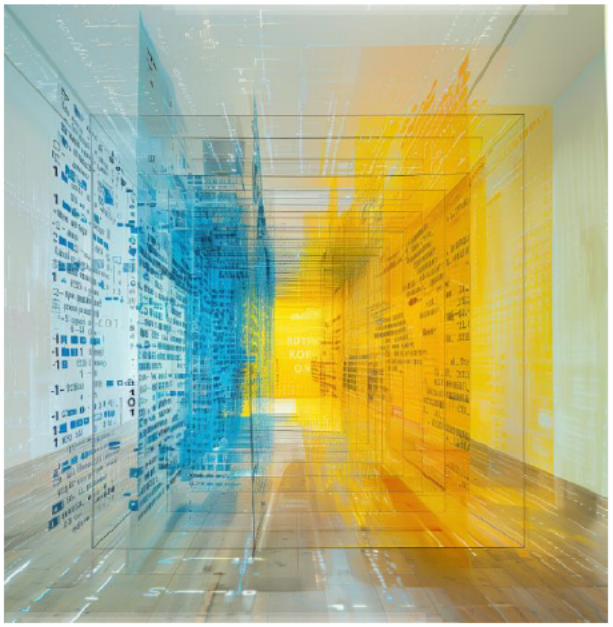
Verheddert im Neuronalen Netz – Aktuelles zu KI und Datenschutz



Sommerakademie
am 9. September 2024
in Kiel

Angelika Martin, Christian Krause, Benjamin Walczak
0431 988-1200
mail@datenschutzzentrum.de
https://www.datenschutzzentrum.de/

ULD

Unabhängiges Landeszentrum für
Datenschutz Schleswig-Holstein

Überblick

- **Personenbezug in großen Sprachmodellen**
  Benjamin Walczak
  - „Sieht man doch!"
  - „Ist auch so!"
- **Offene Diskussion**
  Angelika Martin
  Christian Krause
  Benjamin Walczak

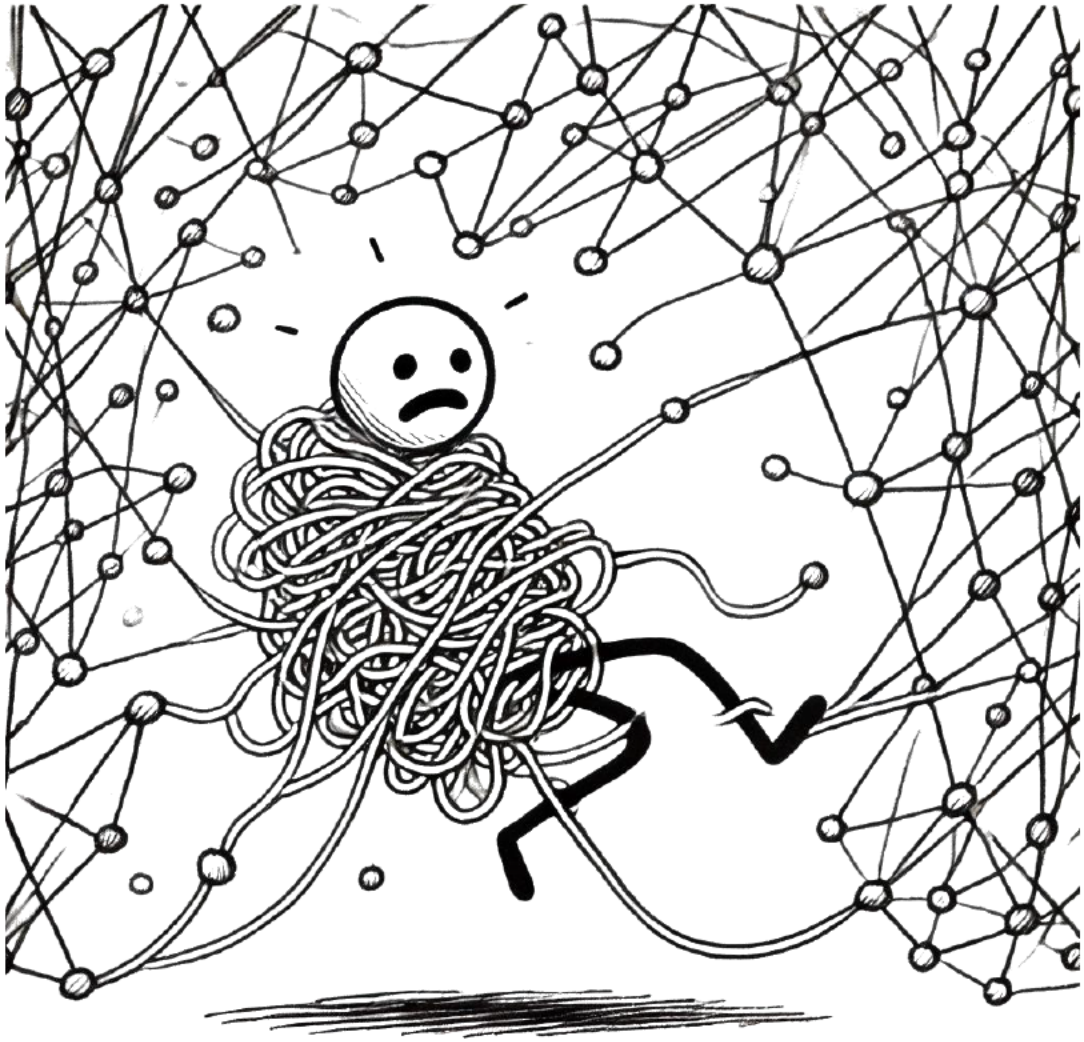


Bild erstellt mit DALL-E, OpenAi.

Save the date

Mi., 6. November, 18 Uhr

ULD, Holstenstr. 98, Kiel

KI: Fragen für`n Freund

Fachaustausch über aktuelle Fragen

und Entwicklungen

Personenbezug in großen Sprachmodellen

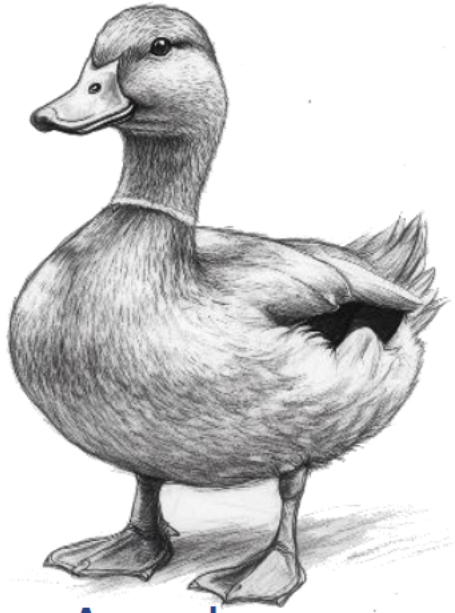
(Large Language Models, LLM)

- **„Sieht man doch!"**
  eine phänomenologische Annäherung

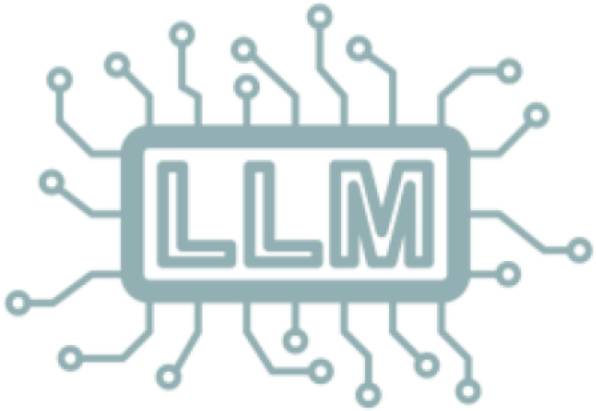
- **„Ist auch so!"**
  eine technische Annäherung

- **Fazit und neue Überlegungen**

„Sieht man doch!"



Aussehen
+ Watscheln
+ Quaken
→ **Ente**



Trainingsdaten personenbezogen
+ Ausgaben personenbezogen
+ Attacken möglich
→ **LLM ist personenbezogen**

„Sieht man doch!"

Stärke der LLMs:
Aufbauend auf großer Menge an Trainingsdaten einen gegebenen Text Wort für Wort verlängern.

Wie kocht man Nudeln?

1. Wasser kochen.
2. Salz hinzufügen.
3. Nudeln ins Wasser geben.
4. Nach Packungsanweisung kochen.
5. Abgießen und servieren.

→

Wirkt wie eine Konversation.
LLMs können so tlw. sehr komplexe Aufgaben erledigen.

„Sieht man doch!"

Habe nun, ach!

Philosophie,
Juristerei und Medizin,
Und leider auch Theologie
Durchaus studiert, mit heißem Bemühn.

→

Bekannte Texte (d.h. oft in den Trainingsdaten) oder sehr markante Texte, können wiederholt werden („Memorization").

„Sieht man doch!"

Neue Verknüpfungen:

Beim Training und beim Abarbeiten von Anfragen an das System können neue Verknüpfungen entstehen.

Der Wahrheitsgehalt ist kaum abbildbar.

Einsatz von LLMs von wissenschaftlichen Texten bis Märchen – konkrete Anfrage ist für das System schwer zu bewerten.

„Ist auch so!"

Speichern ist vielfältig:

Daten können sequentiell gespeichert werden.
(z.B. Texte, Bilder)

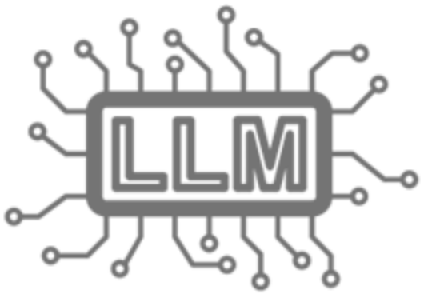
Daten können in einer Baumstruktur gespeichert werden.

(z.B. Präfixbaum, "Trie")

Es gibt probabilistische Datenstrukturen.
(z.B. Bloom-Filter)

Verschlüsselte Daten sind ohne Schlüssel nicht mehr lesbar strukturiert.

„Ist auch so!"

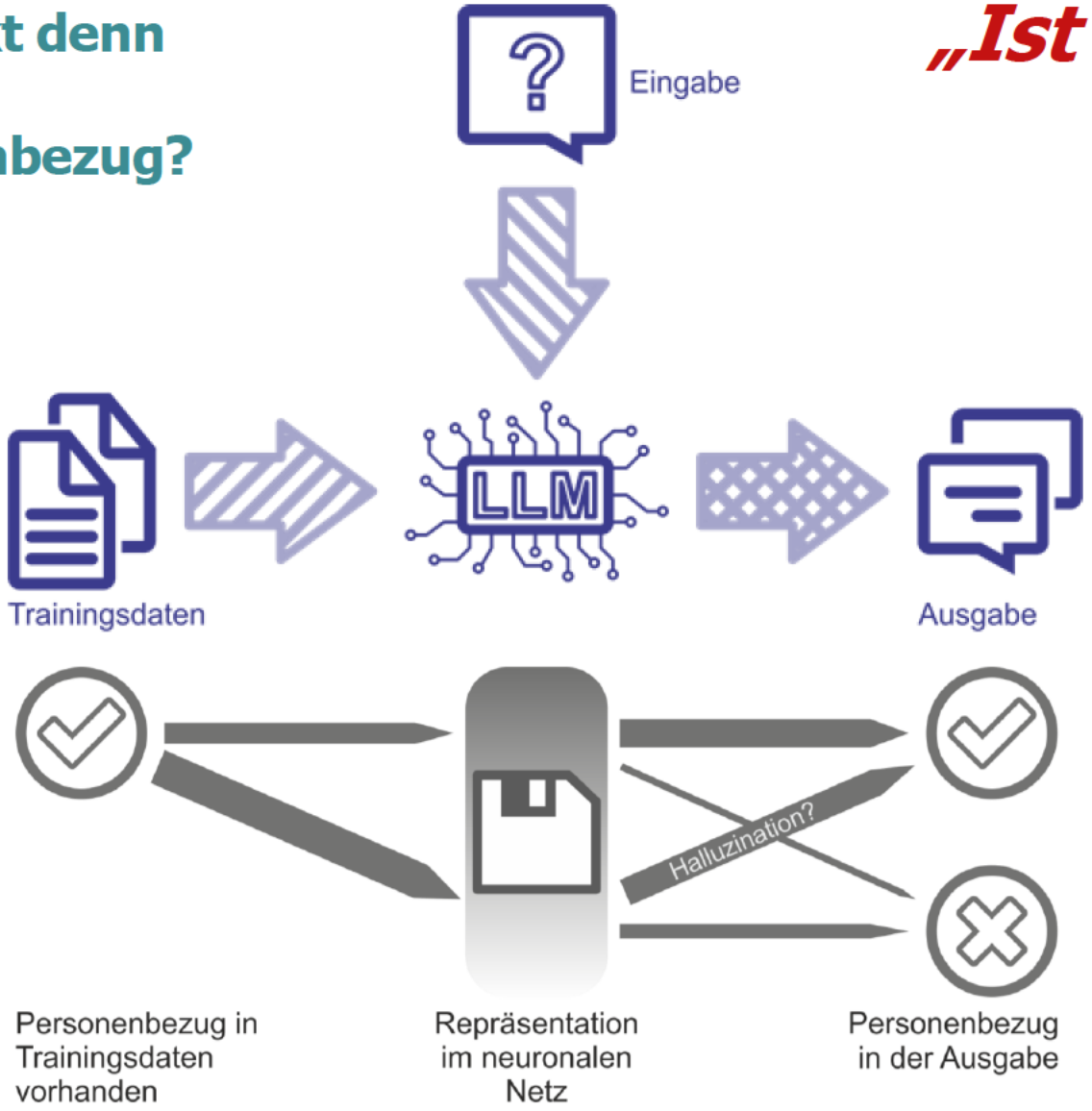


Kompressions-Theorie:
Speicherplatz für LLMs ist deutlich kleiner als die Trainingsdaten.
Viele Informationen gehen beim Training verloren, vermutlich oft personenbezogene Daten.

Halluzinationen:
Schwach repräsentierte (tatsächliche) Informationen können ähnliche Wahrscheinlichkeit der Ausgabe haben wie andere zufällig generierte Informationen.

Wo steckt denn nun der Personenbezug?

„Ist auch so!"



Fazit

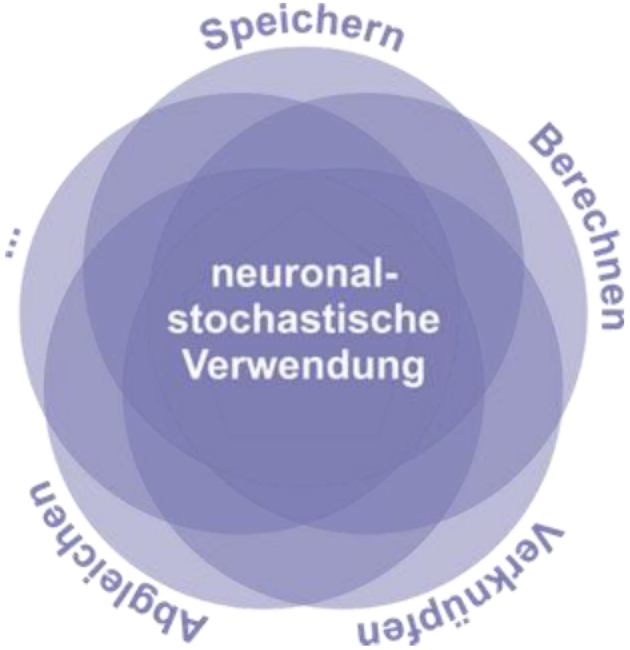
Die Funktionsweise von KI-Modellen ist sehr komplex.

- Keine pauschalen Aussagen möglich.
- LLM mit Personenbezug nachweislich herstellbar.

KI und DSGVO

- DSGVO ist technologieoffen, …
- … aber viele Assoziationen zu Kompression, Scoring oder Big Data – aber sie decken nicht die Komplexität ab.

Überlegungen



Neuartige Verarbeitung auch so benennen!

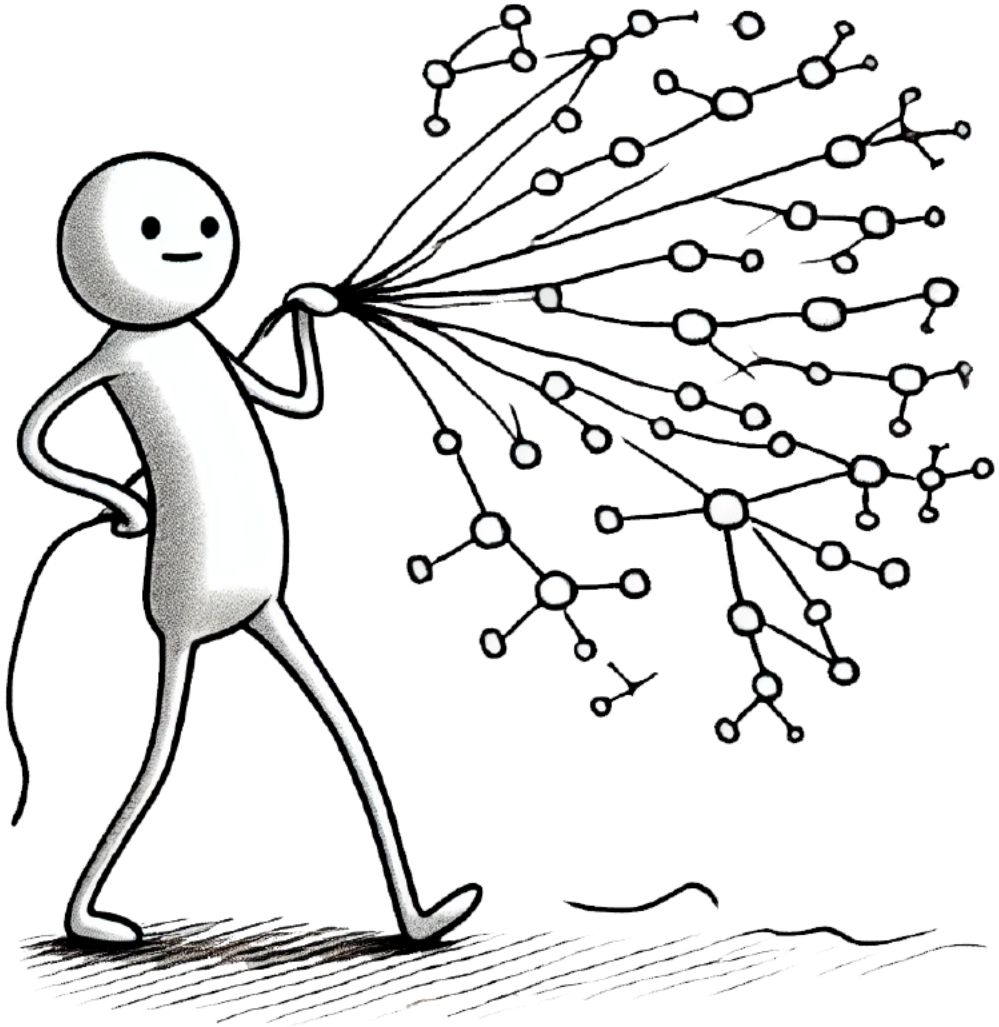
- **Training** (Berechnung von Verbindungen im Neuronalen Netz, Kontexte berücksichtigen, Verknüpfungen speichern, …)
- **Einsatz** (Berechnung der Ausgabe anhand der Eingabe und des Kontexts, neue Verknüpfung von Datenpunkten, …)

Perspektive (Diskussionsgrundlage)

- Rauschen (analog zu Differential Privacy) 👍
- Stärkere Verkettbarkeit 👎

Offene Diskussion

Wie erreichen wir Kontrolle über die KI?



mit

Angelika Martin

Christian Krause

Benjamin Walczak

Bild erstellt mit DALL-E, OpenAi.

Save the date

Mi., 6. November, 18 Uhr

ULD, Holstenstr. 98, Kiel

KI: Fragen für`n Freund

Fachaustausch über aktuelle Fragen

und Entwicklungen