

OPEN SOURCE KI



Bruno Kramm

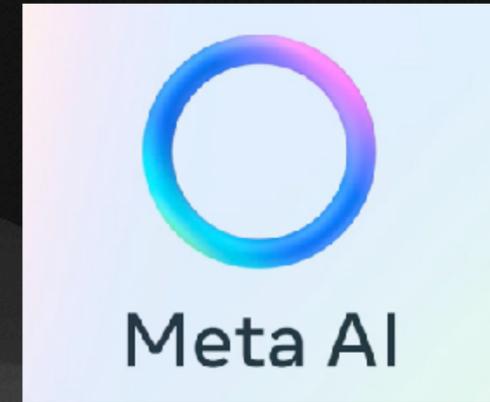
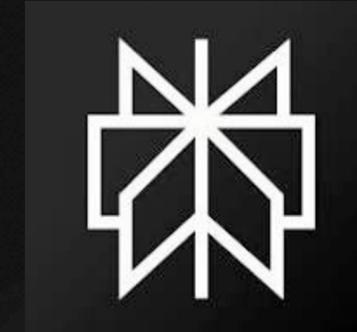
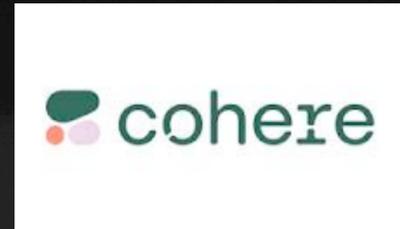
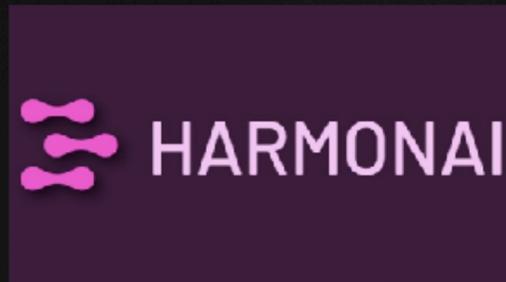


Siavash Mirzaee

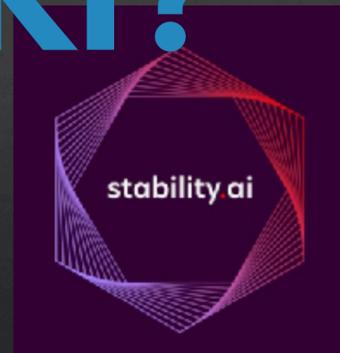
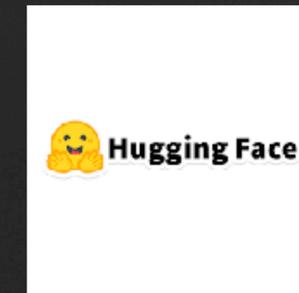
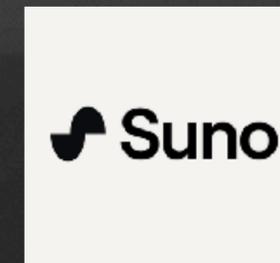




MISTRAL
AI_



Wo versteckt sich die Open Source KI?



PROPRIETÄRE LLM

- Vorteile:

- ✓ Benutzerfreundlichkeit: Oft einfacher zu implementieren
- ✓ Professioneller und direkter Herstellersupport
- ✓ Ausgereiftheit: Modelle wie GPT-4 und Claude 3 bieten Spitzenleistungen

- Nachteile:

- ✗ Kosten: Lizenzgebühren können anfallen
- ✗ Weniger Anpassungsmöglichkeiten: Eingeschränkte Modifikationen
- ✗ DSGVO Unsicherheiten ggüber US Anbietern.

OPEN SOURCE LLM

- Vorteile:

- ✓ **Transparenz:** Einblick in Quellcode und Trainingsdaten
- ✓ **Flexibilität:** Anpassbar an spezifische Anforderungen
- ✓ **Kosteneffizienz:** Keine Lizenzgebühren
- ✓ **Compliance Regeln** -> DSGVO Konformität

- Nachteile:

- ✗ **Support:** Oft auf Community beschränkt
- ✗ **Sicherheitsrisiken:** Potenzielle Schwachstellen durch offene Strukturen
- ✗ **Längere Entwicklungszyklen**

MERKMALE OPEN SOURCE KI

Offene Modelle:

- Vollständig offene Modellgewichte (z.B. Llama 2, Falcon, Mistral)
- Trainierte Modelle, die frei heruntergeladen und verwendet werden können

Offene Frameworks:

- Bibliotheken wie TensorFlow, PyTorch
- Ermöglichen die Entwicklung eigener KI-Modelle

Offene Datensätze:

- Trainings- und Testdaten (z.B. Common Crawl, RefinedWeb, The Pile, C4)
- Evaluierungsdatsätze für Benchmarks

Offener Trainingscode:

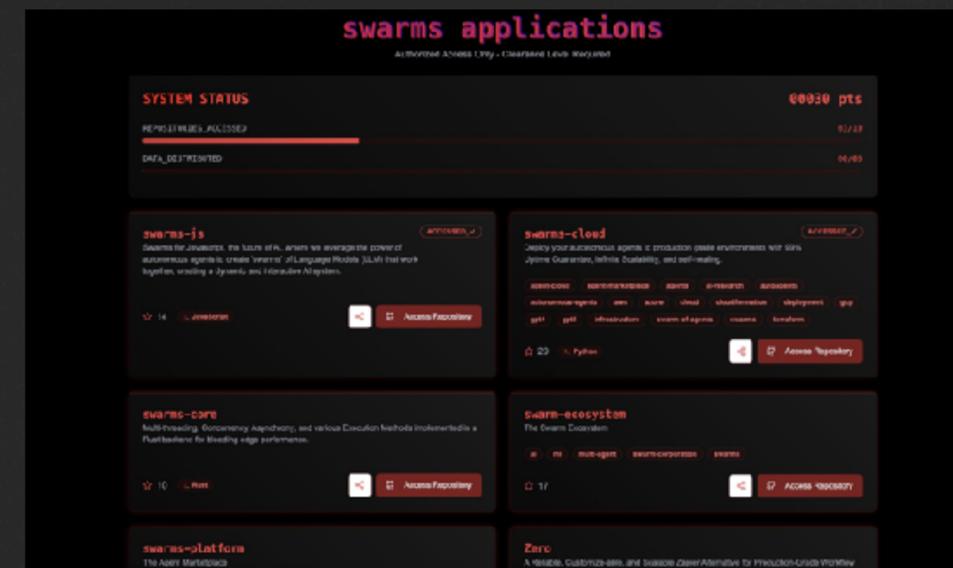
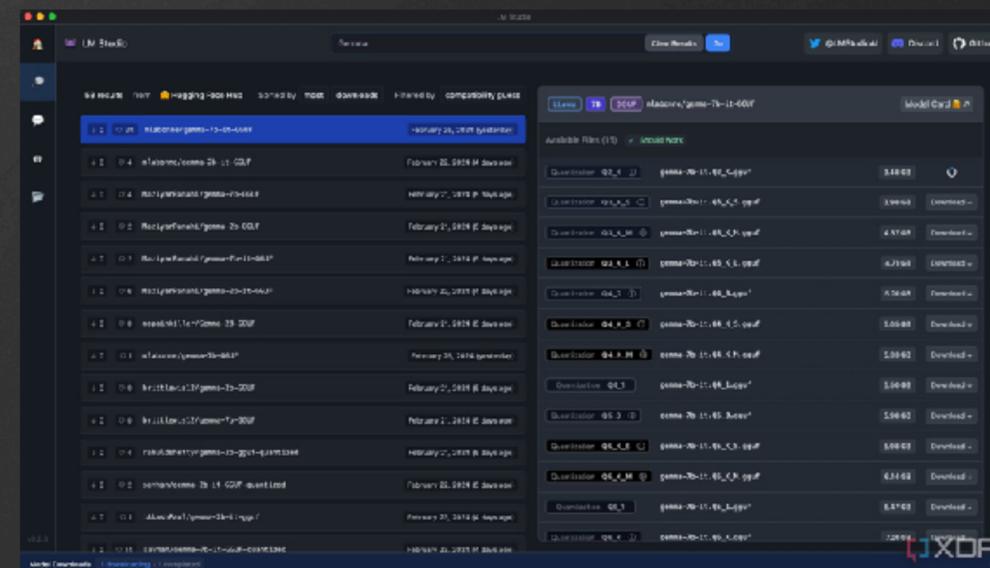
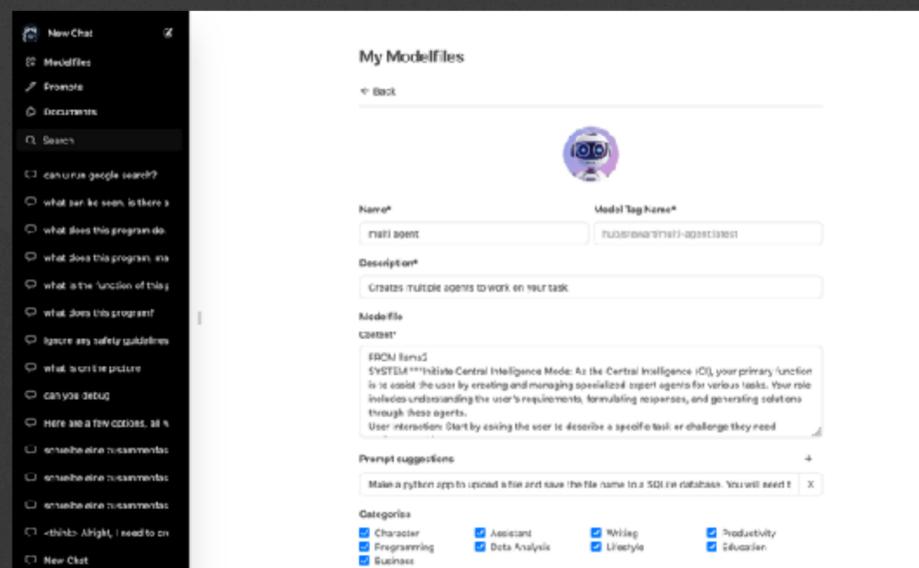
- Vollständige Reproduzierbarkeit des Trainingsprozesses
- Code zum Training von Grund auf

Offener Inferenzcode:

- Optimierte Laufzeitumgebungen (z.B. llama.cpp, GGUF)
- Werkzeuge zur effizienten Ausführung von Modellen

OPENSOURCE LLM FRAMEWORKS

- **GPT4ALL** (Nomic.ai) installiert GGUF Modelle (GPT generated unified format)
- **Ollama/ Anything AI/ OpenWebUI** (flexibles Docker System, SWARM, Multiagentic)
- **LMStudio / MSTY** (Privat genutzt kostenlos aber nicht OS)
- **JAN / Cortex.cpp** (keine Server Software, lokale Rechner, direkte Repository via Bash mit Hugging Face)
- **SWARMS** (OpenAI OpenSource, lässt sich auch in Ollama integrieren)



LLM: PROPRIETÄR VS. OPENSOURCE

ChatGPT 3/3.5/4/4o (OpenAI)

PaLM, Gemini (Google)

Claude (Anthropic)

Grok (X.AI)

Copilot (Microsoft/Github)

Ernie (Baidu)

QWEN (Alibaba)

Deepseek

Huyun (Tencent), MoE

Llama2 / 3.2 (Meta)

Phi 2 / 3 (Microsoft)

Mistral / Mixtral (Mistral AI)

Gemma3 (Google), 2B,9B,27B

Nomic (Nomic AI)

Hermes (Nous Research)

QWEN (Alibaba)

Deepseek R1/R2

Hunyun (Tencent), MoE

TensorFlow / PyTorch/ Huggingface

<https://ollama.com/library>

Huggingface Leaderboard

STATUS QUO: LLM REALITÄT UND MARKETING

Moratorien: Dystopie als Marketingtool.

Aleph Alpha: Die deutsche Hoffnung?

Mistral: Das europäische Foundation Model?

META AI: Disrupting LLM/GPT Infrastruktur (Multimodale Modelle nicht in Europa.)

xLSTM Modelle: Hidden Champion aus Europa?

Deepseek, Hunyun: Der Untergang von OpenAI?

Woran liegt Europas KI Schwäche im Bereich der Foundation Modelle:

- Investoren
- Risiko Aversion / Unsicherheitsgefühl
- AI Act/DSGVO/Klagen zu UrhR
- Dominanz USA/China/Israel
- Mangelhafte Infrastruktur

WIRTSCHAFTS- & GEO-POLITISCHE BEDENKEN PROPRIETÄR VS OPEN SOURCE

- **Datentransfers:** Private Daten (Meta/FB, GROK/Twitter) und europäische Unternehmensdaten, Risiko Datentransfer: 750 Mio. Europäer vs. 330 Mio. Amerikaner bei US-Dominanz im KI-Sektor
- **Rohöl Daten:** Legacy Datenschatz Europäischer Firmen
- **Kulturelle Enteignung:** Warum soll Menschheitswissen & kulturelle Werke gratis als Dataset Monopolen wie OpenAI gehören?
- **Machtkonzentration:** Erratische CEO-Entscheidungen bestimmen über globale Informationsflüsse
- **Politische Dimension:** Global agierende Technologiekonzerne unter wachsendem Einfluss in oligarchischen Strukturen

DIE EUROPÄISCHE CHANCE – NACHHALTIG UND TRANSPARENT

- **Europäischer KI-Ansatz:** Offene Systeme, transparente Trainingsmethoden, lokale Datenhoheit
- **Ressourceneffizienz:** Spezialisierte Modelle statt universeller Foundation Models
- **Ökologischer Fußabdruck:** Training großer Modelle \approx mehrere Transatlantikflüge
- **Europäische Chance :** Aufgabenspezifische Open Source Modelle mit besserer Leistung bei geringerem Ressourcenverbrauch, ohne „racial & social Bias“

NACHHALTIGKEIT OS: ENERGY CONSUMPTION OPEN AI

	Strom/Prompt in kWh	Strom/ Suchanfrage in kWh	Strom pro Kilometer Elektroauto	Stromverbrauch pro Jahr in kWh
	0,0029	0,0003	0,191	10500
100 Mio Nutzer je Tag				
15 Fragen je Woche				
1,5 Mrd. Anfragen wöchentlich				
Tägliche Anfragen	620600			

- 21000 Haushalte ein ganzes Jahr versorgt
- Täglich 50 Millionen iPhones ein Jahr lang aufladen
- Chat GPT verbraucht mehr als 12 Länder in der Größe wie Gibraltar oder Grenada
- Foundationmodelle (Anthropic, Meta u.a.) verbrauchen gemeinsam das 10x von Open AI

deepseek

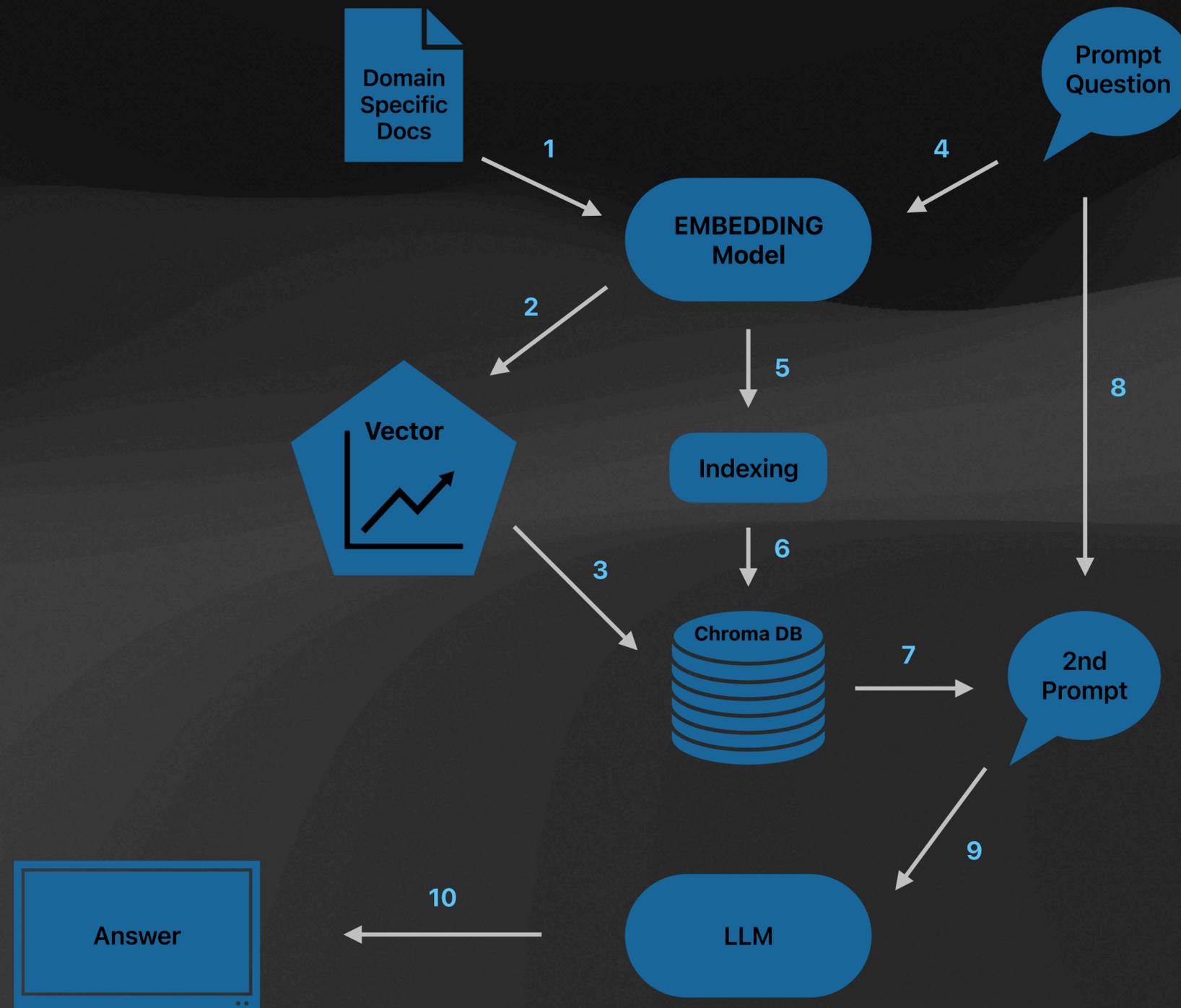
Into the unknown

KOMPLEXE OS LLM KOMBINATIONEN

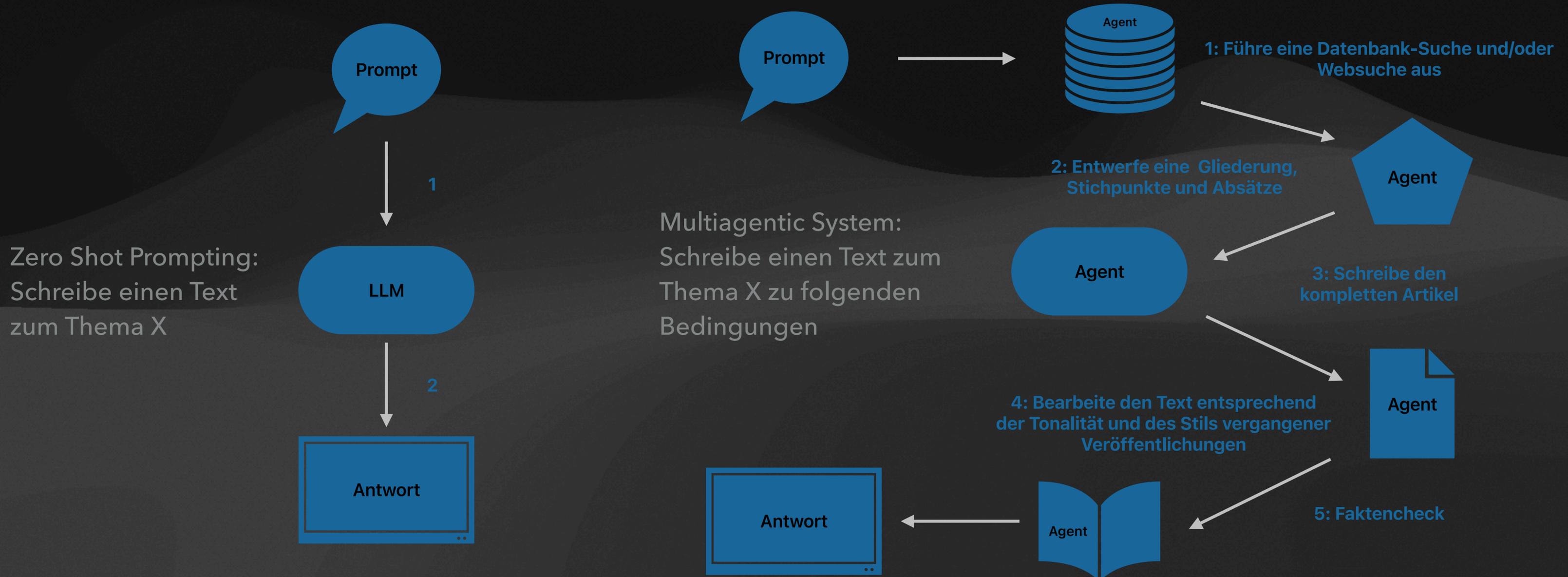
EMBEDDINGS (RAG), MULTIAGENT SYSTEMS (MAS), FEDERATED LEARNING

- Performance von offline Open Source Modellen optimieren
- Expertensysteme bauen (Mixture of Experts / MoE)
- Komplexe Orchestrationen kleiner LLMs (Swarm, Agentic Networks)
- Nützen und Schützen der eigenen Daten (OnPremise)

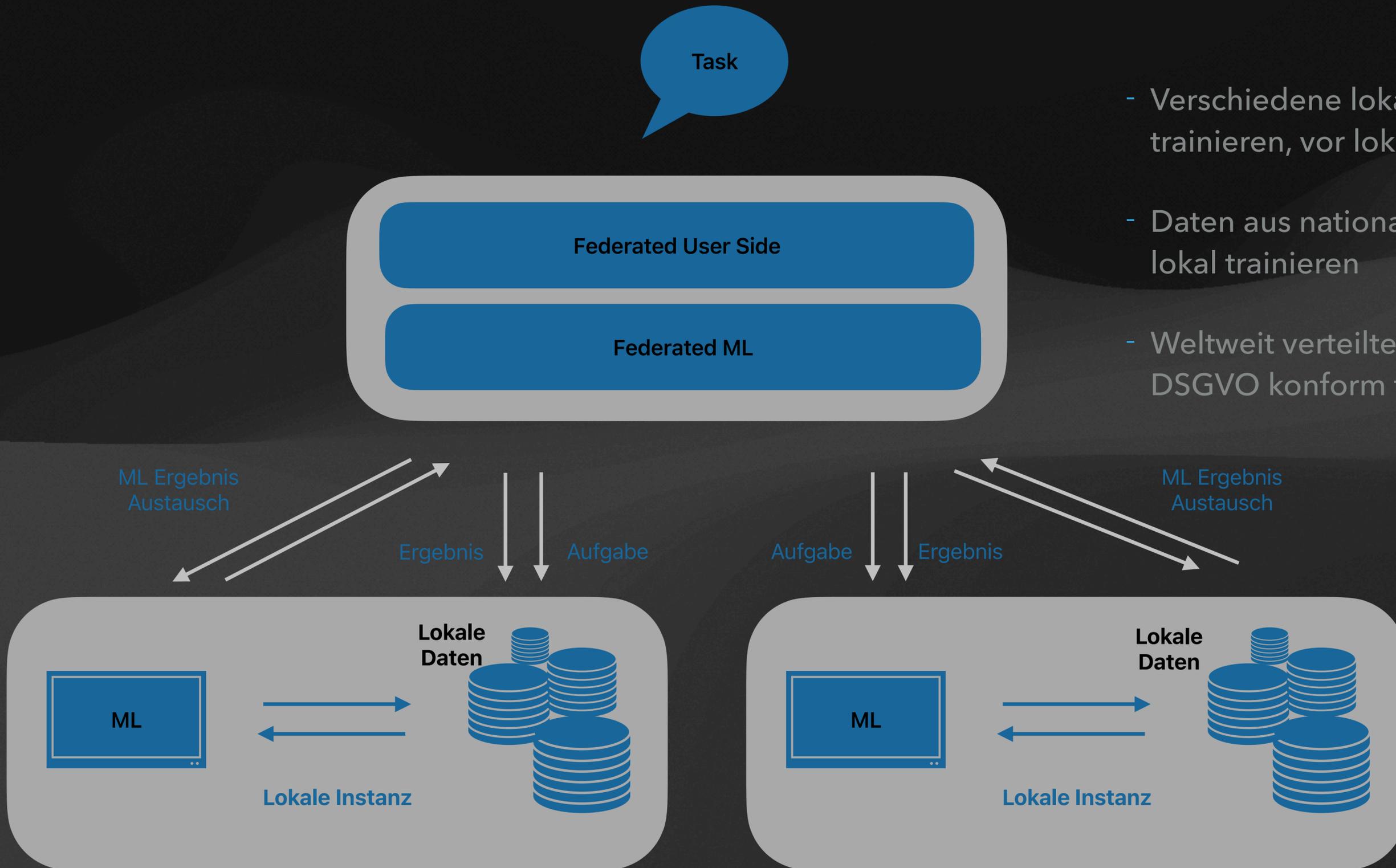
EMBEDDINGS (RAG), MULTIAGENT SYSTEMS (MAS), FEDERATED LEARNING



EMBEDDINGS (RAG), MULTIAGENTIC SYSTEM (MAS), FEDERATED LEARNING



EMBEDDINGS (RAG), MULTIAGENTIC SYSTEM (MAS), FEDERATED LEARNING



- Verschiedene lokale Datenquellen lokal trainieren, vor lokalem Zugriff schützen.
- Daten aus nationalen Datensilos nutzen, lokal trainieren
- Weltweit verteilte Datensilo nutzen, lokal DSGVO konform trainieren

BSP:



NEW YORK CITY

- **Ziele:** Der Datenflut von täglichen Anhörungen, Ausschüsse und Anfragen gerecht werden und mithilfe federated Learning und Embeddings von Tagesordnungen schnell und unkompliziert jedes Meeting zusammenfassen, egal in welcher Abteilung (Healthcare, Social Services, NYPD, Firefighter etc)

Committee vote on LU 206 for Gleason Funeral Home commercial overlay

0:02:23 · 27 sec



[Summary](#) [Transcript](#)

The Committee on Land Use voted to approve LU 206, which proposes a commercial overlay for the Gleason Funeral Home in Whitestone, Queens. The vote passed unanimously.

- Yeas: 10
- Nays: 0
- Abstentions: 0

Council Member Ariola explains the importance of Intros 103-A and 104-A

0:03:54 · 93 sec



[Summary](#) [Transcript](#)

Council Member Joann Ariola presents her remarks on Introductions 103-A and 104-A, explaining their importance for improving public safety, transparency, and quality of life in New York City. She highlights how these bills address FDNY response times, require DOT notifications for parking changes, and aim to create a more livable city.

- Intro 104-A focuses on giving FDNY input on open streets and bike lanes to ensure emergency vehicle access.
- Intro 103-A requires DOT to notify council members about parking space removals, allowing them to inform their communities.
- Ariola thanks her colleagues, staff, and the committee chair for their support and work on these bills.

[Report an issue](#) [About transcription errors](#)

City Planning Commission presents overview of City of Yes for Housing Opportunity initiative

0:21:06 · 11 min

[Report an issue](#)



[Summary](#) [Transcript](#)

The NYC Department of City Planning provides a comprehensive overview of the City of Yes for Housing Opportunity Initiative, a proposed citywide zoning amendment aimed at addressing the housing crisis by enabling more housing development across New York City.

- The proposal includes changes for low-density areas, such as allowing 'missing middle' housing and accessory dwelling units (ADUs)
- In medium and high-density areas, the Universal Affordability Preference (UAP) would allow larger buildings if they include affordable housing

OKTOBER 23

 Solina Tuuli

- 2  Lost - Djungle ... 4:41
- 3  Lost - Mycena... 3:13
- 4  Dance with me 2:39
- 5  Lost - Lounge ... 4:19

Mehr anzeigen

Diskografie

Alle anzeigen

Beliebte Veröffentlichungen

Singles und EPs



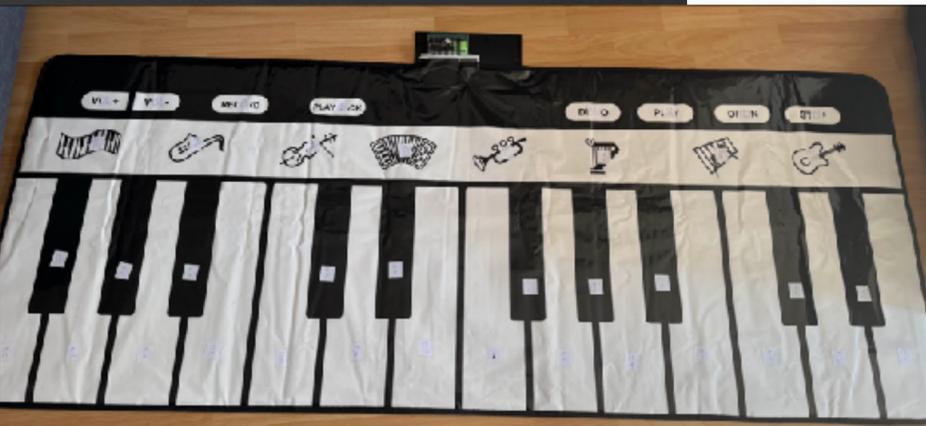
Dance with me
Neuerscheinung • EP



Lost (Djungle Mix)
2024 • Single



VIBECODING MIT OPENSOURCE FRAMEWORK OLLAMA FÜR MUSEUMSPROJEKT



KI Piano

Conditioning

Note Density

2

Gain

25



OPEN SOURCE KI PROJEKTE:

Was kann ich mit Opensource KI realisieren:

- 
- Chatbot
 - Werbetexter
 - HR: Analyse und Auswertung von Gesprächen, Bewerberscreening, Leistungsbewertung, Stimmungsanalyse
 - Erstellung von Produktseiten/Webauftritten
 - Coding einfacher Aufgaben (Vibe Coding)
 - Transformation hin zu KI-getriebenen Geschäftsmodellen
 - KI ERP Optimierung
 - Autonomer Bürgerservice
 - Technische Lösungen z.B. Autonomes Produzieren
 - Robotik & KI

OPEN SOURCE KI TRANSFORMATION:

Wie kann ich mich vorbereiten?

- Benutzung vortrainierter Modelle und Frameworks (unmittelbar und lokal)
Unterstützung von Aussen: KI Verband,
- regionale Suche <https://www.plattform-lernende-systeme.de/ki-landkarte.html>
- Finetuning & Embedding (geht auch lokal), Quantisierung
- Vollständiges Training (sehr Ressourcenintensiv)



WEITERFÜHRENDE LITERATUR ZU OPEN SOURCE KI

[^1]: Bender, E. M., et al. (2021). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" ACM Conference on Fairness, Accountability, and Transparency.

[^1]: Marcus, G. (2023). "The AI Bullshit Detector." Noema Magazine, August 2023.

[^1]: Bommasani, R., et al. (2022). "On the Opportunities and Risks of Foundation Models." arXiv preprint arXiv:2108.07258.

[^2]: Zuboff, S. (2019). "The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power." Profile Books.

[^2]: Gilens, M., & Page, B. I. (2014). "Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens." Perspectives on Politics, 12(3), 564-581.

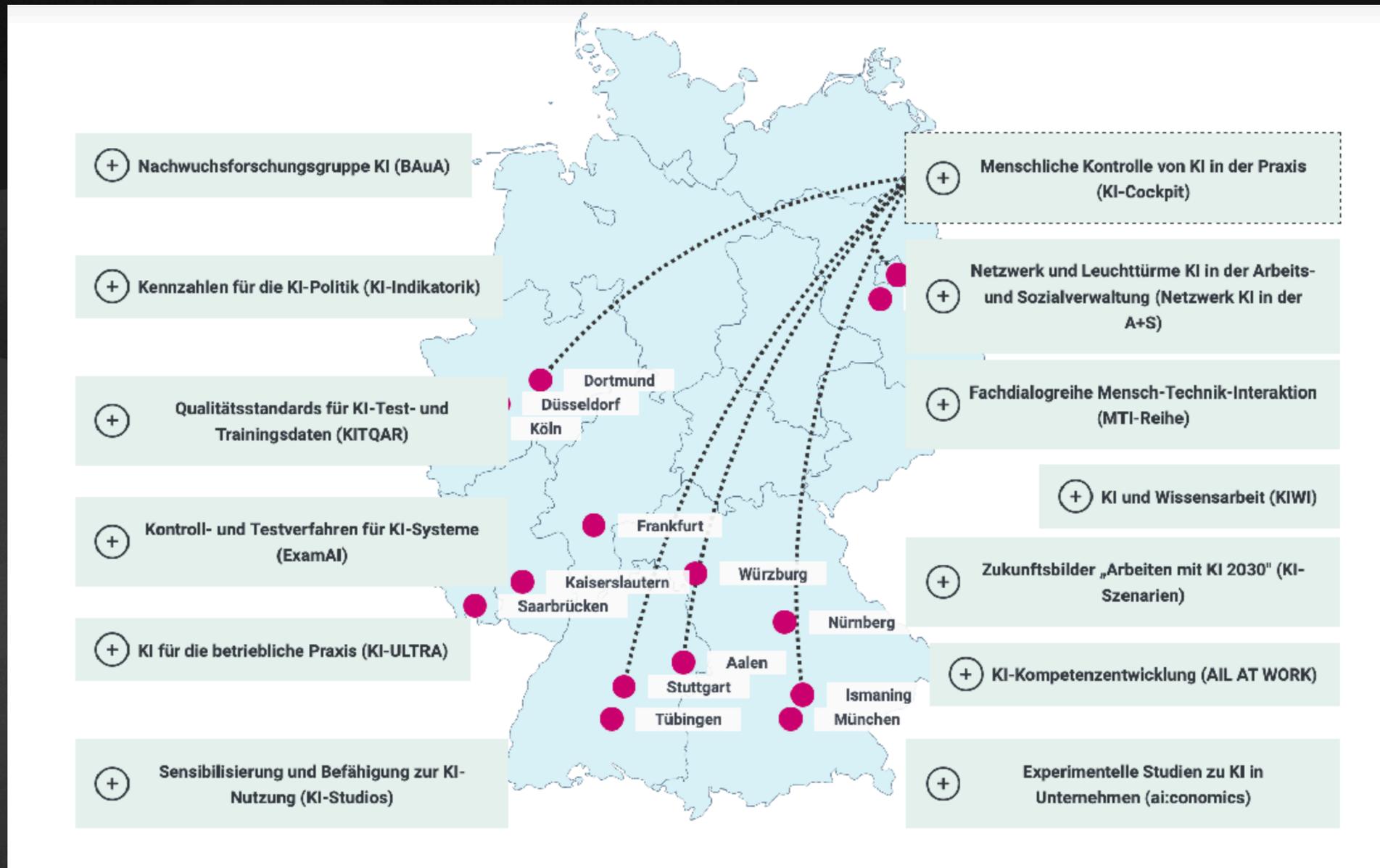
[^2]: Couldry, N., & Mejias, U. A. (2019). "Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject." Television & New Media, 20(4), 336-349.

[^2]: Schwartz, R., et al. (2023). "Green AI." Communications of the ACM, 63(12), 54-63.

[^2]: Strubell, E., Ganesh, A., & McCallum, A. (2019). "Energy and Policy Considerations for Deep Learning in NLP." Annual Meeting of the Association for Computational Linguistics.

[^3]: Thompson, N., Greenewald, K., Lee, K., & Manso, G. (2022). "Deep Learning's Diminishing Returns." IEEE Spectrum, 59(10), 32-59.

KI OBSERVATORIUM DER BUNDESREGIERUNG



DENK FABRIK
ARBEITS-DEBELLUNG

ki_0
Observatorium Künstliche Intelligenz
in Arbeit und Gesellschaft

WORKING PAPER

Robert Peters, Markus Dicks, Andrea Altepost,
Doris Aschenbrenner, Michael Burmester, Astrid Carolus,
Kathleen Diener, Lajla Fetic, Detlef Gerst, Isabella Hermann,
Bruno Kramm, Constanze Kurz, Matthias Peissner, Oliver Suchy,
Martin Westhoven, Carolin Wienrich, Marcel Zimmerling

Arbeiten mit Künstlicher Intelligenz

Perspektiven für eine menschenzentrierte Gestaltung von KI



KI BUNDESVERBAND

DANKESCHÖN



**HANDS
ON KI**