINAL



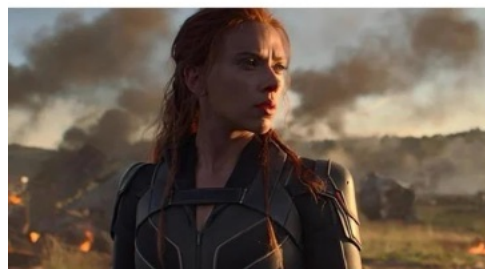
Dune WARNER BROS.

MIDJOURNEY V6

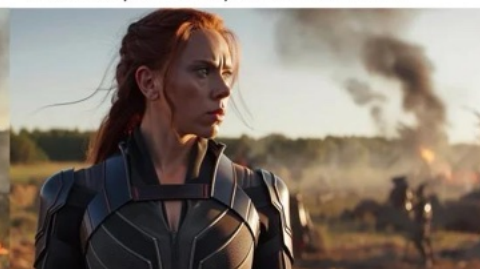
dune movie screencap, 2021, dune movie trailer --ar 16:9 --v 6.0



scarlett johannsen black widow battlefield, 2021, screenshot from a movie, movie scene, official --ar 16:9 --v 6.0



Black Widow MARVEL



the matrix, 1999, screenshot from a movie, movie scene, 4k, bluray --ar 16:9 --v 6.0



The Matrix Resurrections WARNER BROS.



ORIGINAL



MIDJOURNEY V6

Thanos infinity war, 2018, screenshot from a movie, movie scene, 4k, bluray --ar 16:9 --v 6.0



just show me a movie screencap from the avengers infinity war from 2018 halfway through the movie --ar 2:1 --v 6.0 --style raw



Avengers: Infinity War MARVEL

Quelle: Kreye, Es kommt noch geiler, SZ, 9.1.2024, <https://www.sueddeutsche.de/kultur/plagiate-hollywoodbilder-folgen-filme-gefahren-1.6329507>

Manchmal erzeugen generative Large Language Models **Tatsachenbehauptungen**, die nicht in den Trainingsdaten oder Prompts und Kontextinformationen enthalten sind.

Wenn es sich um **personenbezogene Daten** handelt, kann das mit **hohen Risiken für Betroffene** einhergehen.

Martin Bernklau wird Opfer der Künstlichen Intelligenz

KI-Chat macht Tübinger Journalisten zum Kindersch

Copilot-Halluzinationen über Martin Bernklau:

"[...] in einem Missbrauchs-Fall gegen Kinder [...] angeklagt."

"[...] spektakulär aus dem Zentrum für Psychiatrie [...] ausgebrochen [...]"

"[...] skrupellose[r] Bestatter aus Rostock, der trauernde Frauen ausnutzt. [...] Straftaten begangen, darunter Betrug, Diebstahl und unerlaubten Waffenbesitz."

Martin Bernklau hat **als Gerichtsreporter über die Fälle berichtet**.



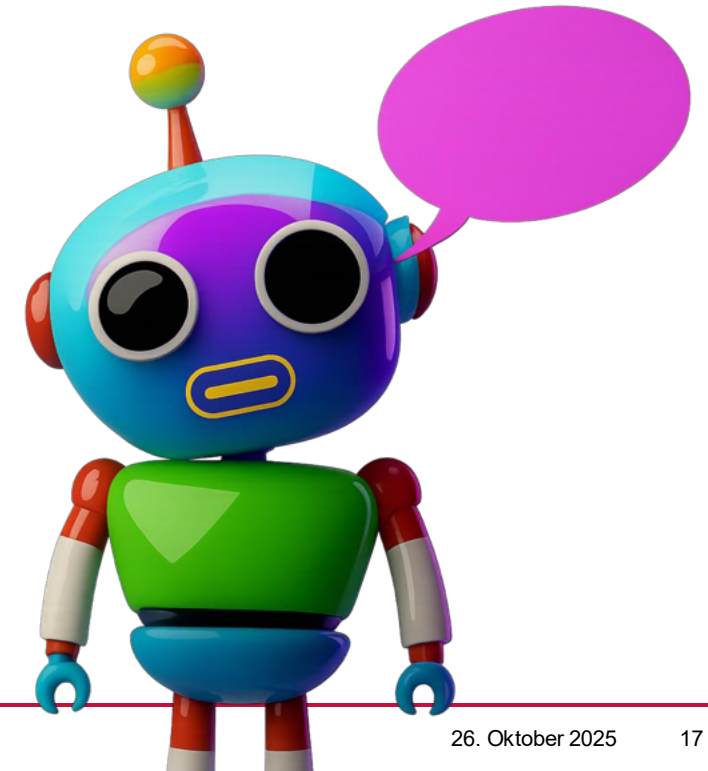
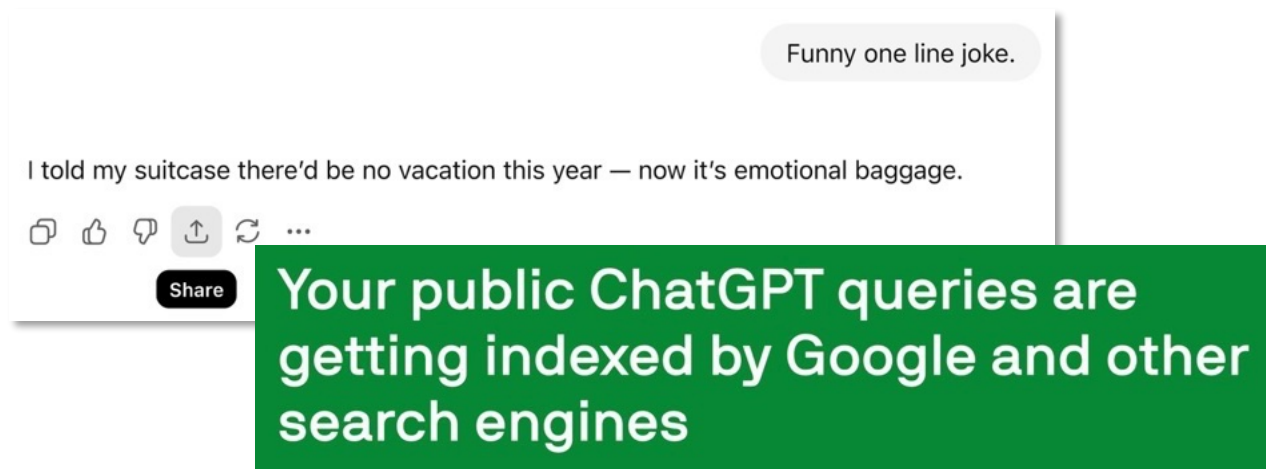
Wie lassen sich Rechtsverstöße vermeiden

Datenschutzrecht

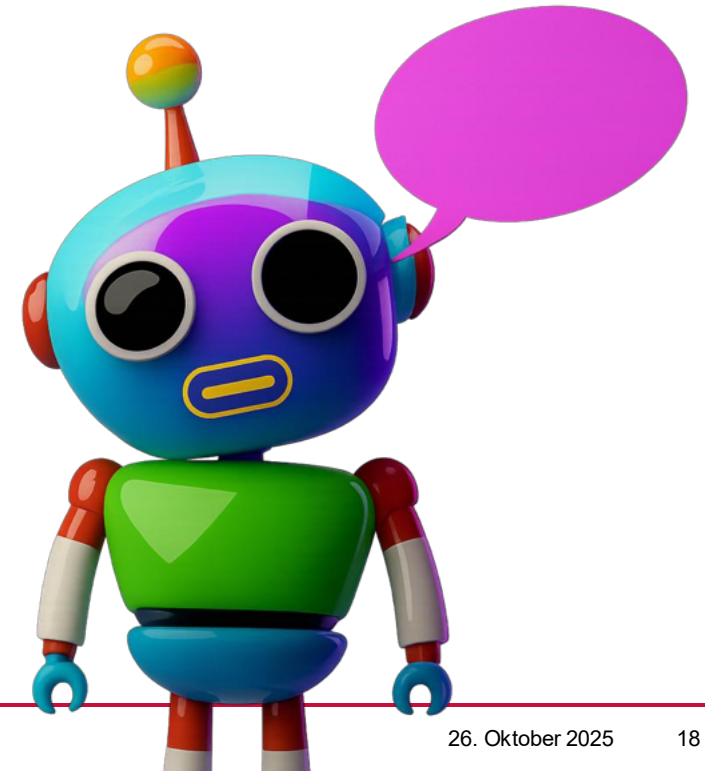
- Anwendbar auf die **Verarbeitung personenbezogener Daten** (Informationen über identifizierbare Personen)
- Verarbeitung z. B. **Einstellen auf Social-Media-Plattformen, Transkription von Sprachnachrichten, Eingabe in Prompt** (z. B. „Wie alt ist...“?)
- Keine Anwendbarkeit des Datenschutzrechts auf **rein persönliche Tätigkeiten**

≠ Teilen mit unbegrenztem Personenkreis; in größeren Gruppen

Chatboteingabe? Tipp: Verläufe nicht speichern und Training mit eigenen KI-Interaktionen unterbinden. Verläufe nicht veröffentlichen.



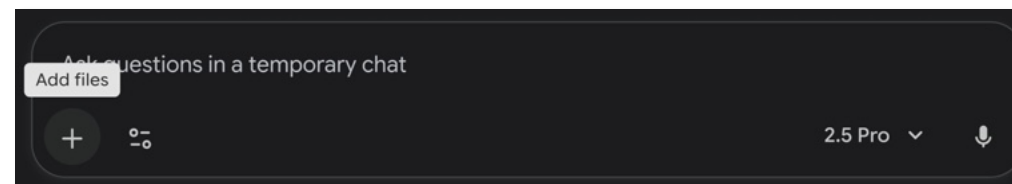
- Anwendbar auf die **Verarbeitung personenbezogener Daten** (Informationen über identifizierbare Personen)
- Verarbeitung z. B. **Einstellen auf Social-Media-Plattformen, Transkription von Sprachnachrichten, Eingabe in Prompt** (z. B. „Wie alt ist...“?)
- Keine Anwendbarkeit des Datenschutzrechts auf **rein persönliche Tätigkeiten**
≠ Teilen mit unbegrenztem Personenkreis; in größeren Gruppen
Chatboteingabe? Tipp: Verläufe nicht speichern und Training mit eigenen KI-Interaktionen unterbinden. Verläufe nicht veröffentlichen.
- Im Anwendungsbereich des Datenschutzrechts Verarbeitung insb. nur mit **Rechtsgrundlage** und unter Einhaltung des **Datenrichtigkeitsgebots** (Überprüfung von Ausgaben vor deren Weiterverwendung!)
- Hohe Anforderungen an die Verarbeitung von **besonderen Kategorien personenbezogener Daten** ≈ bestimmte sensible Daten
- Im Zweifel sollten personenbezogene Daten **nicht in Chatbots eingegeben** werden.



- Eingaben und Ausgaben von Chatbots können **urheberrechtlich geschützte Inhalte** enthalten, z. B. Sprachwerke wie Bücher, oder journalistische Artikel oder Fotos
- **Vervielfältigungen** (≈ Kopien) sind grundsätzlich nur mit Zustimmung der Rechteinhaberin zulässig
- Vervielfältigungen entstehen insb., wenn

- die Modelle (ggfs. unbemerkt) **Trainingsdaten (nahezu) wörtlich wiedergeben**
- geschützte Inhalte in einem Chatbot **als Kontext hochgeladen** werden.

Im Rechtsstreit mit Open AI deutet sich ein Sieg für die Gema an



- Sie können aber aufgrund von Schrankenregelungen **ausnahmsweise ohne Zustimmung zulässig** sein, insb. zu wissenschaftlichen Forschungszwecken.
- Jedenfalls: **Unterbinden des Trainings** mit Eingaben in den Einstellungen der Chatbotanbieterin.
- Rechtsverfolgung droht aber vor allem bei der Weiternutzung von Ausgaben: **Sorgfältiges Prüfen auf etwaige Plagiate.**

'After colonizing the adult world,
AI enters the world of children'

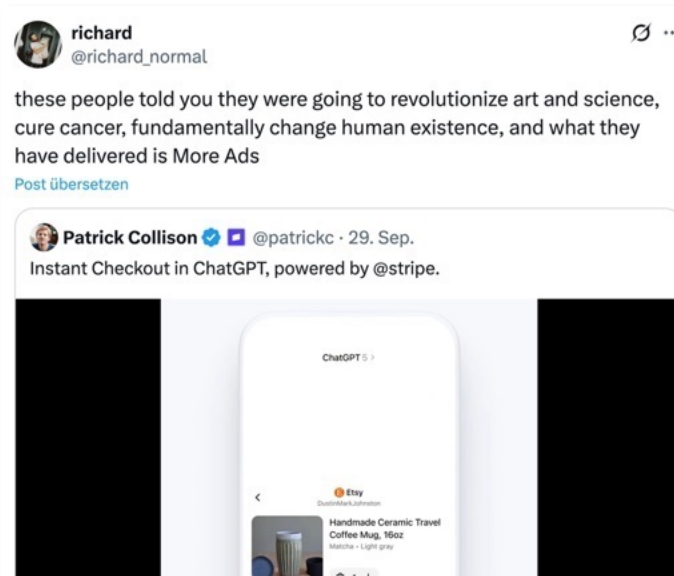
With just a few messages,
biased AI chatbots swayed
people's political views

Generative AI

AI-Generated “Workslop” Is Destroying Productivity

by Kate Niederhoffer, Gabriella Rosen Kellerman, Angela Lee, Alex Liebscher, Kristina Rapuano and Jeffrey T. Hancock

September 22, 2025, Updated September 25, 2025



The Chatbot Culture Wars Are Here

Conservatives, including President Trump, are accusing A.I. companies of left-wing bias, following a playbook that worked well against social media platforms.

AI as Normal Technology

An alternative to the vision of AI as a potential superintelligence

BY **ARVIND NARAYANAN & SAYASH KAPOOR**
APRIL 15, 2025

Herzlichen Dank!

