

Lost in Citation

KI-Memorization im Kontext aktueller Entwicklungen des Urheber- und
Datenschutzrechts

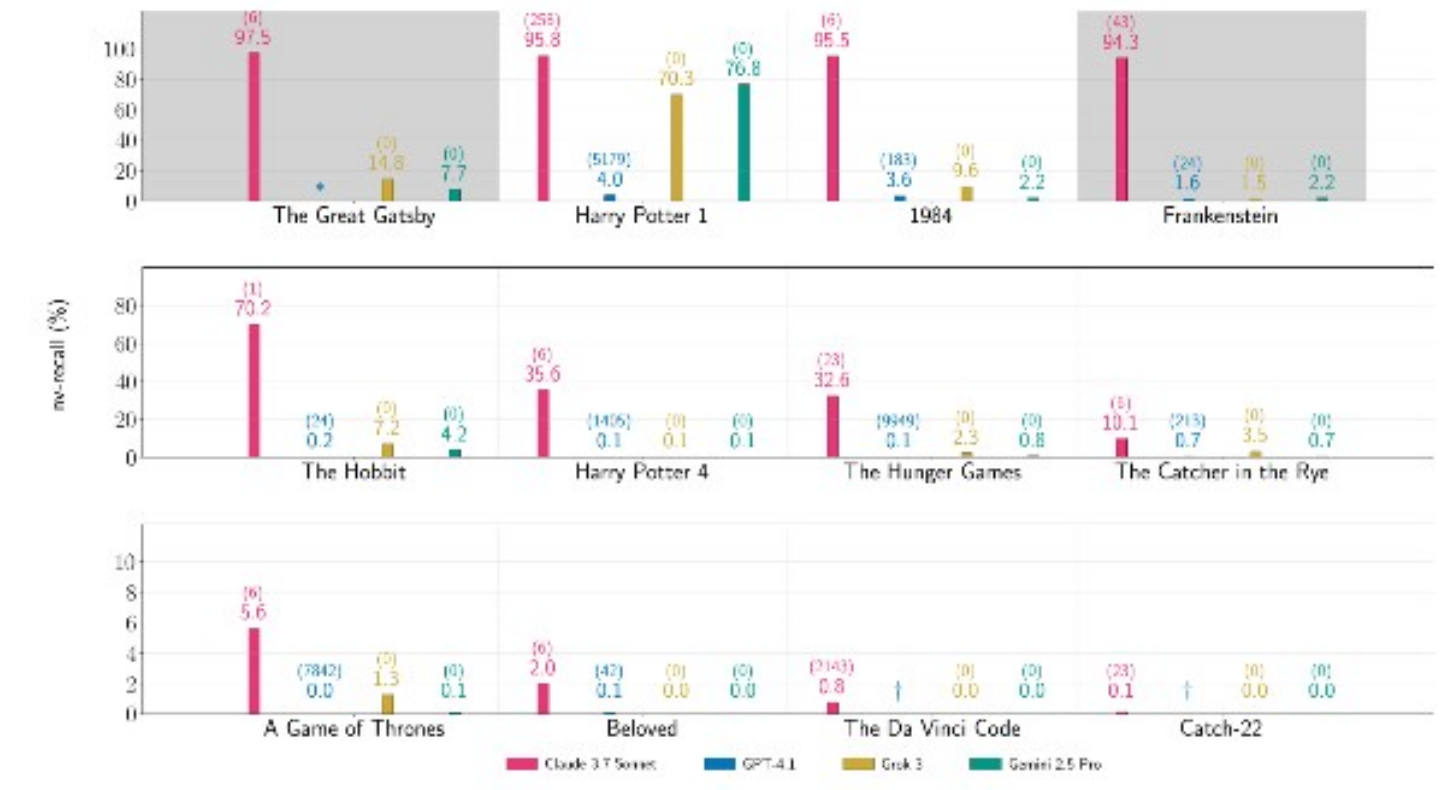
Dr. Michael Funke | Dr. Lisa Käde

JBViniol

Herbstakademie 2026

The logo for JBViniol features the letters 'j', 'b', 'v', 'i', 'n', 'i', 'o', 'l' in a bold, sans-serif font. The 'i' and 'o' are stylized with red dots above and below them, respectively.

Das Problem / Motivation



Quelle: Ahmed et al. Extracting books from production language models, 2026, S. 12

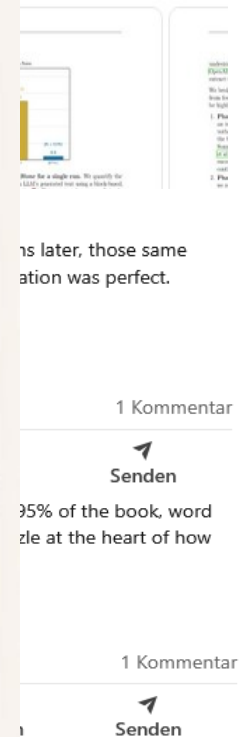
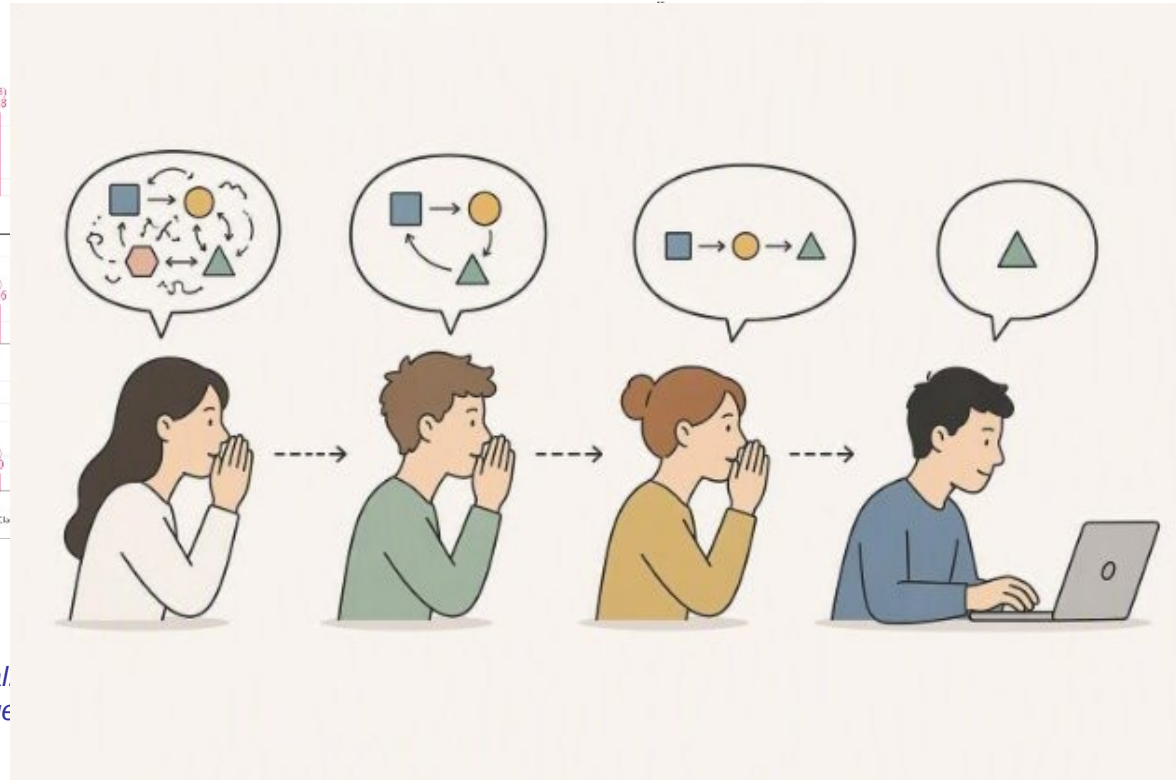
Das Problem / Motivation

New Stanford paper shows production LLMs can leak near exact book text, with Claude 3.7 Sonnet hitting 95.8%.

... mehr

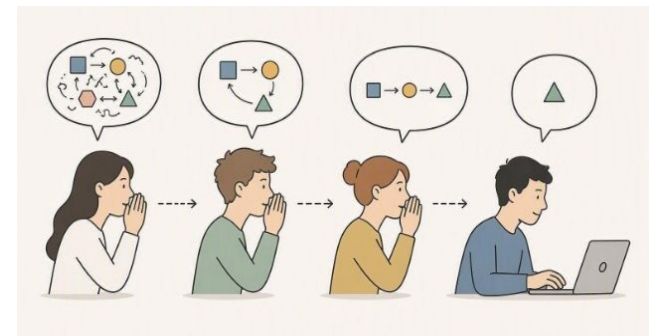


Quelle: Ahmed et al.
production language



Das Problem / Motivation

- ▶ Weltweite Frage: „Speichern“ bzw. „memorisieren“ KI-Modelle ihre Trainingsdaten?
- ▶ Verweis auf „bahnbrechende Studien“ – aber: Beiträge geben sie verkürzt/selektiv wieder
- ▶ Verzerrung pflanzt sich durch Zitationsketten fort → „Lost in Citation“
- ▶ **Take-away:**
Kritische Rezeption der Originalforschung ist entscheidend



Grundlagen und Beispiele

Memorization: Beispiele von Carlini et al. (2023), „Extracting Training Data from Diffusion Models“



Figure 3: Examples of the images that we extract from inference procedure. The top row shows the original images.



Figure 14: Inpainting-10. Given an image (I) of the image (second column) which was trained on model which was not

Quelle: Carlini et al. (2023), in: Proceedings of the 32nd USENIX Security Conference

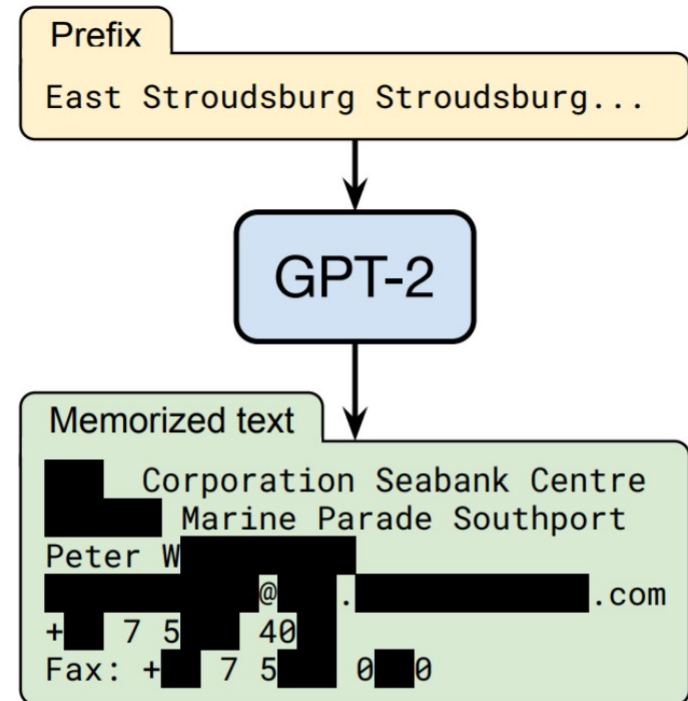


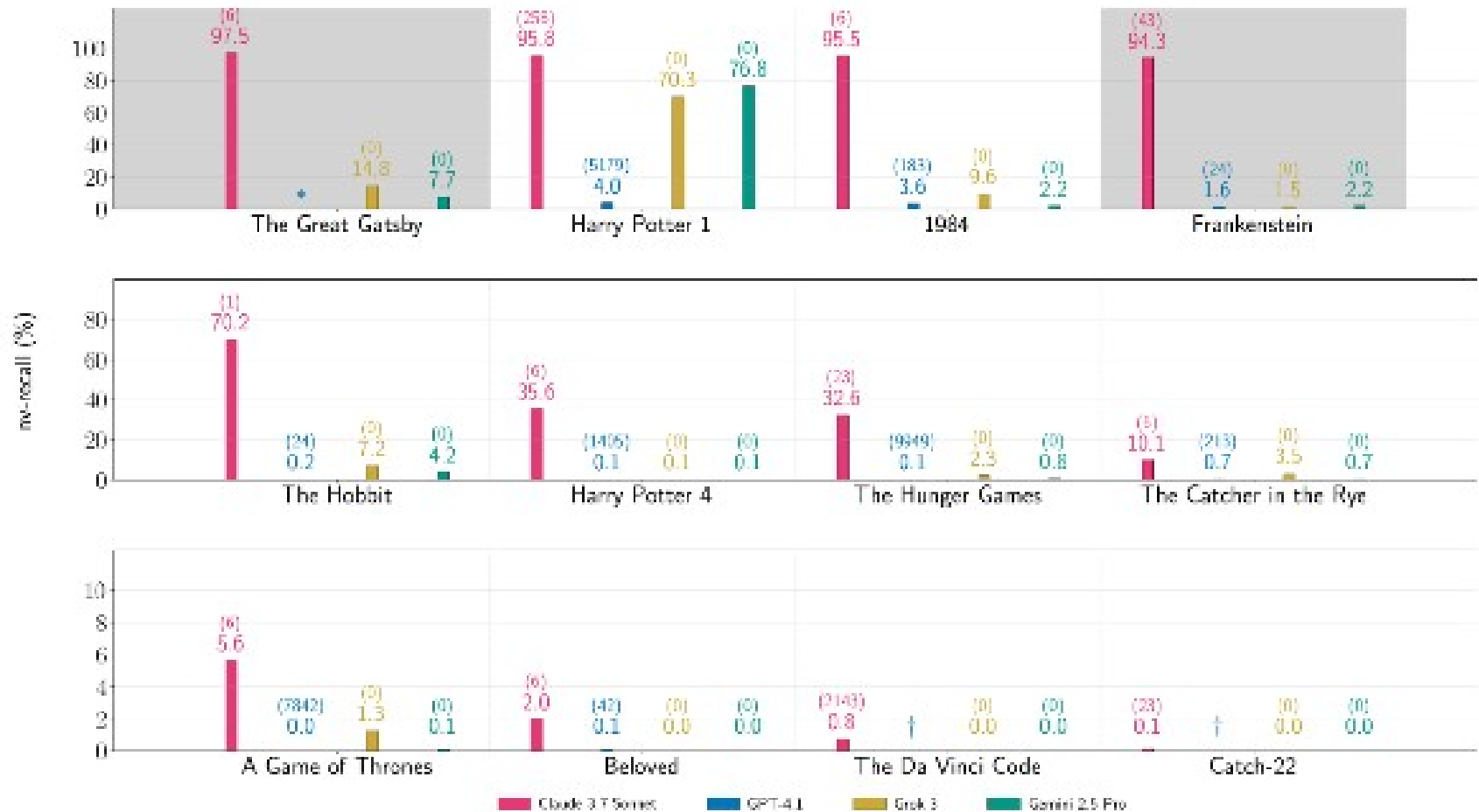
Figure 1: **Our extraction attack.** Given query access to a neural network language model, we extract an individual person's name, email address, phone number, fax number, and physical address. The example in this figure shows information that is all accurate so we redact it to protect privacy.

Quelle: <https://nicholas.carlini.com/writing/2025/privacy-copyright-and-generative-models.html>

Begriff und Funktionsweise

- ▶ „Memorization“ = Auswendiglernen; **keine einheitliche Definition**
- ▶ LLMs = wahrscheinlichkeitsbasierte „Sprachbaumaschinen“ – Texte **nicht** wie auf Festplatte gespeichert
- ▶ Ziel der Modelle: **Generalisieren**, nicht Auswendiglernen

Die Grundaussage



Ursachenforschung und Forschungsstand

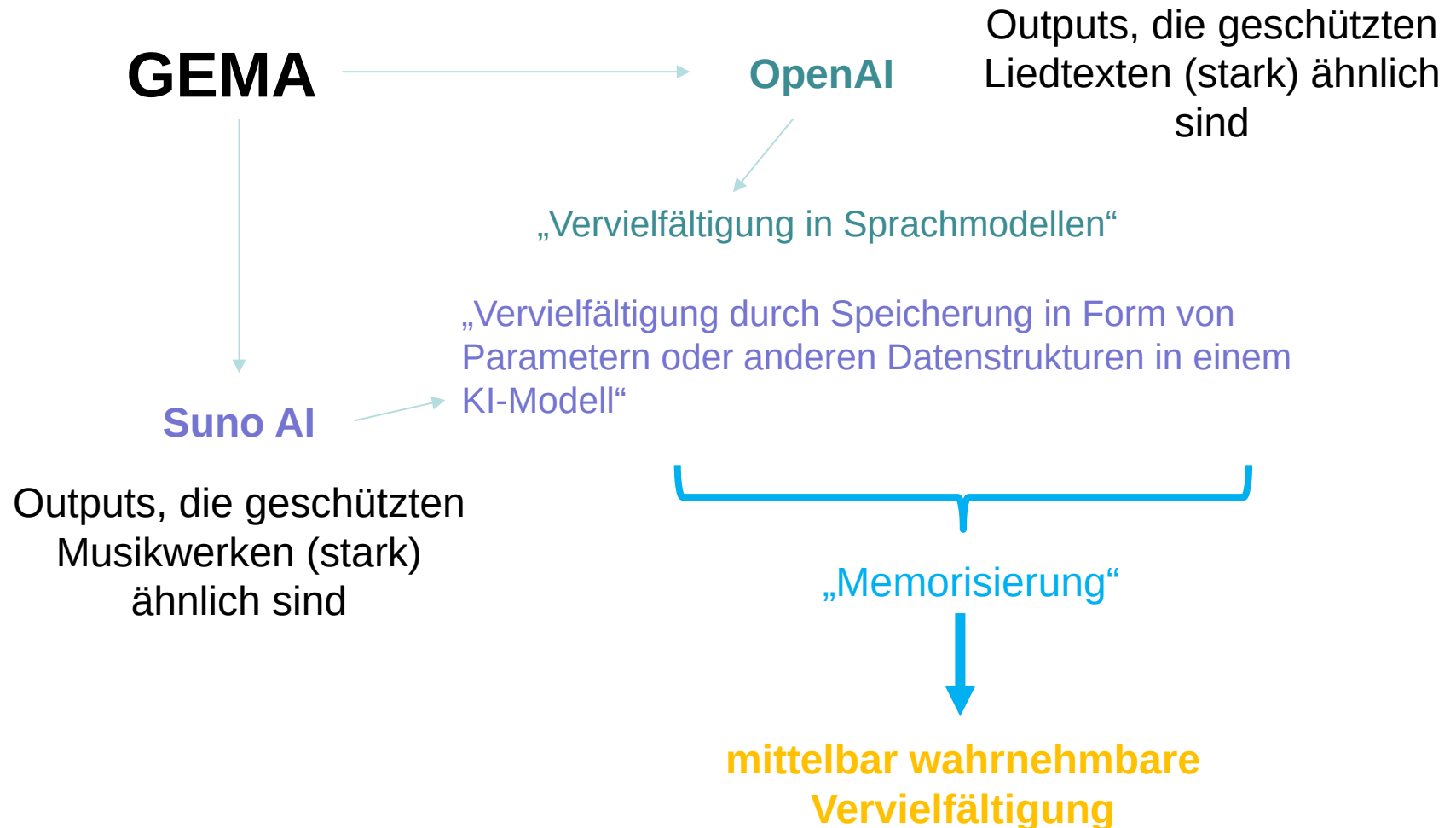
- ▶ **Ursache und Umfang** tatsächlich noch wenig erforscht
 - ▶ Prashanth et al.: Rezitieren / Rekonstruieren / Erinnern
 - ▶ Morris et al.: „Speichergrenze“, danach Generalisierung und keine Memorization
- ▶ **Forschungsstand:**
 - ▶ KI-Modelle können in einigen Fällen Inhalte in Teilen wiederzugeben, die in den Trainingsdaten enthalten waren
 - ▶ Memorization betrifft unterschiedliche KI-Modelle in unterschiedlichem Maße.
 - ▶ Nicht alle Trainingsdaten sind von diesem Effekt betroffen.
 - ▶ Nur weil ein Output Trainingsdaten enthält, kann nicht zwingend darauf geschlossen werden, dass (unbeabsichtigte) Memorization hierfür die Ursache ist.



Rechtlicher Kontext

- ▶ **Urheberrecht:** LG München I, GEMA ./ OpenAI (Nov. 2025); LG München I, GEMA ./ Suno (Juli 2026)
→ Reichweite der TDM-Schranke (44b UrhG), Einordnung von „Memorisierung“ als Vervielfältigung
- ▶ **Datenschutz:** Enthalten Modelle personenbezogene Daten? EDSA-Stellungnahme 28/2024
- ▶ Gemeinsamer Nenner: beide stützen sich auf dieselbe (verzerrt zitierte) Forschung, beide werden verzerrt zitiert

Rechtsprechung: Urheberrecht



Rechtsprechung: Urheberrecht

!! Spannende Entscheidung mit Signalewirkung für künftige #KI Fälle !!

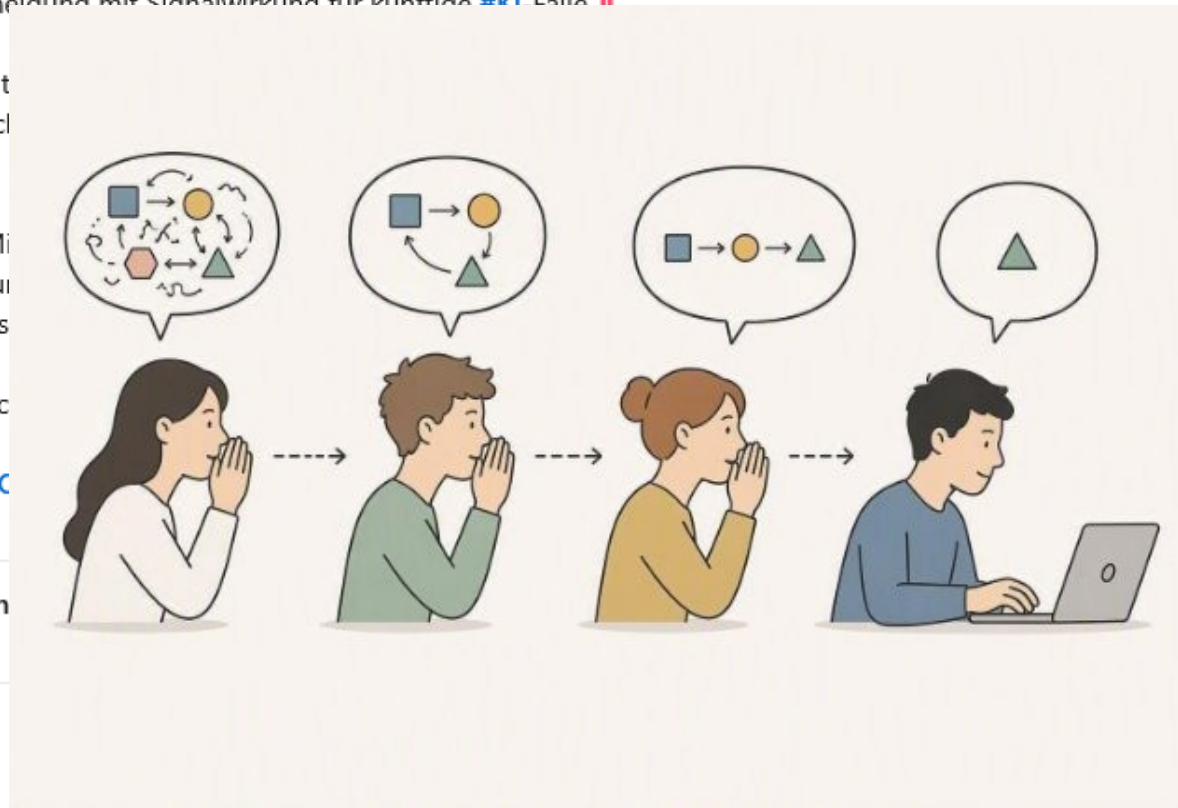
Das LG München I hat urheberrechtlich geschützte InfoSoc-RL darstellt.

Die Text- und Data-Mineralmodell memorisiert und der Texte in Outputs s

Das Urteil ist nicht rec

#Urheberrecht #KI #C
ce #IPLaw #GÖRG

Urteil GEMA gegen
justiz.bayern.de



echtlich geschützten
as hat heute das
GEMA (=... mehr

1.2025

Urteil GEMA gegen Open

Datenschutzrechtliche Bewertung

- ▶ Modelle speichern Wahrscheinlichkeiten, keine Klartext-Daten
- ▶ Identifizierbarkeit (Art. 4, ErwG 26 DSGVO): nur Mittel, die „nach allgemeinem Ermessen wahrscheinlich“ genutzt werden
- ▶ Extraktion: hoher Aufwand, ungewisser Ausgang, Schutzmechanismen
- ▶ **Ergebnis: bessere Argumente gegen pauschalen Personenbezug**
– nicht vorschnell „zur Sicherheit“ das Datenschutzrecht anwenden

Urheberrechtliche Bewertung

- ▶ Memorization dient als „Krücke“ für die These: Modell = Vervielfältigung im urheberrechtlichen Sinne
- ▶ Bei Vervielfältigung im Modell: Löschpflicht nach § 44b Abs. 2 UrhG kaum erfüllbar; Weitergabe nicht mehr gedeckt
- ▶ Problem: keine breite Reproduktion belegt; meist adversariale Experimentalbedingungen
- ▶ **Ergebnis: pauschales „alle Trainingsdaten sind vervielfältigt“ ist nicht haltbar**

Rechtspolitik: Wohin führt das?

- ▶ Umfang der Memorization ist **kein taugliches Maß** für die Rechtslage
- ▶ Eigentliche Interessen offen benennen: Konkurrenzschutz? Vergütung Kulturschaffender? Kontrolle?
- ▶ Lösung gehört in gesetzgeberischen Diskurs (idealerweise EU-weit) – **DSM-Richtlinien-Evaluierung als Chance**

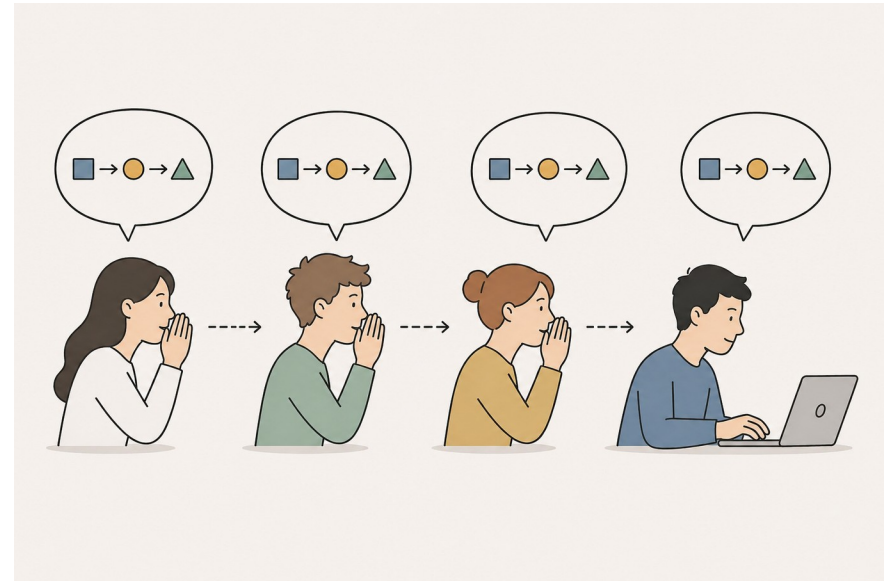
Fazit & Kernbotschaften

- ▶ Memorization ist in *gewissem Maße* vielen Modellen immanent, aber **kein Maß für alle Trainingsdaten**
- ▶ Datenschutz: risikobasiert bewerten statt vorschnell anwenden
- ▶ **Kritische Rezeption der Forschung ist essenziell**
- ▶ KI-Forschung + Rechtswissenschaft brauchen einen **ehrlichen Dialog**

Nothing's really making any sense at all, let's talk

Coldplay – Talk (2005)

→ Es führt kein Weg mehr an interdisziplinärer Zusammenarbeit vorbei, um die komplexen Rechtsfragen zu lösen, die neue Technologien mit sich bringen.



Vielen Dank für die Aufmerksamkeit!

Dr. Michael Funke
Partner | JBVinol
funke@jbvinol.de



Dr. Lisa Käde
Associate | JBVinol
kaede@jbvinol.de



jbvinol

