

Proprietäre vs. Open Source LLM

Was sich wann lohnt und worauf man achten sollte

Dr. Daryoush Vaziri & David Golchinfar

Netzwerk Mittelstand-Digital



Mittelstand-Digital

Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages



Stand: August 2023

Mittelstand-Digital Zentrum Fokus Mensch

Gemeinsam die digitale Zukunft menschzentriert gestalten.


Menschzentrierte Innovation und Gestaltung

 Fraunhofer Institut für Arbeitswirtschaft
und Organisation

Wohlbefinden und Resilienz

 Hochschule der Medien Stuttgart

Fairness und Diversität

 Technische Universität Chemnitz
Technische Universität Berlin
BITMi e. V.


Digitale Souveränität

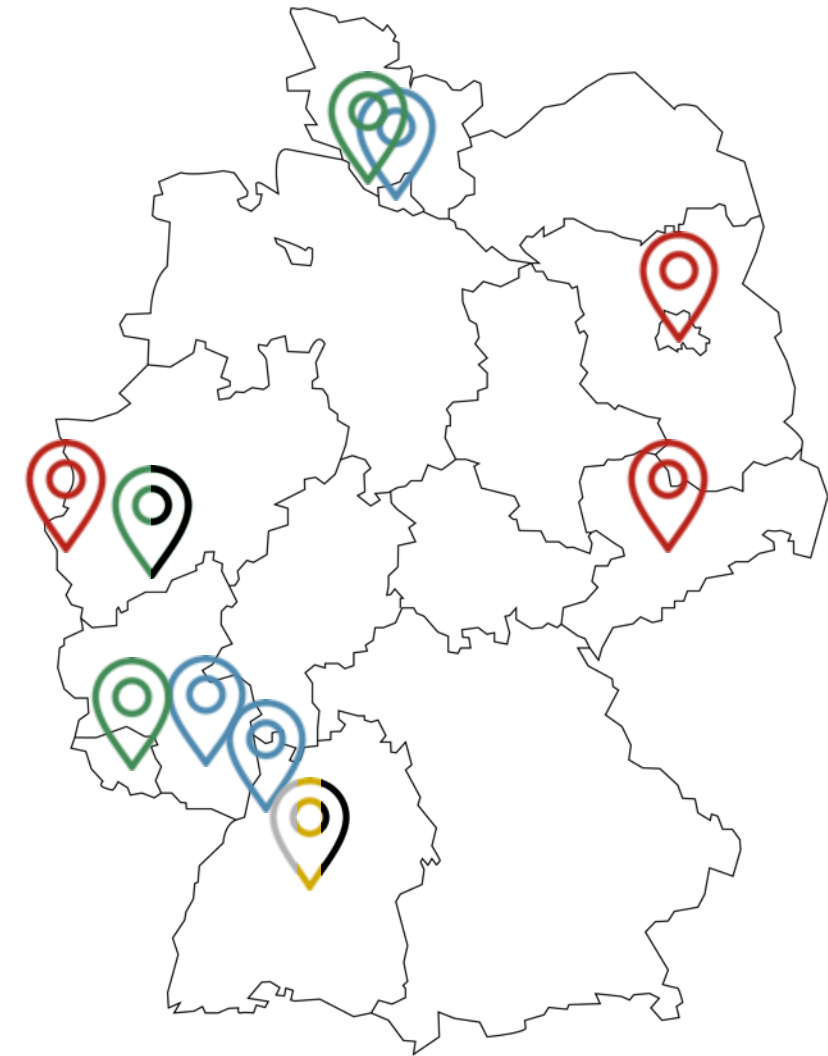
 Hochschule Bonn-Rhein-Sieg
NORDAKADEMIE
Ergosign

Nachhaltige Transformation

 Universität Hamburg
Usability in Germany e. V.
Karlsruher Institut für Technik
Hochschule Kaiserslautern

Künstliche Intelligenz

 Hochschule Bonn-Rhein-Sieg
Hochschule der Medien Stuttgart



Grafik angelehnt an Bruce Jones Design Inc. auf freeusandworldmaps
Icons von Freepik und Mayor Icons auf Flaticon.com

Lernen und vernetzen



Ausschreibungen

Egal ob laufende Umfragen, mögliche Kooperationen oder Ausschreibungen für gemeinsame Projekte: Hier bleiben Sie diesbezüglich informiert.



Events

Egal ob externes oder eigenes Ereignis: wir stehen Ihnen durch unterschiedliche Angebote auf einer Vielzahl von Veranstaltungen zur Verfügung.



KI-Trainer Programm

Das KI-Trainer Programm unterteilt sich in unterschiedliche Themenschwerpunkte und Angebote rund um das Thema Künstliche Intelligenz.



Schulungen

Über diverse Schulungen in verschiedenen Formaten soll Wissen zu unterschiedlichen Themenbereichen vermittelt werden.



Vorträge

Durch unsere ExpertInnen erhalten sie spannende Impuls und Einblicke in die Projekte und Ergebnisse des Mittelstand-Digital Zentrums Fokus Mensch.



Workshops

Durch unsere Workshops werden Ansätze und Methoden der Menschzentrierten Digitalisierung erlebbar und erfahrbar aufbereitet.

Digitalisierungsdialog
Gemeinsam passendes
Angebot für Sie finden.

Kontaktieren Sie uns!

Digitalisierung in der Praxis erleben



Praxisprojekte

Durch unsere Praxisprojekte kann das Mittelstand-Digital Zentrum Sie bei Ihrem individuellen Vorhaben unterstützen. Erhalten Sie hier weitere Informationen zu den Rahmenbedingungen oder lassen Sie sich von bisherigen Projekten für Ihr Vorhaben inspirieren.

[Weitere Informationen](#)



Demonstratoren

Durch unterschiedliche Demonstratoren werden Innovationen im Bereich der menschenzentrierten Digitalisierung für Unternehmen erlebbar und erfahrbar gemacht. Erfahren Sie hier mehr zu den einzelnen Formaten und Möglichkeiten zur Vorführung, Einweisung und Begehung.

[Weitere Informationen](#)

Digitalisierungsdialog
Gemeinsam passendes
Angebot für Sie finden.

Kontaktieren Sie uns!

Erfahren Sie mehr!

Newsletter



Praxisprojekte

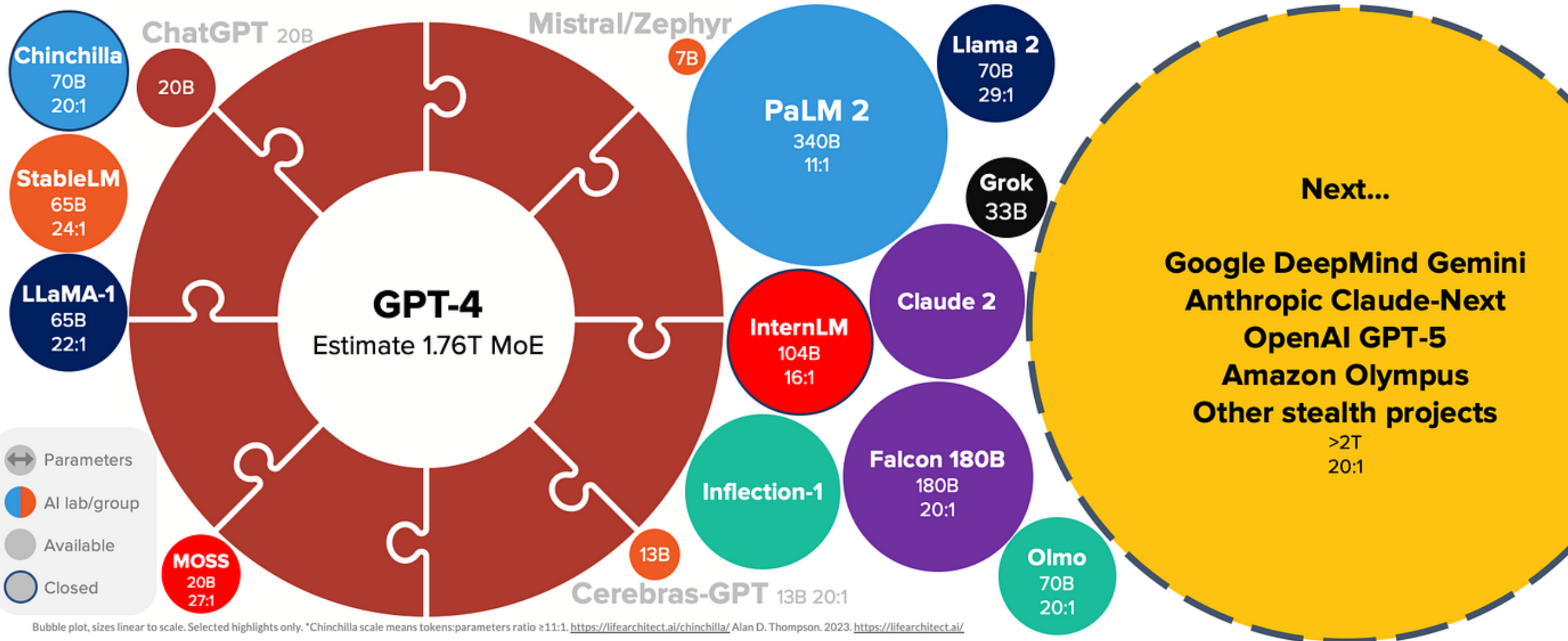




Los geht's

2023-2024 OPTIMAL LANGUAGE MODELS

NOV/
2023



Bubble plot, sizes linear to scale. Selected highlights only. *Chinchilla scale means tokens:parameters ratio $\geq 11:1$. <https://lilearnichillia.com/> Alan D. Thompson. 2023. <https://lilearnichillia.com/>

Proprietäre LLM

- Nicht öffentlich zugängliche LLM
- Merkmale
 - Exklusive Rechte
 - Kommerzielle Nutzung
 - Strenge Datenschutz- und Lizenzbedingungen
 - Hohe Leistungsfähigkeit
 - Je nach Nutzung, hohe Kosten



Open Source LLM

- Öffentlich zugängliche LLM
- Merkmale
 - Frei verfügbarer Quellcode
 - Entwicklung aus der Community heraus
 - Flexible Lizenzierung
 - Individuell anpassbar
 - Transparenz
 - Kosteneffizient

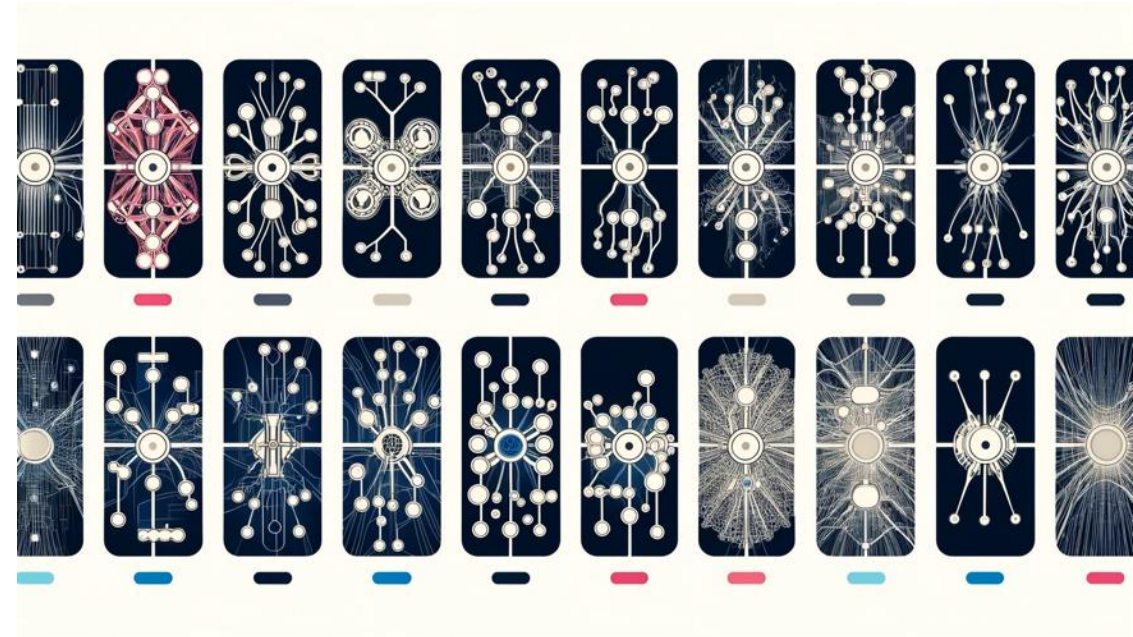


FALCON LLM



Unterscheidungskriterien

- Kontrolle und Transparenz
- Souveränität
- Kosteneffizienz und Prozessorientierung
- Anpassbar- und Erweiterbarkeit (Fine-Tuning)
- Fähigkeiten / Intelligenz / Smartness
- Performanz
- Nachhaltigkeit



(Daten-)Kontrolle und Transparenz

Proprietäre LLM

- Größtenteils Black Box
- Wenig valide Informationen zu Trainingsdaten
- Oft nur über Cloud-Dienste nutzbar
- Hohe Hürden für DSGVO Konformität



Open Source LLM

- Größtenteils White Box
- Informationen zu genutzten Trainingsdaten verfügbar
- On Premise Betrieb immer möglich
- DSGVO Konformität schneller erreichbar

Souveränität

Proprietäre LLM

- Schwer austauschbar, wenn in Prozesse integriert
 - Mit zunehmender Prozessintegration steigende Abhängigkeit von einem Anbieter



Open Source LLM

- Problemlos austauschbar
- Vollständig unabhängig vom Modellanbieter

Kosteneffizienz und Prozessorientierung

Proprietäre LLM

- Gebunden an Anbieterpreis
 - Kosten nehmen bei steigender Nutzung kontinuierlich zu
- Modellgröße je nach Prozessbedarf kaum ausgenutzt



Open Source LLM

- Kostenkontrolle durch eigene Hardware Spezifikation
- Sprungfixe Kosten
- Modellgröße kann an Prozessbedarf angepasst werden

Anpassbar- / Erweiterbarkeit (Fine Tuning)

Proprietäre LLM

- Anpassung in begrenztem Umfang möglich
- Anpassung oft mit Cloud-Diensten verbunden
- Anpassung mit sensiblen Unternehmensdaten kritisch



Open Source LLM

- Anpassung ohne Grenzen möglich
- Anpassung auch On-Premise möglich
- Sensible Unternehmensdaten können genutzt werden

Fähigkeiten / Intelligenz / Smartness

Proprietäre LLM

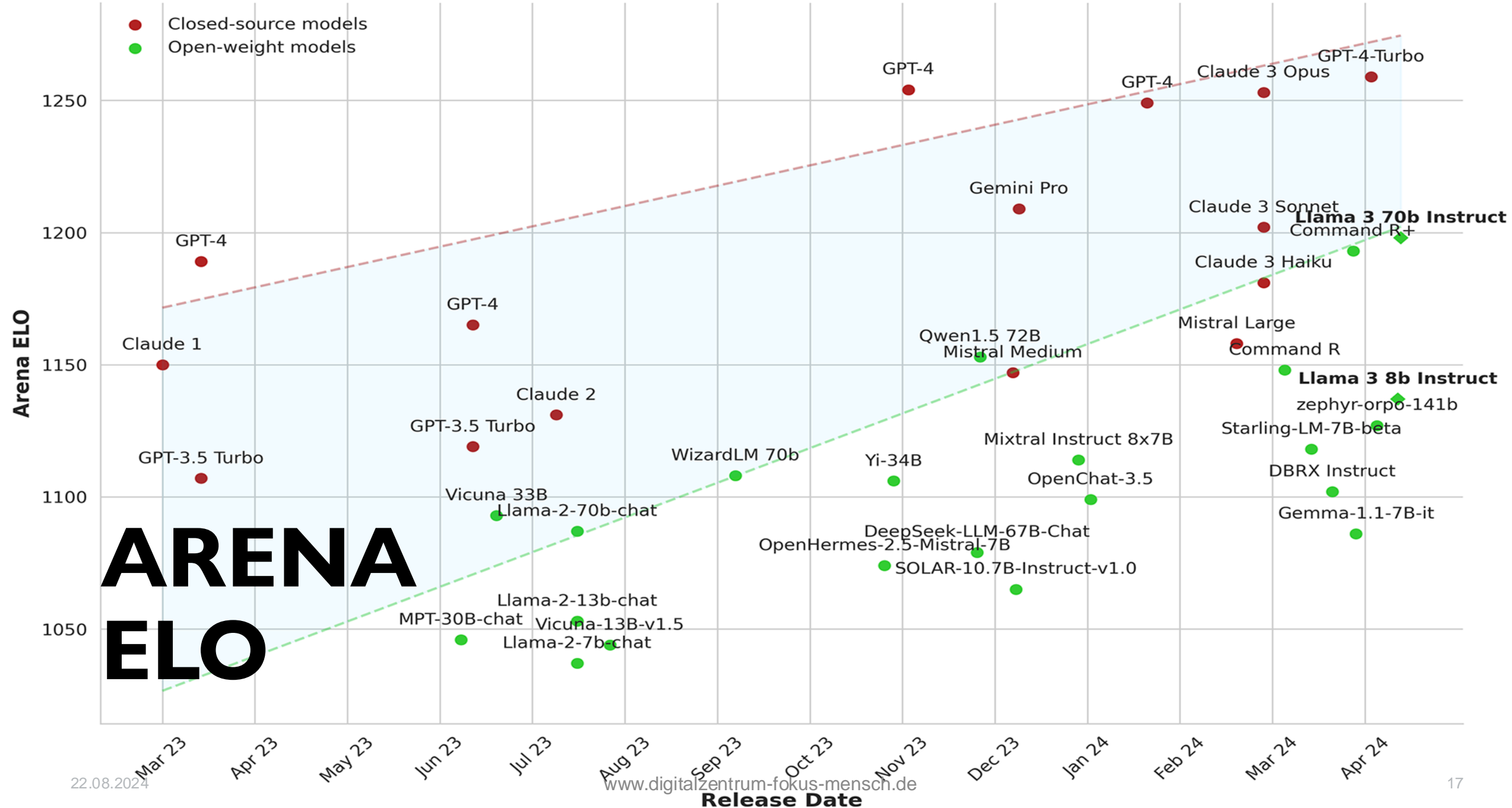
- Durch Vielzahl der Modellparameter in vielen Bereichen überlegen
- Finanzielle Ressourcen ermöglichen kontinuierliche Verbesserungen



Open Source LLM

- Modellintelligenz steigt rasant
- Mixtral bereits stärker als GPT 3.5
- Finanzielle Ressourcen junger Start-Ups begrenzt

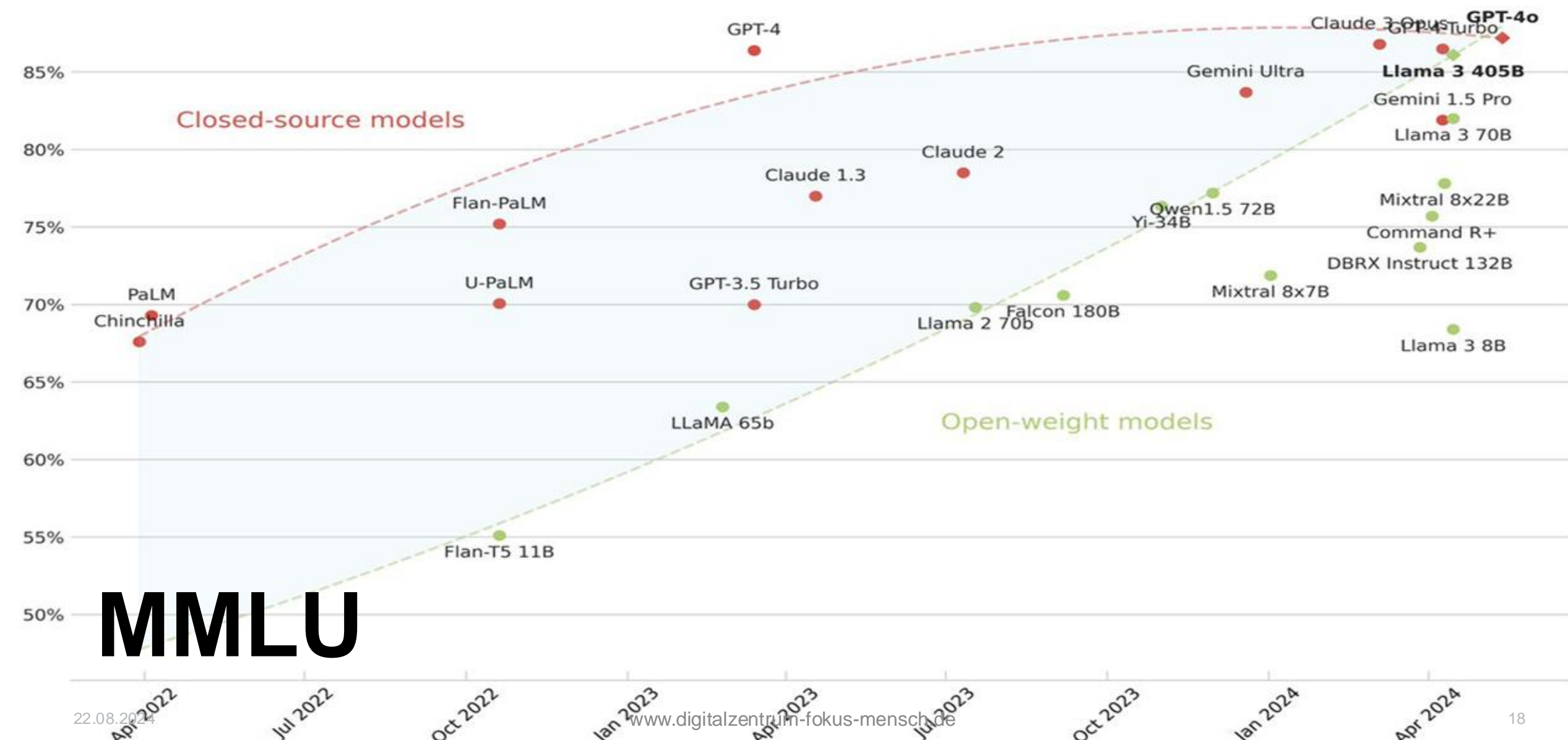
Closed-source vs. Open-weight models (Arena ELO, 22 Apr 24)



Closed-source vs. open-weight models

Llama 3 405B from Meta closes the gap between closed-source and open-weight models.

MMLU (5-shot)



22.08.2024

www.digitalzentrum-fokus-mensch.de

Performanz

Proprietäre LLM

- Durch hohe Anzahl an Modellparametern eher langsam
- Hoher Ressourcenbedarf zur Performanzsteigerung



Open Source LLM

- Höhere Performanz durch kleinere Modelle
- Ressourcenbedarf wesentlich geringer

Nachhaltigkeit

Proprietäre LLM

- Durch ihre Größe hohe Umweltkosten für Training und Betrieb
- GPT4 läuft auf mehreren Clustern mit jeweils 128 A100 GPU's mit 80GB VRAM
- Preis: 12.200 €

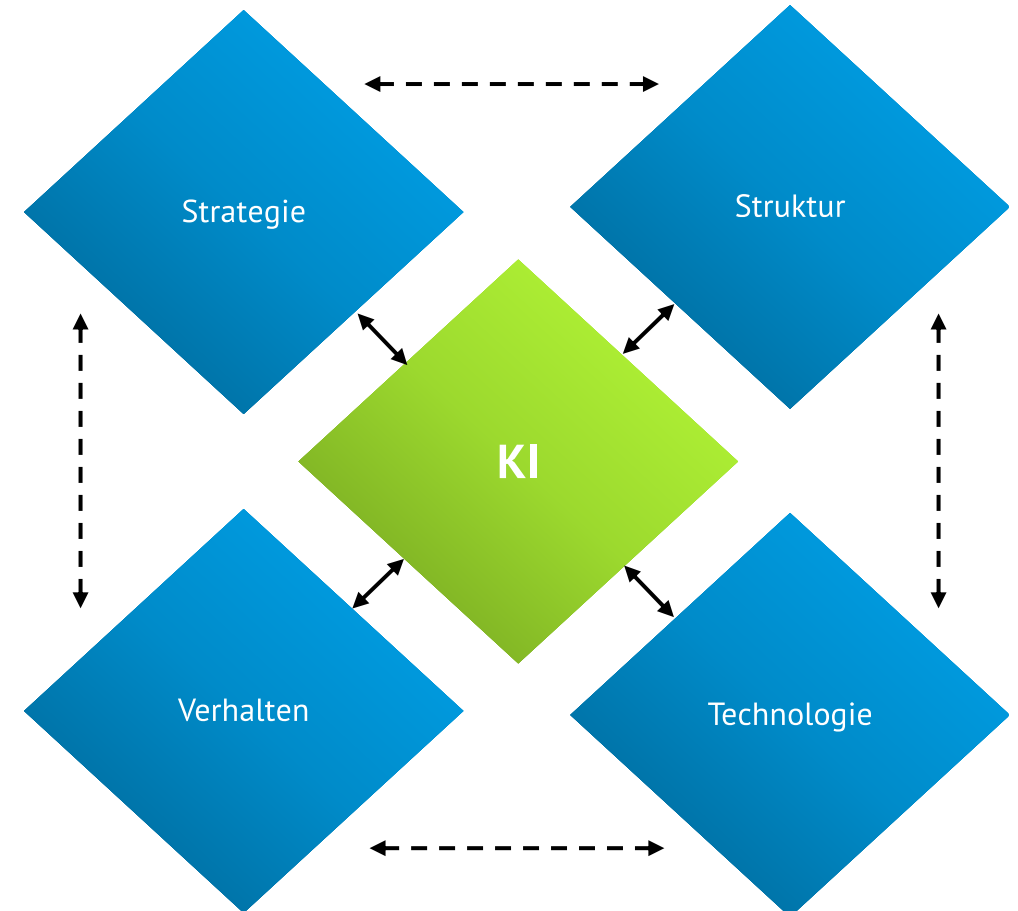


Open Source LLM

- Kleiner und dadurch weniger Umweltkosten für Training und Betrieb
- Ein 70b Modell passt quantifiziert auf eine RTX 3090 mit 24GB
- Preis: 1.200 €

Fazit

- Zahlreiche Faktoren bei LLM-Wahl zu berücksichtigen
- Essenziell ist ein geplantes und strukturiertes Vorgehen
- Dieses sollte u.a. folgende Themen umfassen
 - KI-Strategie
 - Prozessorientierung
 - Verhalten und Kompetenzen von Mitarbeitern
 - Rechtliche Rahmenbedingungen
 - Technologische Rahmenbedingungen



DANKE



Dr. Daryoush Vaziri
Leiter Digitale Souveränität und
KI-Trainer im Mittelstand-Digital
Zentrum Fokus Mensch

[d.vaziri@digitalzentrum-
fokus-mensch.de](mailto:d.vaziri@digitalzentrum-fokus-mensch.de)



David Golchinfar
KI-Trainer im Mittelstand-Digital
Zentrum Fokus Mensch

[d.golchinfar@digitalzentrum-
fokus-mensch.de](mailto:d.golchinfar@digitalzentrum-fokus-mensch.de)

Feedback-Runde

Bitte unterstützen Sie uns mit Ihrer Evaluation der heutigen Veranstaltung und hinterlassen Sie Ihr Feedback unter:

<https://form.typeform.com/to/nu30bgBI>

