



ALGORITHMISCHE WERTSCHÖPFUNG STATT REGELBASIERTER WELTORDNUNG: WARUM KI-LIEFERKETTEN SCHWER ZU REGULIEREN SIND

Prof. Dr. Matthias Krauledat //
Fakultät Technologie und Bionik //

Prof. Dr. Matteo Große-Kampmann
Fakultät Kommunikation und Umwelt

HOCHSCHULE RHEIN-WAAL

LARGE LANGUAGE MODEL (LLM)

Sprachmodell aus der KI

Fähigkeit zur Textgenerierung

Wahrscheinlichkeitsbasiert

Lernt Wort- und Satzfolgebeziehungen aus Textdokumenten



WARUM BRAUCHEN WIR REGELN FÜR KI?

Beispiel: GPT-4chan

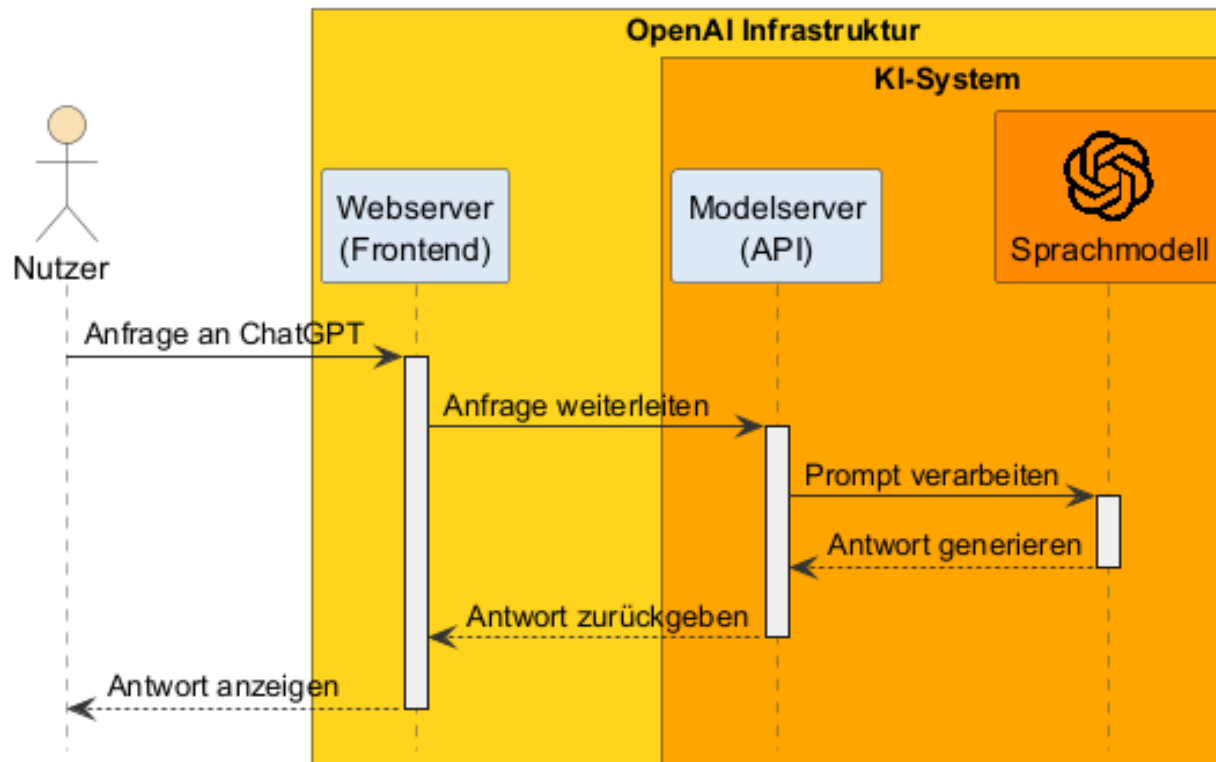


WAS SIND DIE RISIKEN?

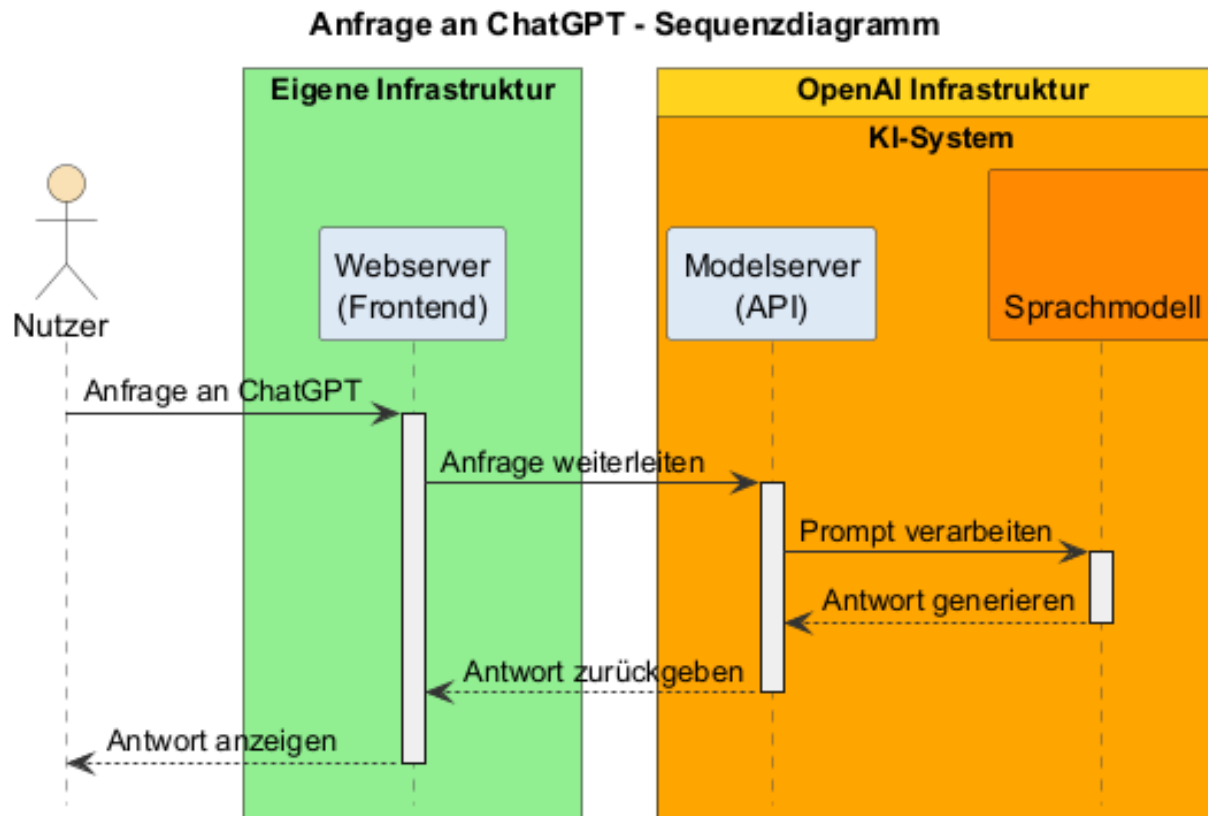
- Ungenauigkeiten
- Bias (Voreingenommenheit)
- Desinformation
- Hate Speech
- Erzeugen von kinderpornographischem Material
- Rassistische Äußerungen
- ...

WIE NUTZT MAN LLMS?

Anfrage an ChatGPT - Sequenzdiagramm

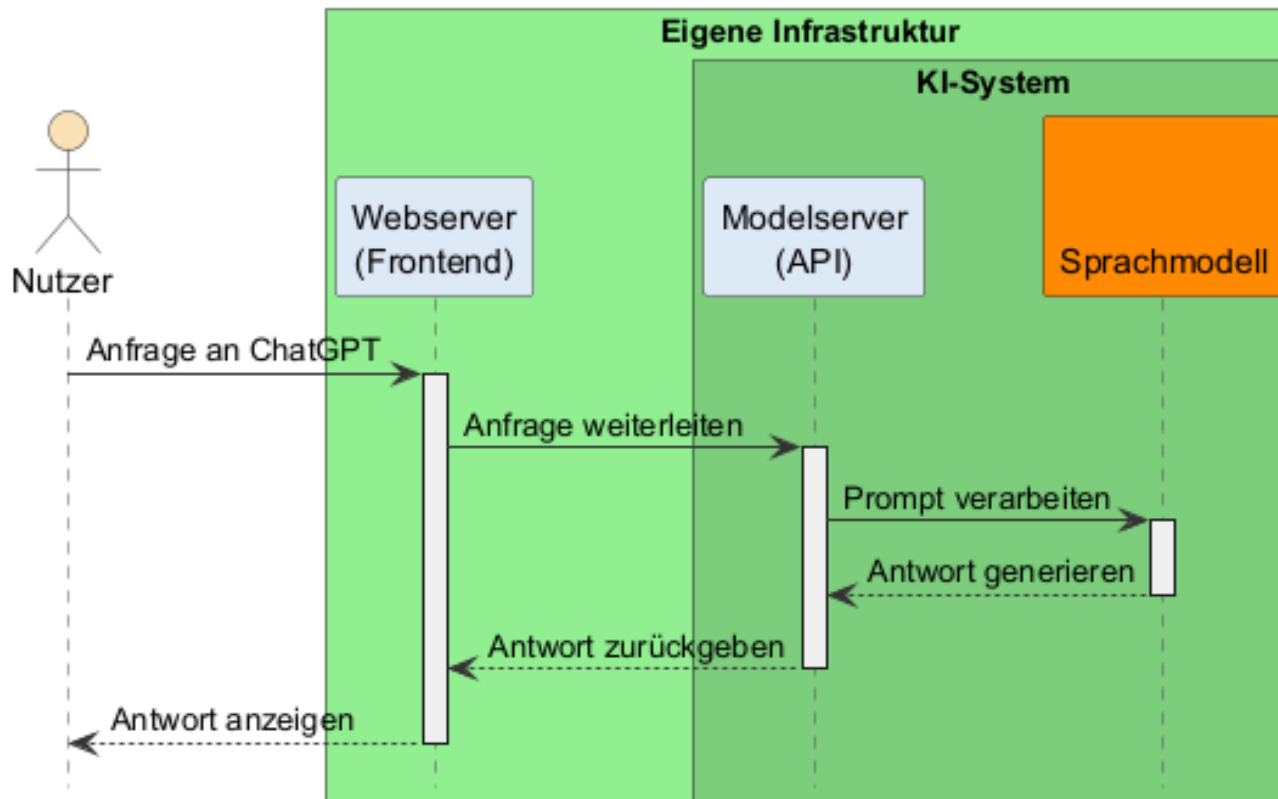


ALTERNATIVER SYSTEMAUFBAU 1



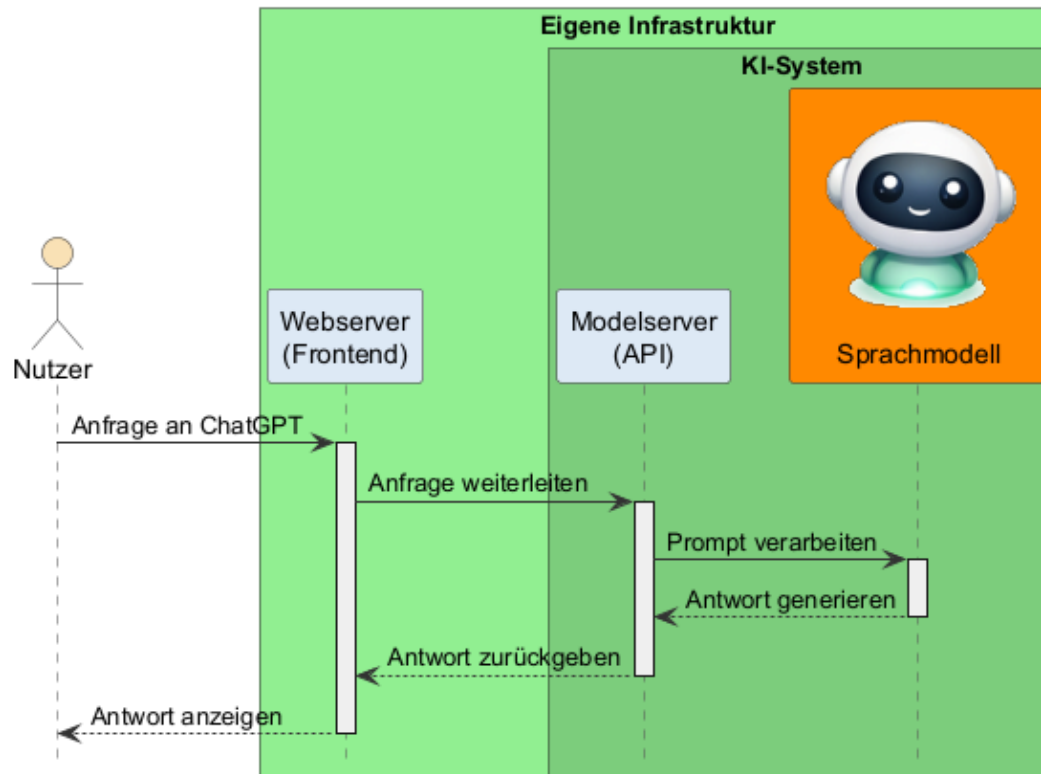
ALTERNATIVER SYSTEMAUFBAU 2

Anfrage an ChatGPT - Sequenzdiagramm



ALTERNATIVER SYSTEMAUFBAU 2

Anfrage an HuggingFace LLM - Sequenzdiagramm

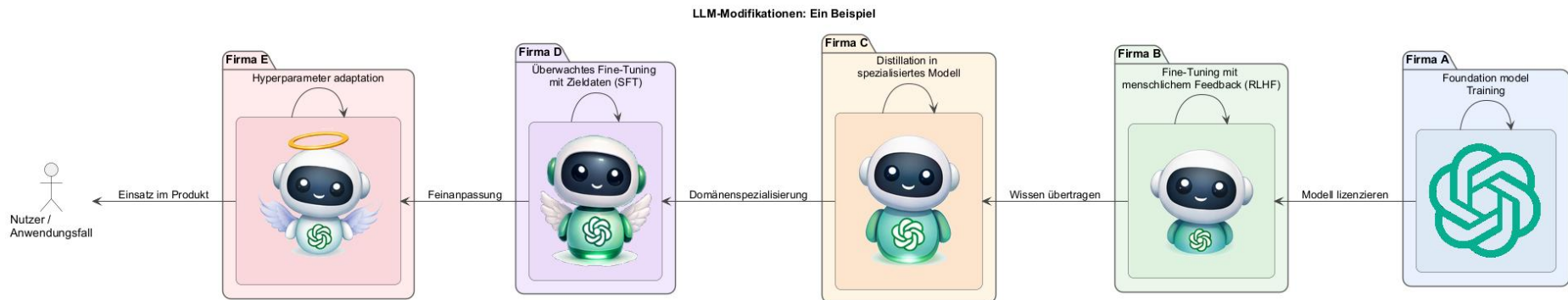


WIE KOMMEN DATEN INS LLM?

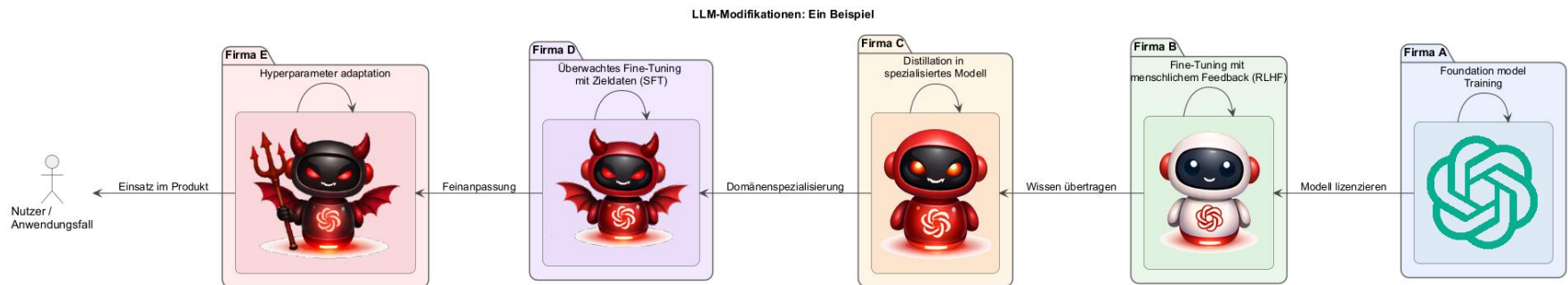
Verschiedene Stufen der Modifikation:

- Originalmodell
- Hyperparameter tuning
- Domänenspezifische Anpassung
- Fine-tuning (SFT/RLHF/RLAIF)
- Distillation

...oder eine Kombination dieser Techniken.



WIE KOMMEN DATEN INS LLM?



Jeder Schritt in der Modifikationskette kann problematisch sein und Bias einführen!

WAS KANN MAN TUN?

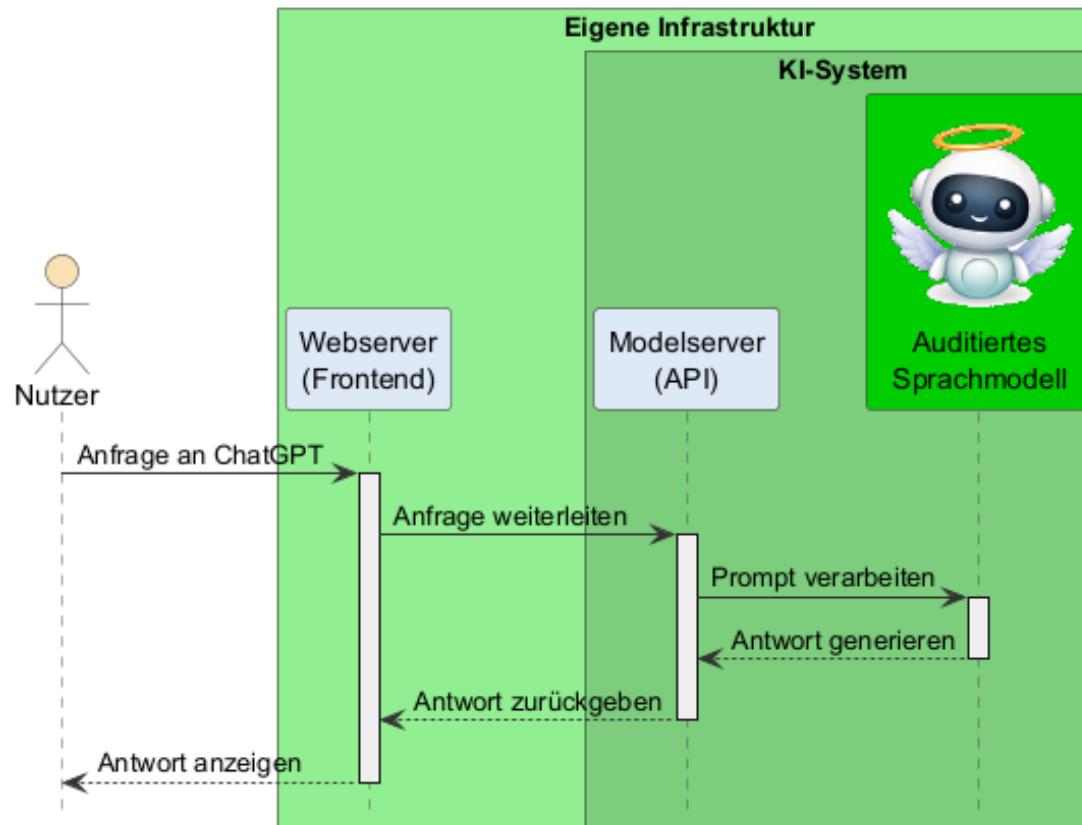
- Transparenz
- Dokumentation
- Ownership/Verantwortung
- X-AI (erklärbare KI)
- Auditierbarkeit



Gesetzliche Regelung: EU AI act

WIE SIEHT ES IDEALERWEISE AUS?

Anfrage an eigenes LLM - Sequenzdiagramm



AUSBLICK: WORKSHOP



REFERENZEN

R. Gorwa, M. Veale: "Moderating model marketplaces: platform governance puzzles for AI Intermediaries", *Law, Innovation and Technology*, 16(2), 2024.

P. Hacker, M. Holweg: "The regulation of fine-tuning: Federated compliance for modified general-purpose AI models", *Computer Law & Security Review*, vol. 60, 2026.

A. Vee: "GPT-4chan: A pre-ChatGPT time capsule", *substack.com*, 2024.

M. Krauledat, R. Hartanto: „Robotik und künstliche Intelligenz“, *Denkwerkstatt 3*, 2019.

M. Hardt, E. Price, N. Srebro: „Equality of Opportunity in Supervised Learning“, <https://arxiv.org/abs/1610.02413>, 2016.

T. Miconi: „The impossibility of ‘fairness’: a generalized impossibility result for decisions“, <https://arxiv.org/abs/1707.01195>, 2017.

S. Barocas, M. Hardt, A. Narayanan: „Fairness and Machine Learning“, *fairmlbook.org*, <http://www.fairmlbook.org>, 2018.

A. Narayanan: “21 Fairness definitions and their politics”, *Conference on Fairness, Accountability and Transparency*, 2018.

HOCHSCHULE RHEIN-WAAL

Campus Kleve

Marie-Curie-Straße 1
47533 Kleve
Deutschland
Tel.: +49 2821 806 73 – 0

info@hochschule-rhein-waal.de
www.hochschule-rhein-waal.de

Campus Kamp-Lintfort

Friedrich-Heinrich-Allee 25
47475 Kamp-Lintfort
Deutschland
Tel.: +49 2842 908 25 – 0



Folgen Sie uns auf Twitter:
www.twitter.com/HochschuleRW



Werden Sie Fan auf Facebook:
www.facebook.de/hochscherheinwaal