ALEPH ALPHA

Generative KI
Chancen und Risiken in der Nutzung

09 December 2023

Intro

Hans-Jörg Schäuble

VP Customer

Unsere Mission

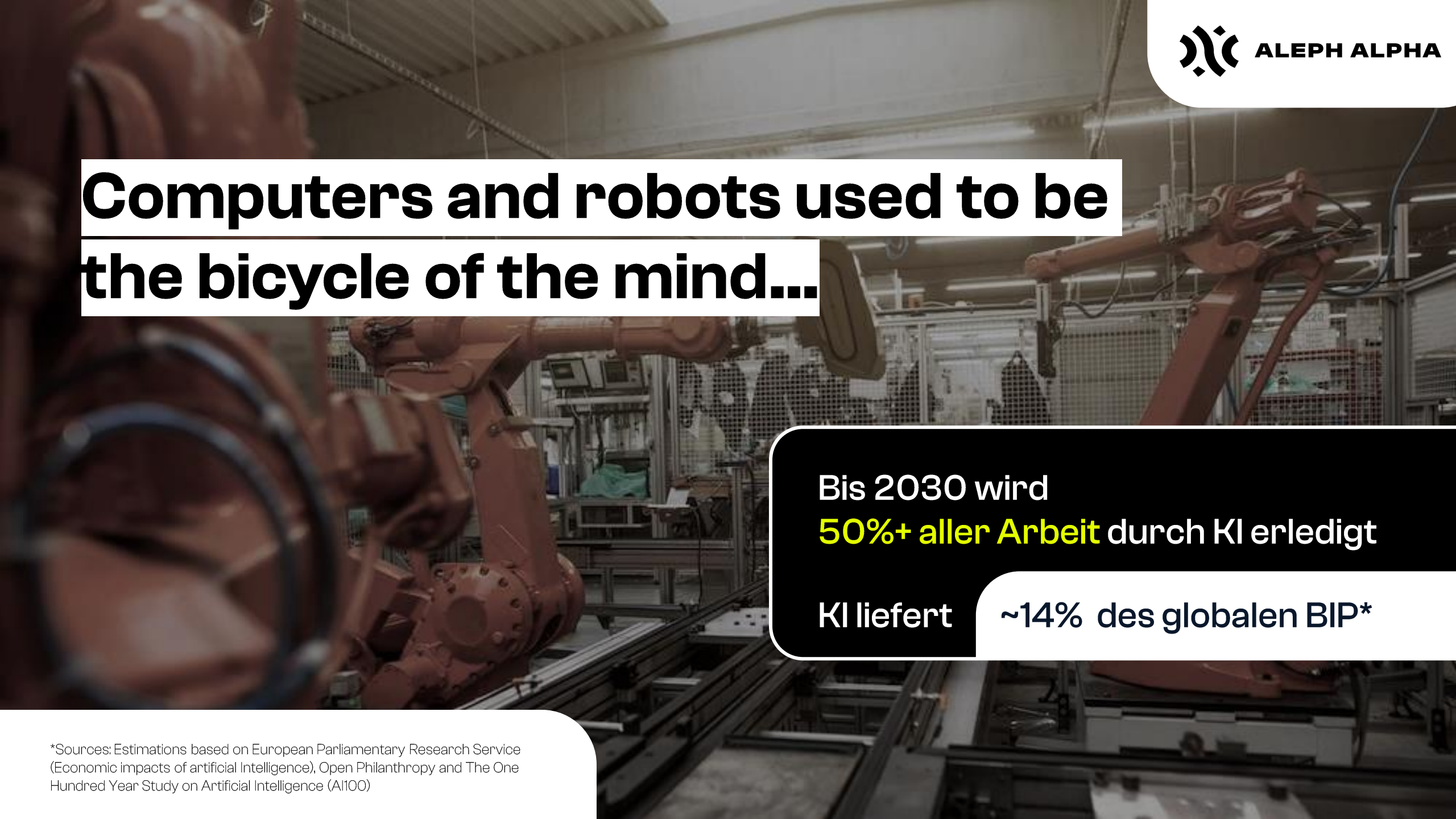
Europa als entscheidenden Akteur im Bereich der künstlichen Intelligenz zu etablieren und digitale Souveränität zu sichern.

Unser Kern

Wir sind ein unabhängiges Unternehmen, das generative KI-Sprachmodelle erforscht, entwickelt und in reale Anwendungen einführt.

Key Data

- Gegründet im Jahr 2019
- Standorte:
  - Heidelberg (Hauptsitz)
  - Berlin
  - Bayreuth (Rechenzentrum)
- Ca. 70 Mitarbeiter
- CEO: Jonas Andrulis



ALEPH ALPHA

Computers and robots used to be the bicycle of the mind...

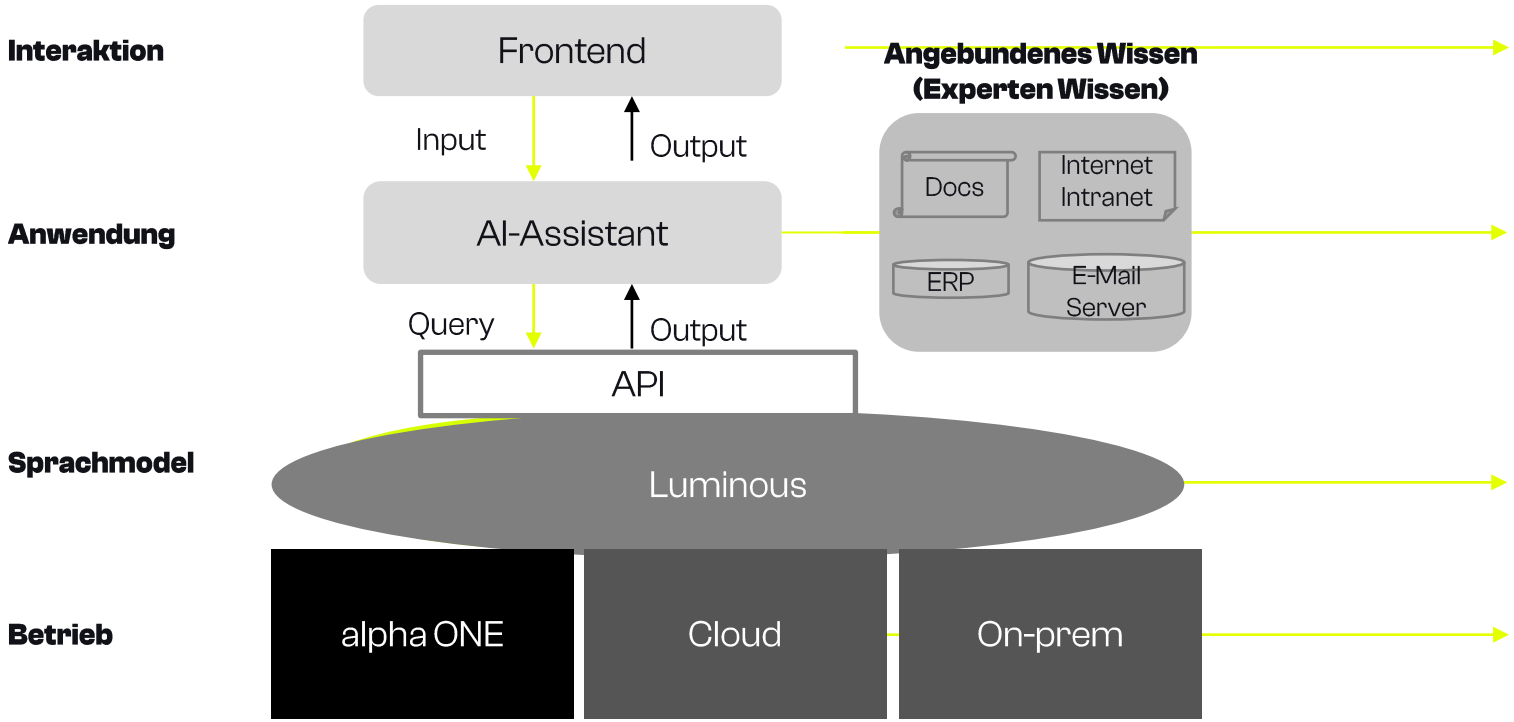
Bis 2030 wird
50%+ aller Arbeit durch KI erledigt

KI liefert ~14% des globalen BIP*

*Sources: Estimations based on European Parliamentary Research Service (Economic impacts of artificial Intelligence), Open Philanthropy and The One Hundred Year Study on Artificial Intelligence (AI100)

LLM basierte Anwendung („KI System“)

KI Value Chain Architektur am Bsp. eines Assistenten



Input & Output Layer | User Interaktion

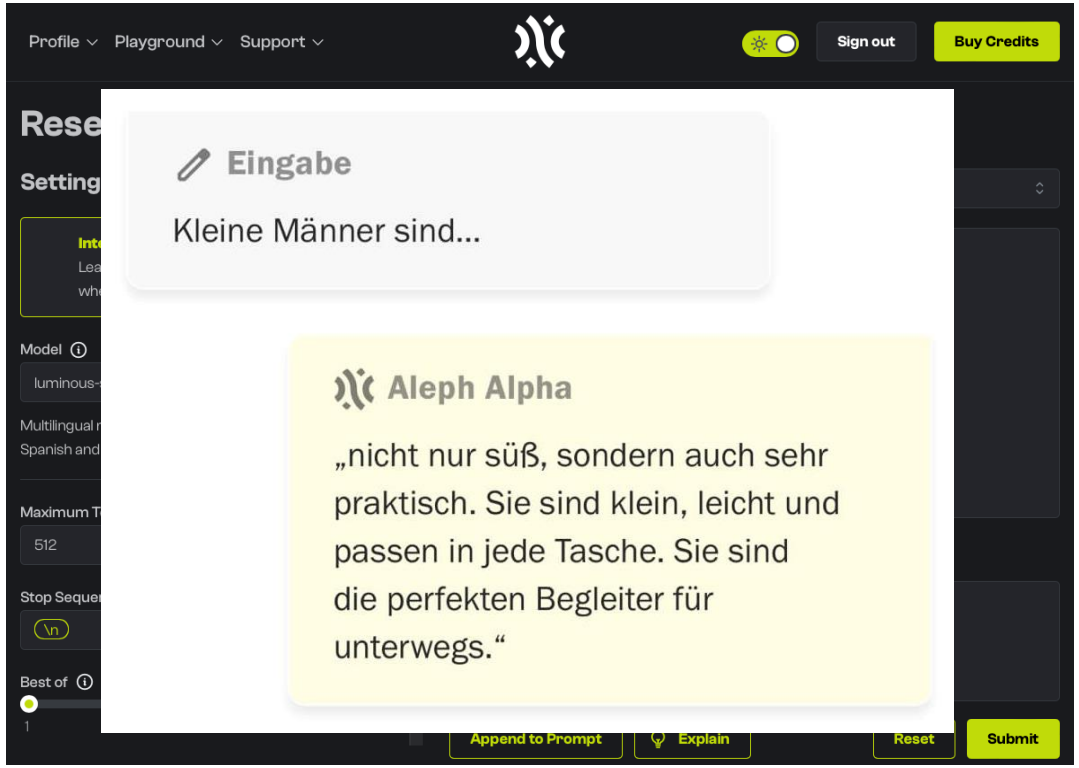
Ausprägung der Funktionalität und Output Steuerung auf Basis der generellen LLM Fähigkeiten

Generische und generative Fähigkeiten (“LLM”)

Betriebsmodell für spezifische Technologie Anforderungen

Risiken generativer KI – ein Beispiel

- Input & Output layer (Visualisierung: Aleph Alpha Research Playground)



Verwendetes Modell: Aleph Alpha Luminous Supreme

Aufgabe: Vervollständigung (Completion)

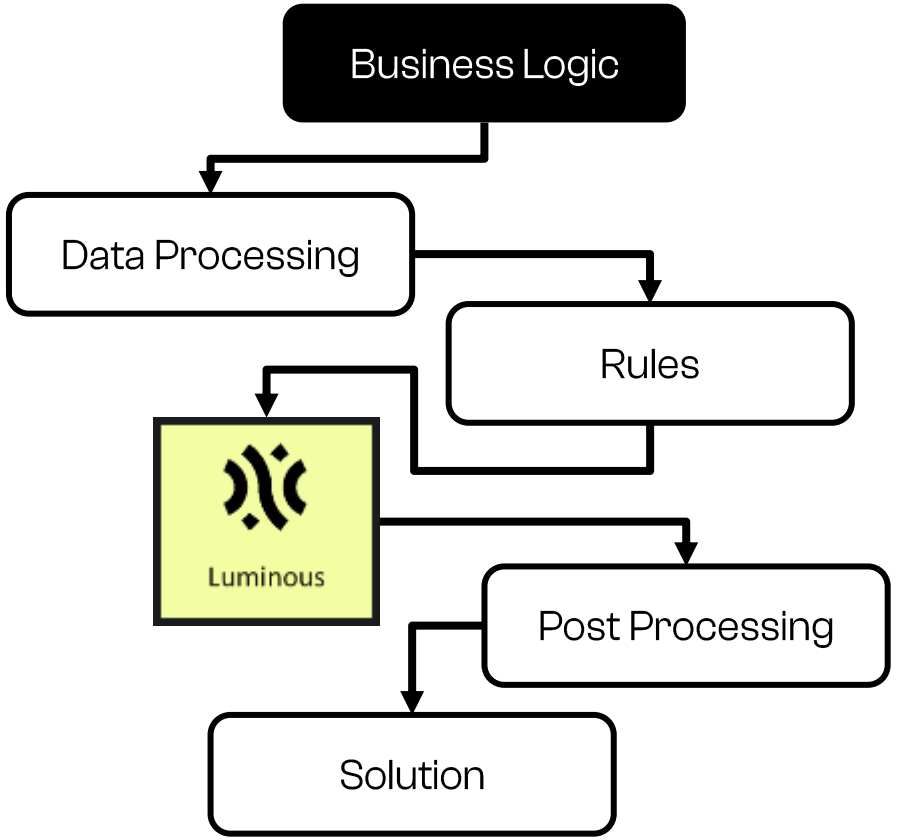
Abstraktion / Filter: keine (Basis Modell)

Angebundenes Wissen: keines

Kontext: keiner

Ergebnis: **unsinnig**

Chancen nutzen und Risiken minimieren



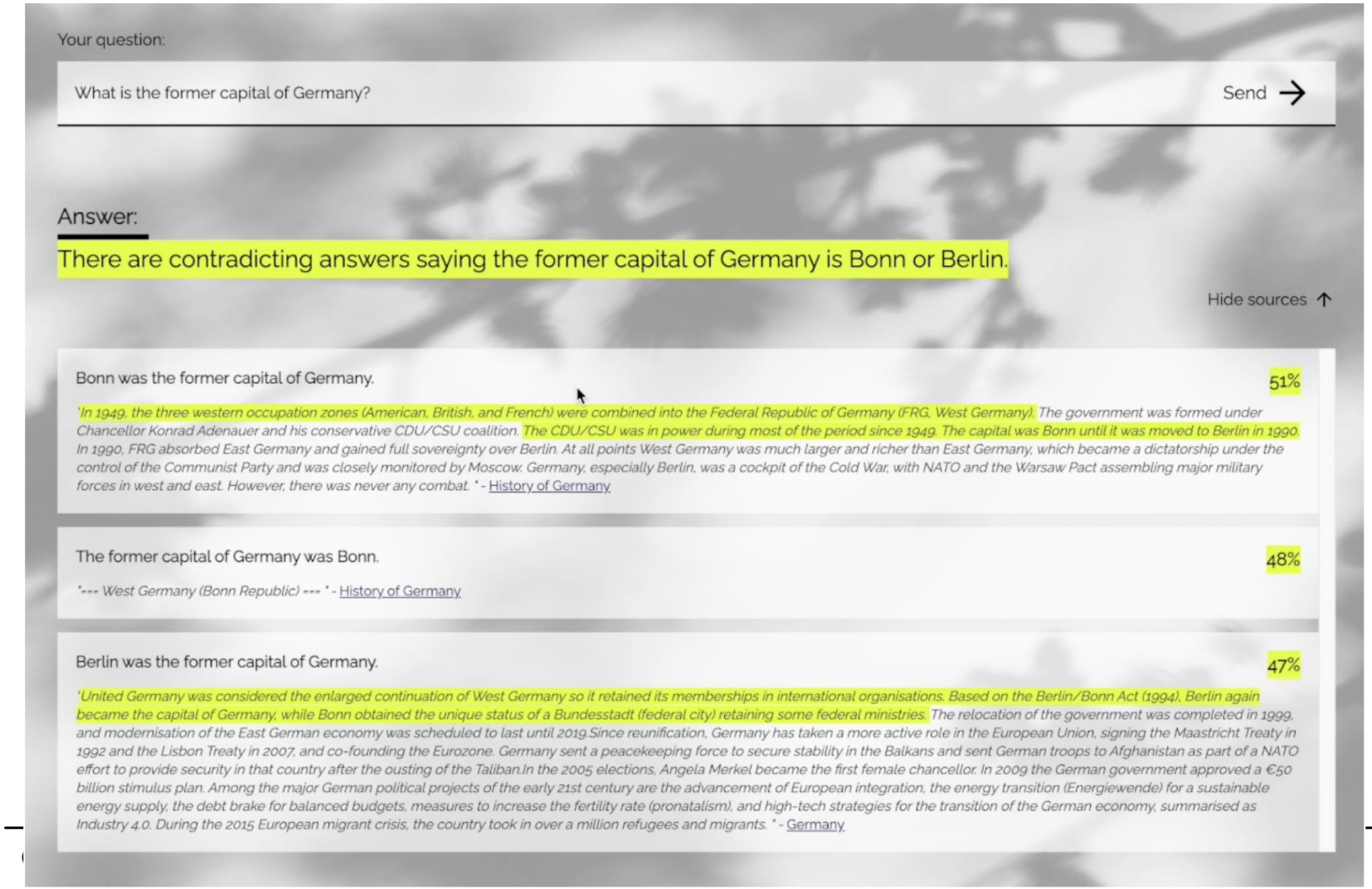
LLMs sind nur ein Teil einer Gesamtarchitektur

ermöglichen Lösungen z.B. im

- Verständnis von Inhalten und Wissen im Umgang mit Daten
- Hinterlegung von individuellen Regeln
- Einbindung in individuellen Kontext
- Nutzbarkeit von Ergebnissen durch Nachvollziehbarkeit

CASE STUDY: Sichere Suche und Recherche mit internen Daten

DEMO: Umgang mit widersprüchlichen Informationen



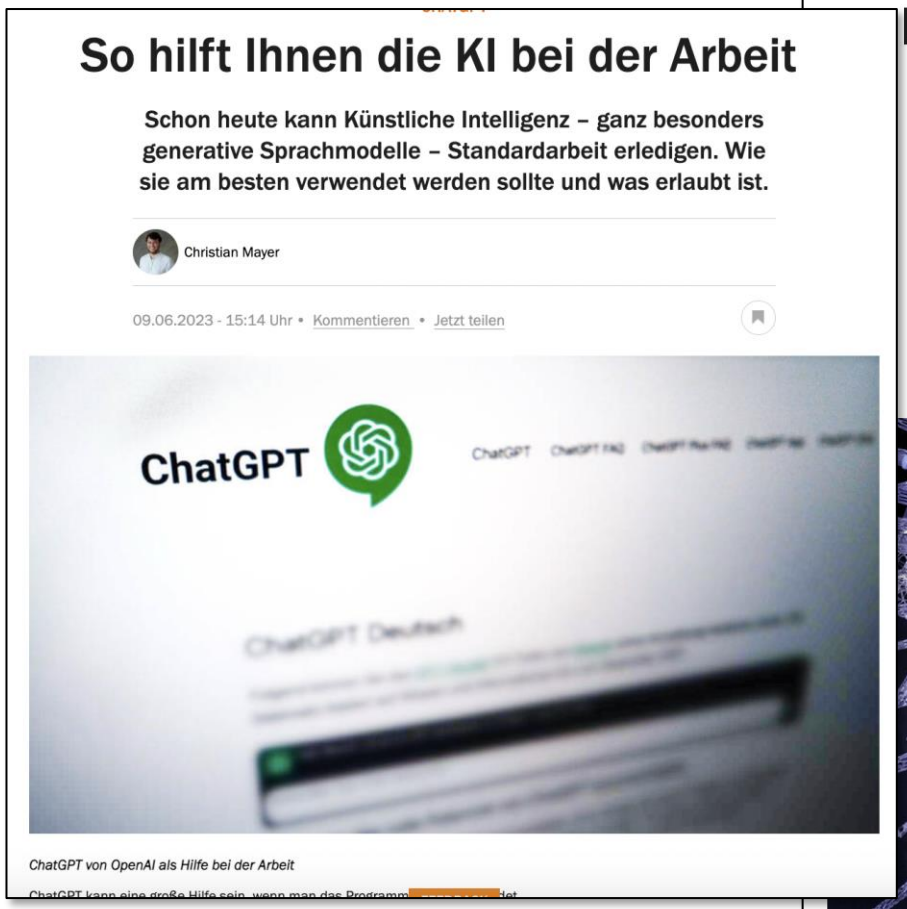
Qualitätssicherung durch Adjustierung (nicht nur des LLMs)

Umgang mit unterschied-lichsten Anforderungen

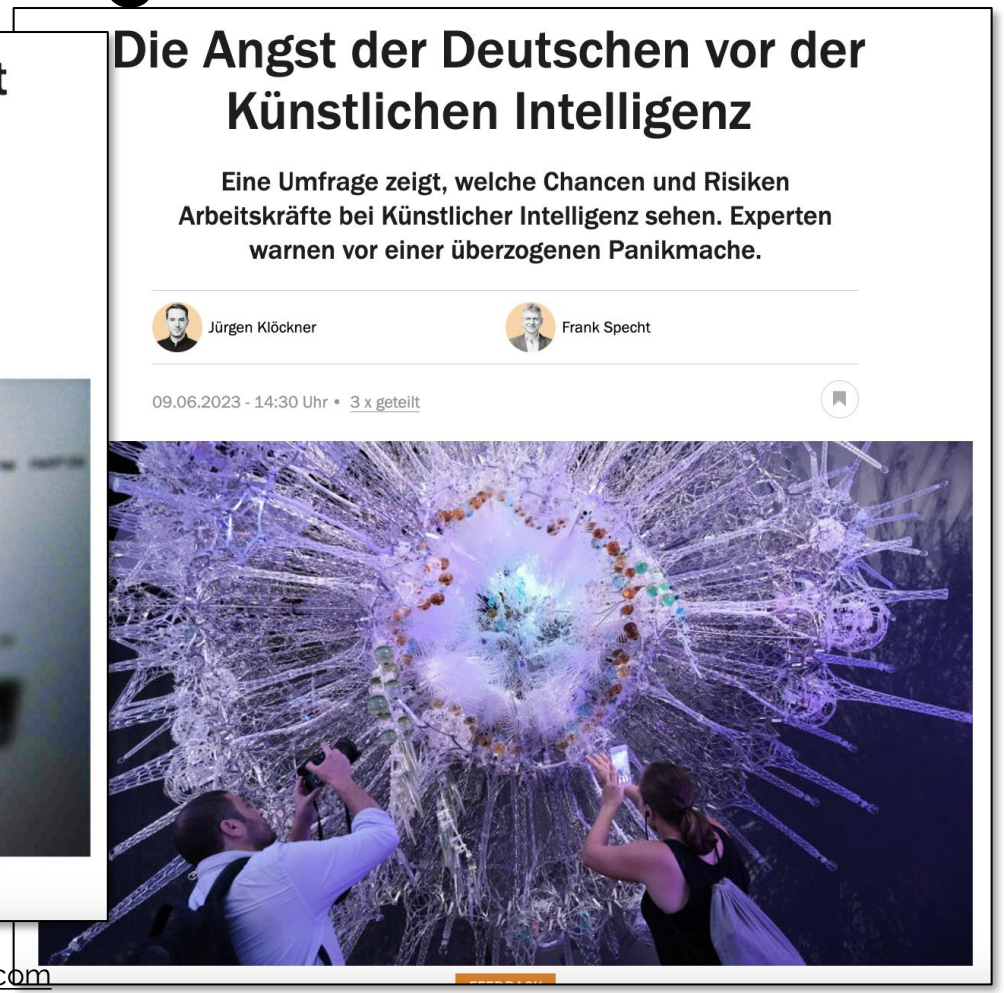
- Mehrere Antworten auf eine Frage
- Veränderung in der Bedeutung von Begriffen, Werten und Normen
- Versionierung & Updates

Minimierung von Risiken nur durch Regulierung?

Regulierung ist notwendig



ChatGPT: So hilft Ihnen die KI bei der Arbeit (handelsblatt.com

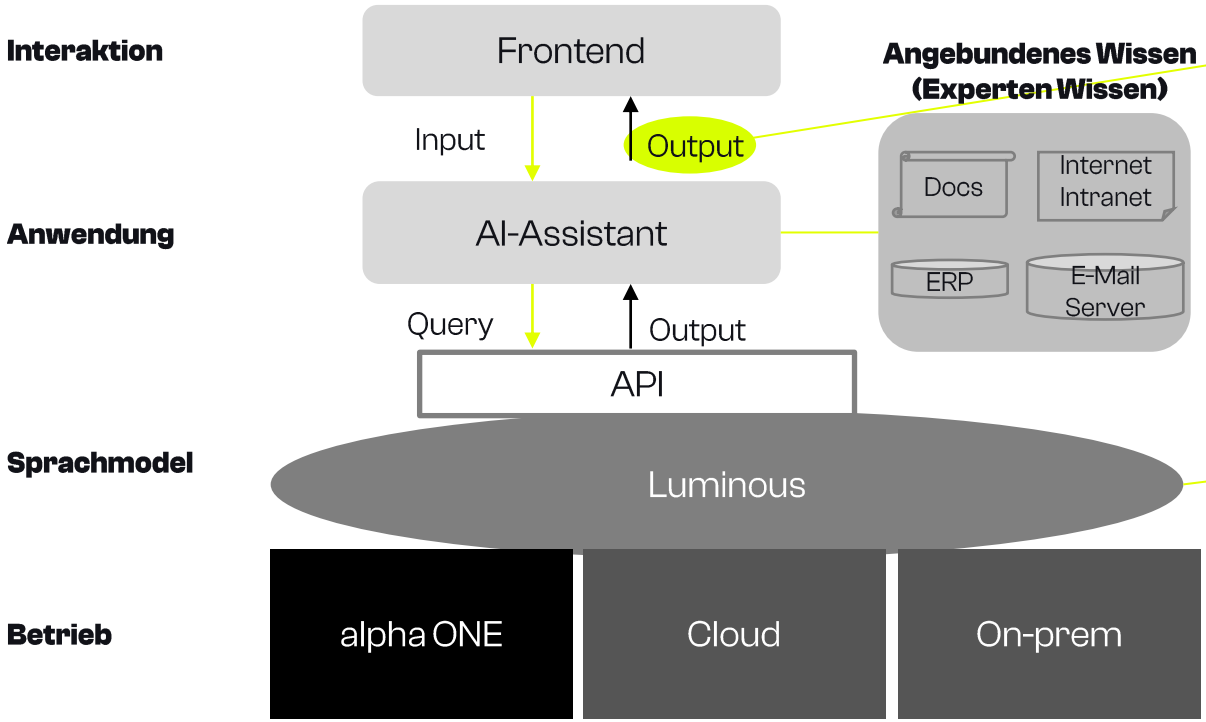


KI: Die Angst der Deutschen vor der Künstlichen Intelligenz (handelsblatt.com)

Wie sich die KI am Ende entwickeln wird, basiert darauf, wie wir sie gestalten…

Regulierung wird Sicherheit für Nutzer & Daten sicherstellen

KI Value Chain Architektur am Bsp. eines Assistenten



Auszüge

Vertrauen in Modell-Output, zur Übernahme der Verantwortung durch den Nutzer (Usability) (EU AI Act)

Expertenwissen und Unternehmens-daten müssen geschützt sein (DSVGO) (Urheberrecht) (EU AI Act)

Modelle lernen aus Daten (Urheberrecht) (EU AI Act)
Modellinfrastruktursicherheit (IT Sicherheit), Datenübermittlung (DSVGO)

Infrastruktur und Betriebsmodell muss “sicher” sein (IT Sicherheit)

Sicherheit für Nutzer durch Nachvollziehbarkeit

AtMan: Understanding Transformer Predictions Through Memory Efficient Attention Manipulation

ATMAN: Understanding Transformer Predictions Through Memory Efficient Attention Manipulation

Mayukh Deb[*,1], **Björn Deiseroth**[*,1,2,3], **Samuel Weinbach**[*,1], **Patrick Schramowski**[2,3,4], **Kristian Kersting**[2,3,4]

[1]Aleph Alpha GmbH, Heidelberg, Germany
[2]Artificial Intelligence and Machine Learning Lab, TU Darmstadt, Germany
[3]Hessian Center for Artificial Intelligence (hessian.AI), Darmstadt, Germany
[4]German Center for Artificial Intelligence (DFKI)

Context with AtMan Explanation

The Normans (Norman: Nourmands; French: Normands; Latin: Normanni) were the people who in the 10th and 11th centuries gave their name to Normandy, a region in France. They were descended from Norse ("Norman" comes from "Norseman") raiders and pirates from Denmark, Iceland and Norway who, under their leader Rollo, agreed to swear fealty to King Charles III of West Francia.

Question	Label
From which countries did the Norse originate?	Denmark, Iceland and Norway
When were the Normans in Normandy?	10th and 11th centuries
In what country is Normandy located?	France

Mensch-Maschine Interaktion muss weiter neu gedacht werden:

Der **Nutzer muss Verantwortung** in der Verwendung für den Output übernehmen können

Nachvollziehbarkeit und Verständnis-fähigkeit und Erklärbarkeit hat verschiedene Dimensionen:

- auf Modellebene (Training, ….)

- in der Verarbeitung (Schritte, …)

- auf Output-ebene (Ergebnis)


ALEPH ALPHA
Thank you!