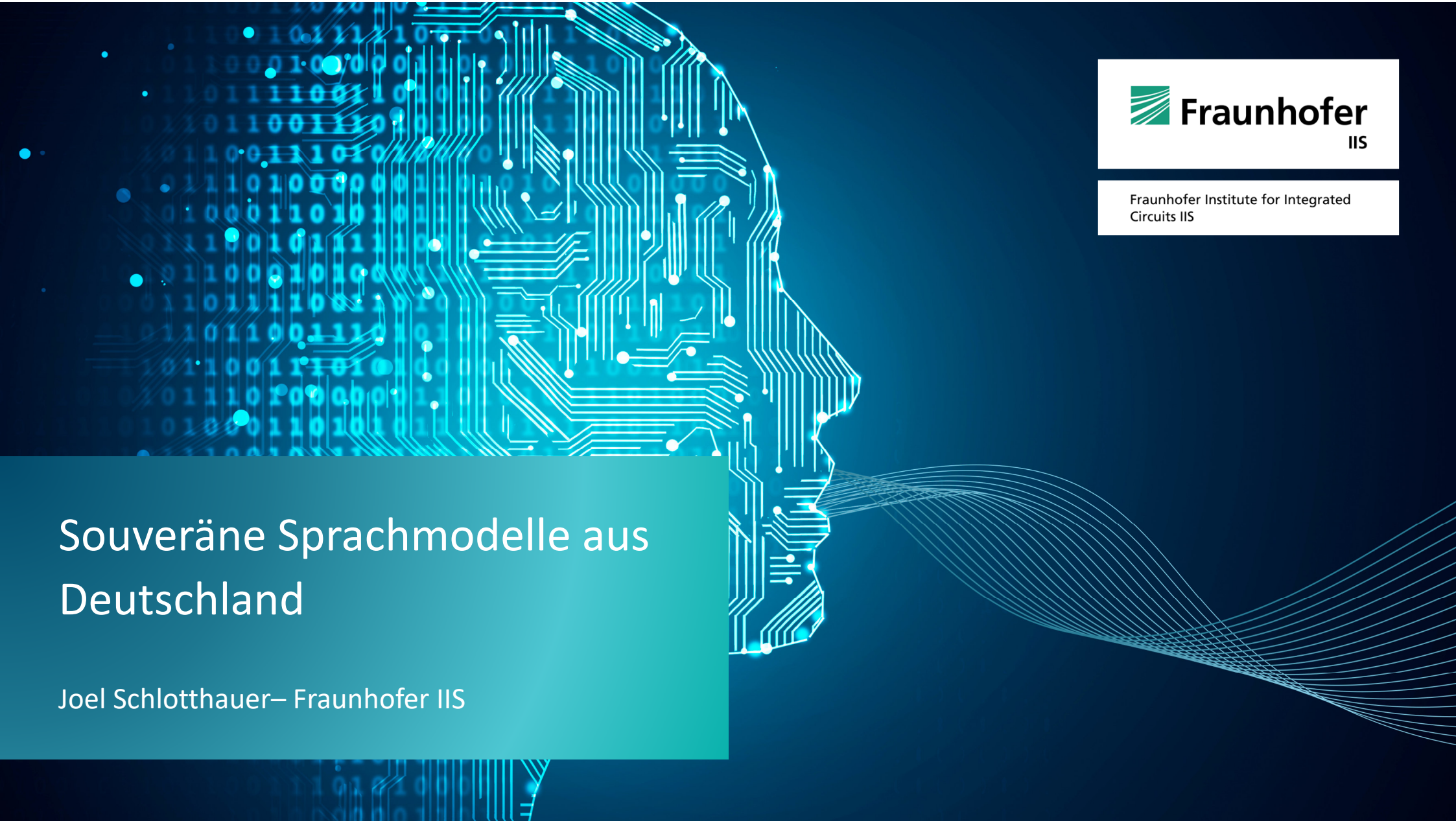


Fraunhofer IIS

Fraunhofer Institute for Integrated Circuits IIS

Souveräne Sprachmodelle aus Deutschland

Joel Schlotthauer– Fraunhofer IIS

Warum brauchen wir deutsche Sprachmodelle?

Souveränität, Wertschöpfung und technologische Basis

1. **Digitale Souveränität**
   Unabhängigkeit bei kritischen Technologien.

2. **Wettbewerbsfähigkeit**
   industrielle Wertschöpfung entlang der gesamten KI-Kette.

3. **Technologische Basis**
   verstehen, anpassen, betreiben und weiterentwickeln können.

4. **Angepasste Modelle**
   für deutsche und industrielle Kontexte.

84%

der Unternehmen, die generative KI nutzen oder es planen, geben an, dass das Herkunftsland des Anbieters „sehr wichtig“ oder „eher wichtig“ ist.

bitkom

Quelle: Bitkom Research 2024

Leistungsniveau auf Standard-Benchmarks vs. industrielle Anforderungen

80 %
Leistungsniveau erreicht

20 %
zur industriellen Reife

Zwei Modellklassen - Ein Kreislauf

Große offene Basismodelle

- An der technologischen Front
- Für komplexe, generalistische Aufgaben
- Erzeugen hochwertige synthetische Daten
- Destillieren und verbessern kleinere Modelle

Kleine spezialisierte Modelle

- In der industriellen Fläche
- Lokal, effizient, latenzarm, kosteneffizient
- Datennah – sensible Daten bleiben im Haus
- Ideal für Geräte, Edge und agentische Systeme

Basiskompetenz • Synthetische Daten • Distillation

Reale Anwendungen • Use Cases • Feedback

ELMOD 2.7B

Sprachmodell aus Deutschland

Prototyp lauffähig auf Smartphone Chip

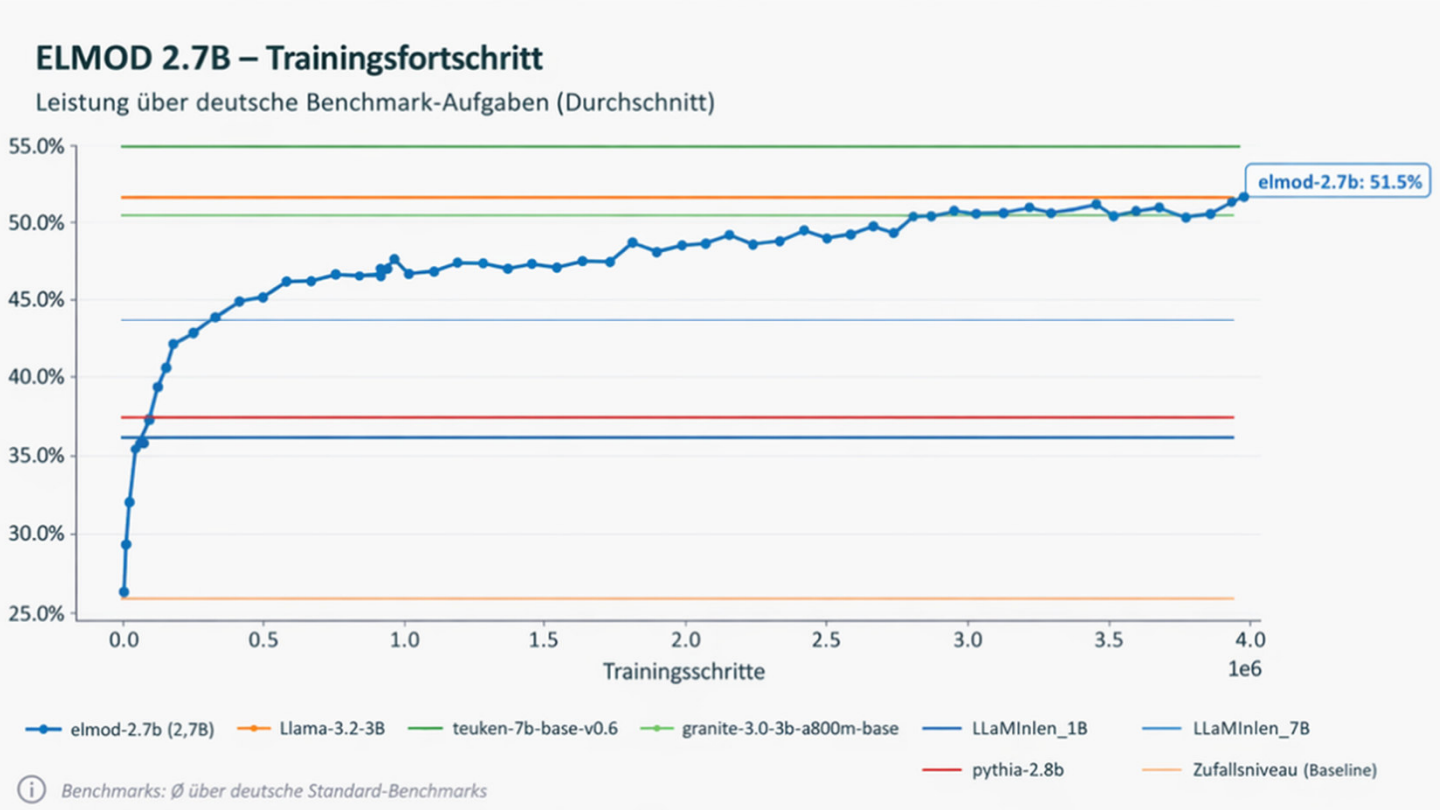
ELMOD 2.7B Referenzprojekt – Lokal-lauffähige Sprachmodelle aus Deutschland

TECHNIK

2,7 Mrd.
Parameter
trainiert auf 3,8T Tokens
(DE + EN + Code)

52k GPUh
auf H100 @ FAU Helma
End-to-end-Pipeline

5,5 PB
kuratierter deutscher
Rohdaten-Korpus



FORSCHUNGSPROTOTYP

ELMOD ist ein **Forschungsprototyp.**
Wir entwickeln heute die Grundlagen für souveräne KI-Modelle, die künftig an vielen Orten sinnvoll eingesetzt werden können.

ELMOD 2.7B – GenAI direkt in der Hosentasche

Referenzprojekt

ANWENDUNGSFALL — GEDANKENSPIEL

Der Servicetechniker beim Kunden.

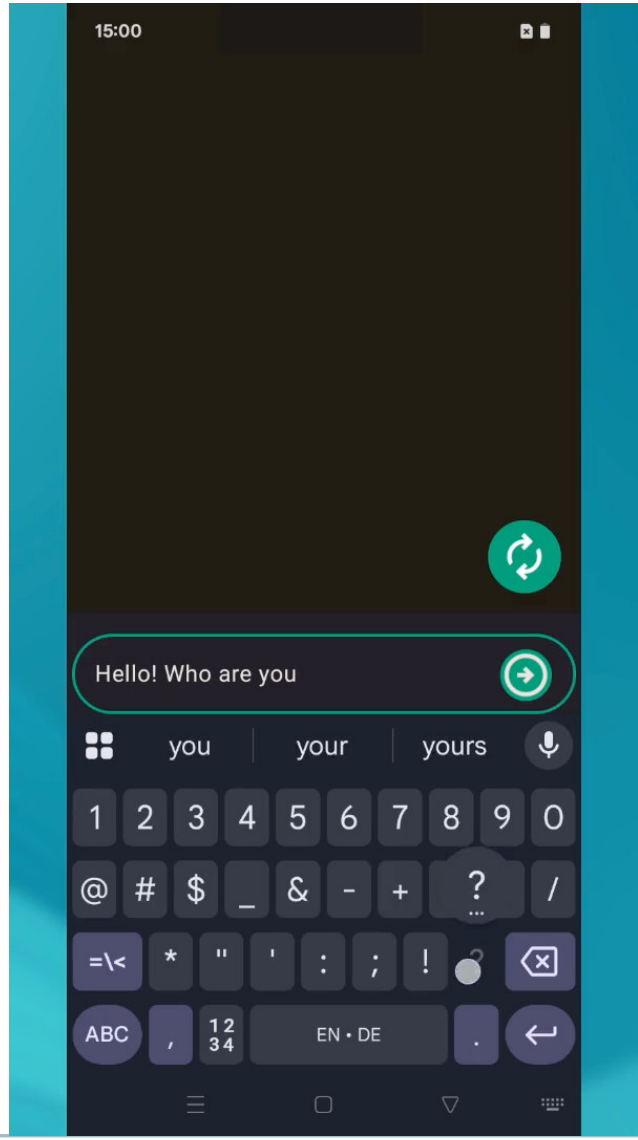
Ein mögliches Zukunftsszenario –
nicht der Entwicklungszweck von ELMOD.

Fehlerlogs | Sensordaten | Zeitreihen | Handbücher

Das Modell **schlussfolgert** über heterogene Quellen.

Offline. Sicher. Auf Deutsch. Datennah.

Lokale Ausführung



SOOFI

Europas Reasoning LLM

Ziel: Europäische KI-Souveränität

SOOFI – Europas Reasoning LLM

Referenzprojekt: Anschluss an die technologische Spitze

Bundesministerium für Wirtschaft und Energie

- SOOFI soll ein offenes KI-Sprachmodell mit rund **100 Mrd. Parametern** entwickeln
- Ziel: **europäische KI-Souveränität** — weniger Abhängigkeit von US-Anbietern.
- Gefördert mit ca. **21 Mio. €** durch das Bundesministerium für Wirtschaft und Energie
- **Reasoning-Modelle** — für komplexe Aufgaben in Industrie & Verwaltung

soofi

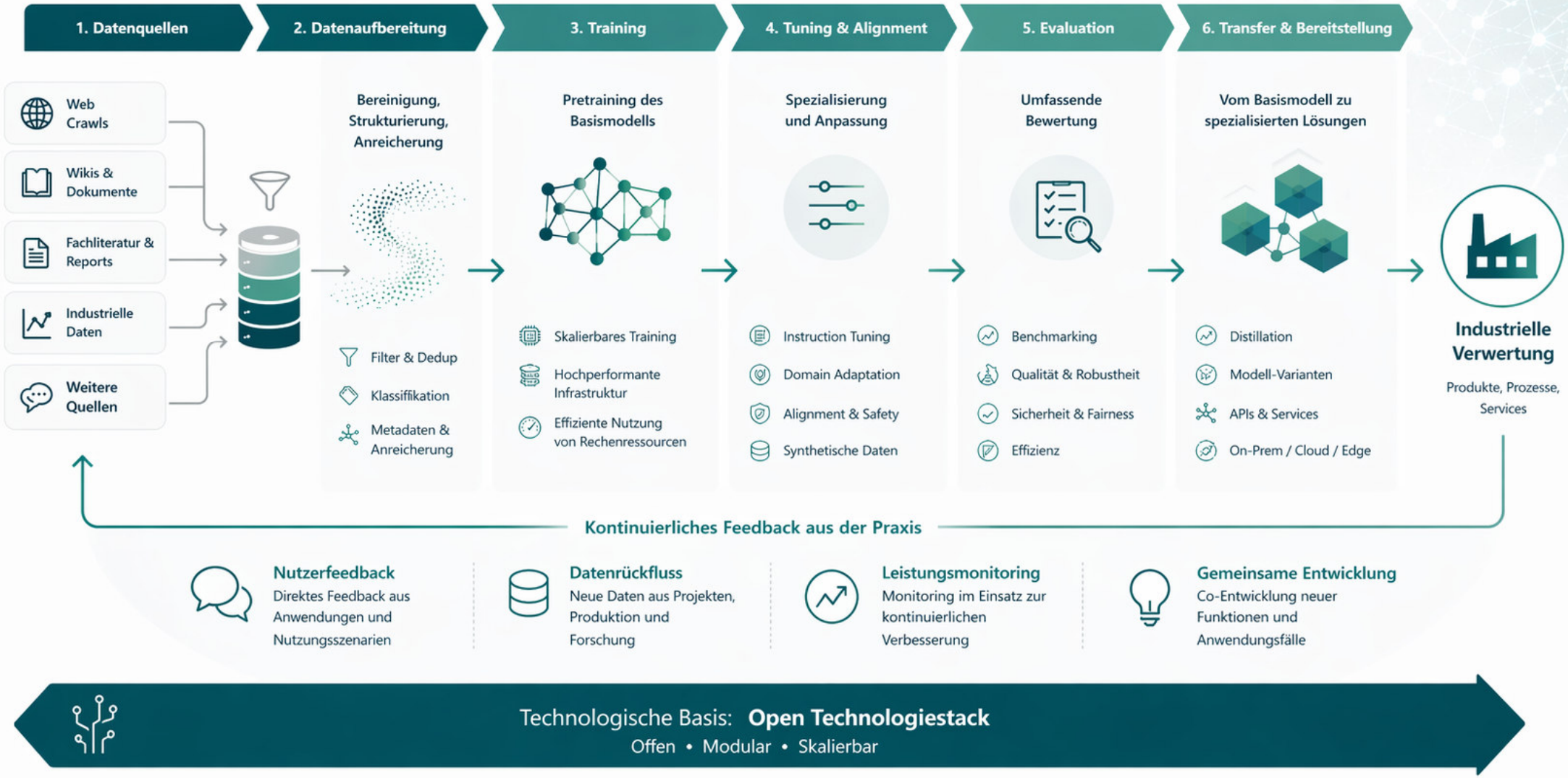
SOOFI – Europas Reasoning LLM

Referenzprojekt

- Training auf der neuen **Industrial AI Cloud der Deutschen Telekom**
- 130 NVIDIA DGX B200 Systemen mit insgesamt über **1.000 GPUs** exklusiv für SOOFI
- Seit **März 2026**

Die gesamte Kette zählt

- Von Daten bis zur Anwendung und zurück



Souveräne KI entsteht durch Kopplung von Forschung und Anwendung

- Große, generalistische Modelle **zunehmend möglich in DE.**
- Deutschland baut **souveräne Fähigkeiten und Infrastruktur** auf.

→ **Forschung ist die Grundlage.**

Im Zusammenspiel entsteht Wirkung.

Nutzung ist entscheidend. Industrielle Modelle für die Praxis.

- **Spezialisierte Modelle** für Industrie und Verwaltung.
- Zugang zu **Daten, Domänenwissen und Anwendungsschicht.**

→ **Anwendung ist der Hebel.**

Die letzten 20% entstehen in der Kopplung von Forschung, Industrie und gemeinsamer Infrastruktur.

Vielen Dank für Ihre Aufmerksamkeit!

Kontakt

Joel Schlotthauer
Gruppenleiter Natural Language Processing
Department Generative KI
joel.schlotthauer@iis.fraunhofer.de

Fraunhofer-Institut für Integrierte Schaltungen IIS
Am Wolfsmantel 33
D-91058 Erlangen

www.iis.fraunhofer.de