



**ATLANTIC Hotel Kiel**

08.09.2025

## ***KI mit eigenen Datenquellen – unterstützen RAG-Systeme den Datenschutz?***

- **Dr. Heinke Bastek (Kontor Business IT)**
- **Benjamin Walczak (ULD)**
- **Angelika Martin (ULD)**

*Die große Frage zu RAG:*

# Rockt KI damit den Datenschutz?



# *Playlist*

- **Intro**  
Einführung in KI und RAG
- **Refrain**  
Datenschutzrahmen und Herausforderungen mit LLMs
- **Gitarrensolo**  
RAG-Systeme in der Praxis
- **Bridge**  
Erste Datenschutz-Einordnungen und Handlungsempfehlungen
- **Outro**  
Ausblick
- **Zugabe**  
Offene Diskussion

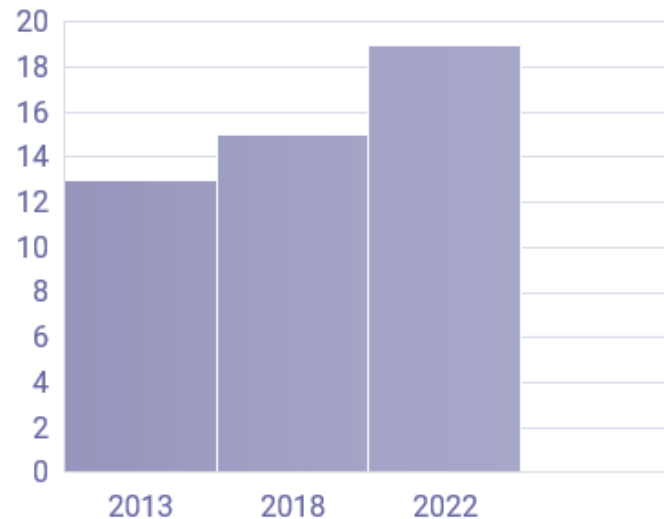


## Intro

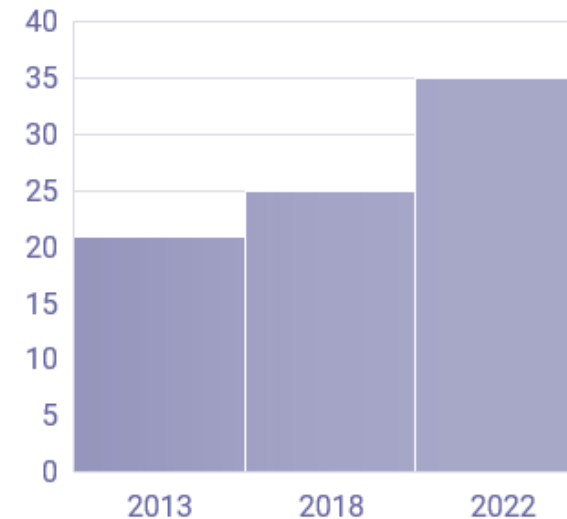
# Ausgangslage

Bürokräfte müssen die Hälfte ihrer Arbeitszeit allein für E-Mailbearbeitung und Informationssuche aufwenden.

**Prozent der Arbeitszeit für Informationssuche**



**Prozent der Arbeitszeit für E-Mailbearbeitung**



**> 50 % der Arbeitszeit werden nicht wirtschaftlich genutzt.**

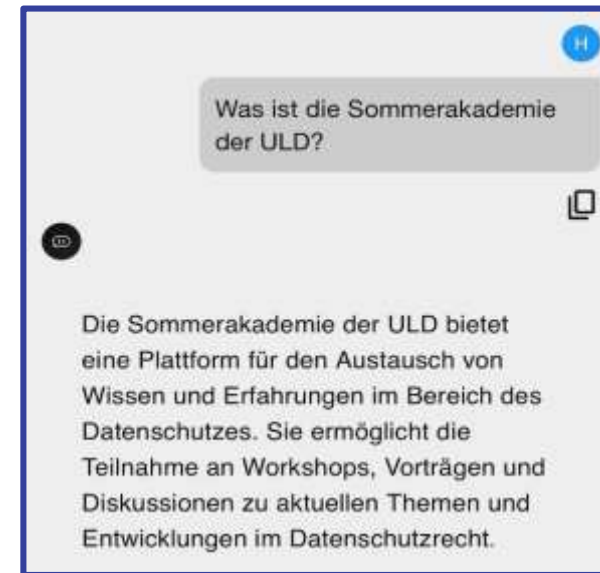
<https://www.buero-kaizen.de/presse/studienergebnisse/studie-2022/>

## Intro

# Wie kann Generative KI unterstützen?

Generative KI ist wie ein sehr schlauer Assistent, der Inhalte erstellen kann.

- LLM (Large Language Model): KI-Modell, das trainiert ist, um mit menschlicher Sprache umzugehen
- Unabhängig von Kontext und Sprachstil
- Einsatz bei
  - E-Mail-Assistenz
  - Content-Erstellung
  - Kundenservice-Chat
  - Code-Assistenz
  - ... vieles, vieles mehr...



→ **Aber: LLMs kennen meine eigenen Daten nicht!**

## Intro

# Wie mache ich internes Wissen verfügbar?

Sprachmodelle mit internen Informationen trainieren erfordert organisatorischen Aufwand für Datenaufbereitung, Governance und Compliance-Management

**Lösung: Informationen in den Prompt verpacken...,  
... insofern der Inhalt in das Kontext Fenster passt.**

| Modell         | Tokens    |
|----------------|-----------|
| GPT-4o         | 128.000   |
| Llama 3.2      | 128.000   |
| Claude Sonnet  | 200.000   |
| Gemini 1.5 Pro | 2.000.000 |

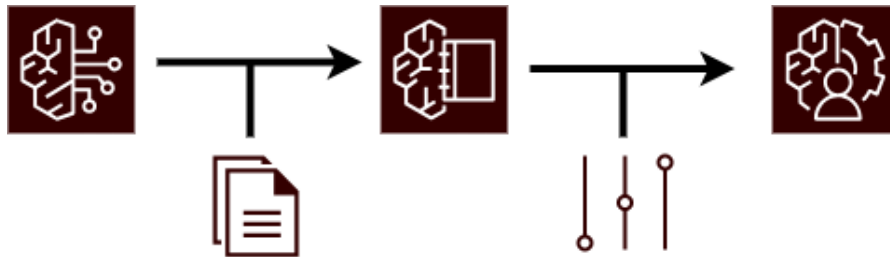
Was ist das Kontext Fenster:

- „Kurzzeitgedächtnis“
- Begrenzt auf bestimmte Länge (Tokens)
- 1 Token  $\approx$  4 Zeichen

**→ der Platz für interne Informationen im Kontext Fenster ist begrenzt!**

## Intro

# Finetuning



- spezialisiert vortrainierte Modelle durch weiteres Training auf spezifische Aufgaben
- bessere Leistungen in dem gewünschten Anwendungsbereich

## Dennoch: Halluzinationen und fehlende Transparenz

Plausibel klingende, aber faktisch falsche Informationen

- Sachliche Fehler
- Fabrizierte Inhalte
- Unsinnige Ausgaben

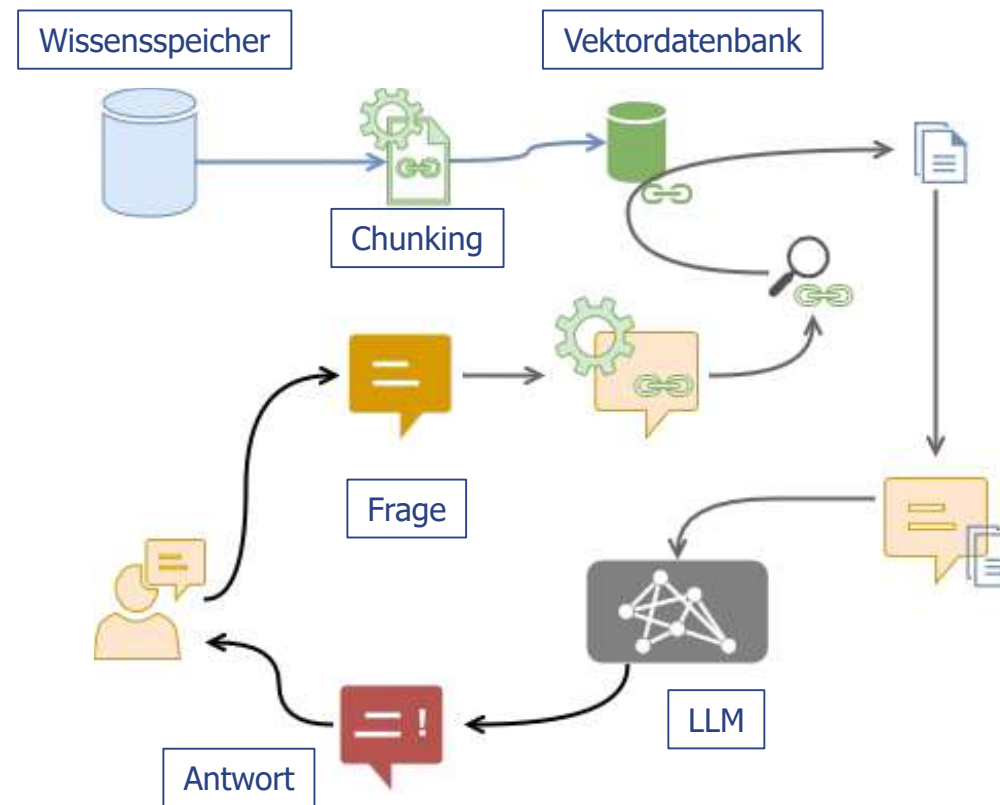
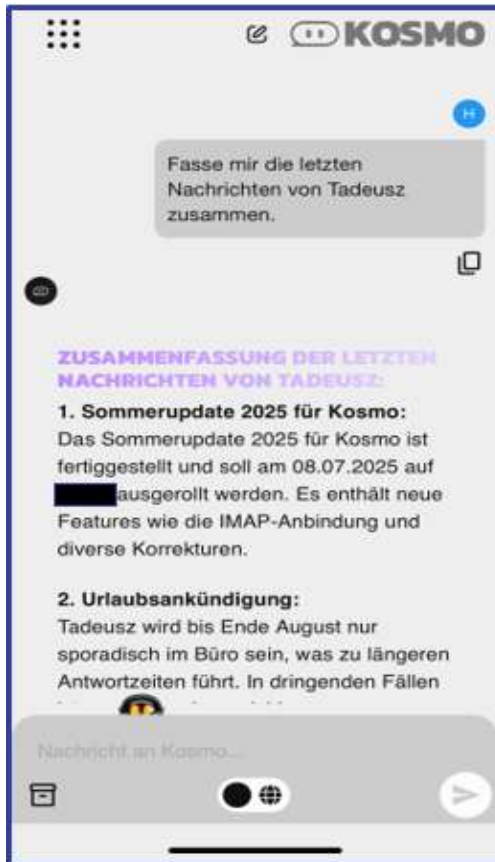


→ **Lösung: Interne Daten nachvollziehbar zur Verfügung stellen**

## Intro

# Retrieval Augmented Generation

RAG ermöglicht den Zugriff auf aktuelle, (nicht-öffentliche) Informationen, wodurch Antworten präziser und faktisch korrekter werden



## Intro

# Finetuning vs. RAG

|                              | Finetuning                                     | RAG                                      |
|------------------------------|------------------------------------------------|------------------------------------------|
| Aktualität                   | Statisch zum Trainingszeitpunkt                | Dynamisch durch externe Datenquellen     |
| Löschkonzept                 | Erst nach neuem Training                       | Sofort umsetzbar                         |
| Nutzer- und Rechtemanagement | Alle Nutzer greifen auf die gelernten Daten zu | Individuell erstellte Wissensdatenbanken |
| Wechsel des Sprachmodells    | Neues Training erforderlich                    | Wechsel ohne Datenverlust möglich        |
| Transparenz eigener Daten    | „Black Box“                                    | Bekannt, was an Daten verarbeitet wird   |
| Nachvollziehbarkeit          | Keine Quellenangaben                           | Direkte Quellenreferenzen                |

Egal ob Finetuning oder RAG:

- On-Premise vs. Cloud: Bleiben Daten im Unternehmen? EU-Datengrenze?
- Datenverarbeitung: Werden In- und Outputs zum Training verwendet?
- Logging: selbst wenn nicht beantwortet, wird die Anfrage geloggt

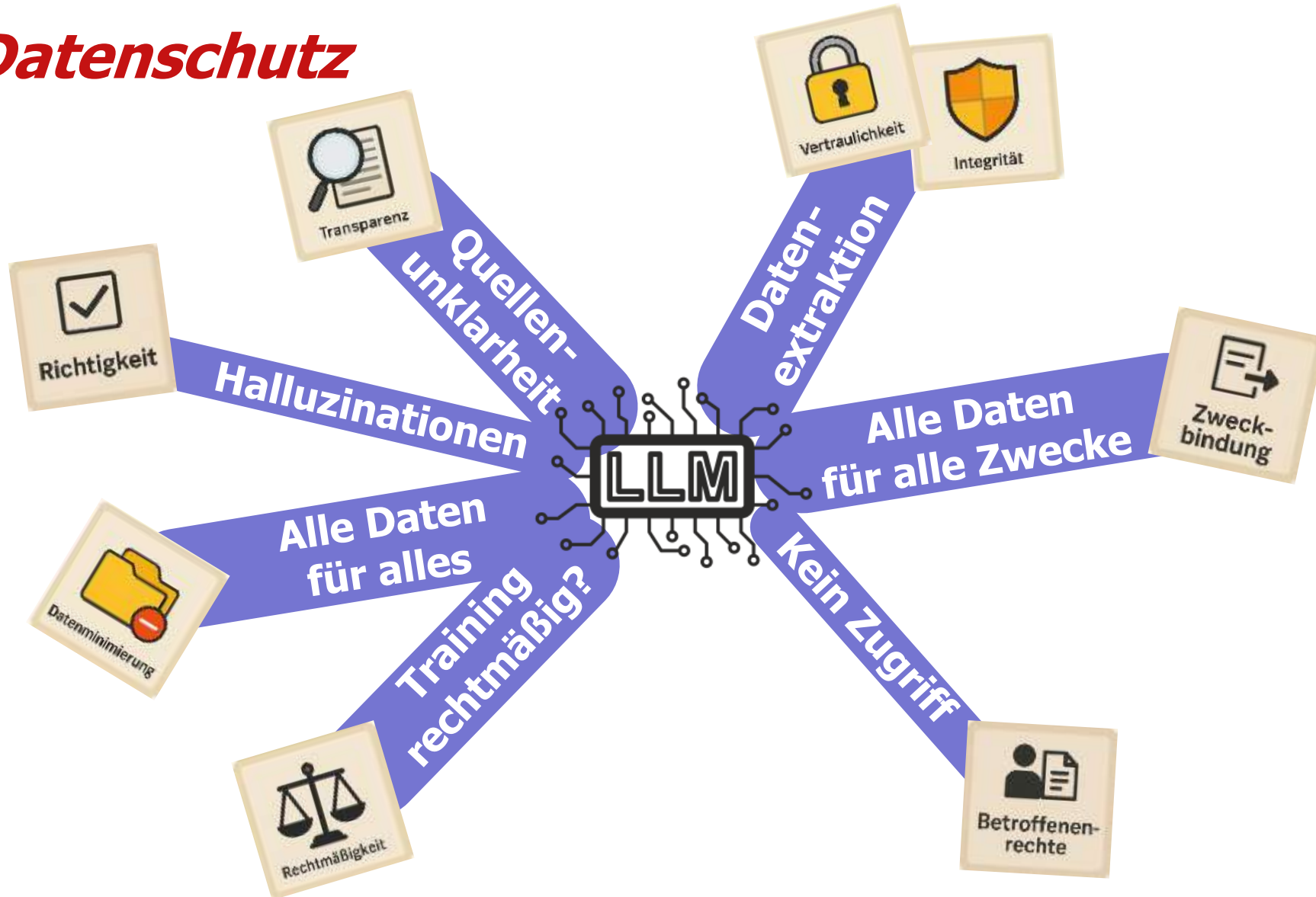
*Refrain*

# ***Datenschutzrahmen***



Refrain

# LLM + Datenschutz



*Gitarrensolo*

# ***Datenquellen im Retrieval***



WEBSEITEN

**Interne Quellen:**

- Unternehmensdokumente, Wikis, E-Mails
- Vollständige Kontrolle über Inhalte
- Keine externen Abhängigkeiten



GESCHÄFTSDATEN

**Externe Quellen:**

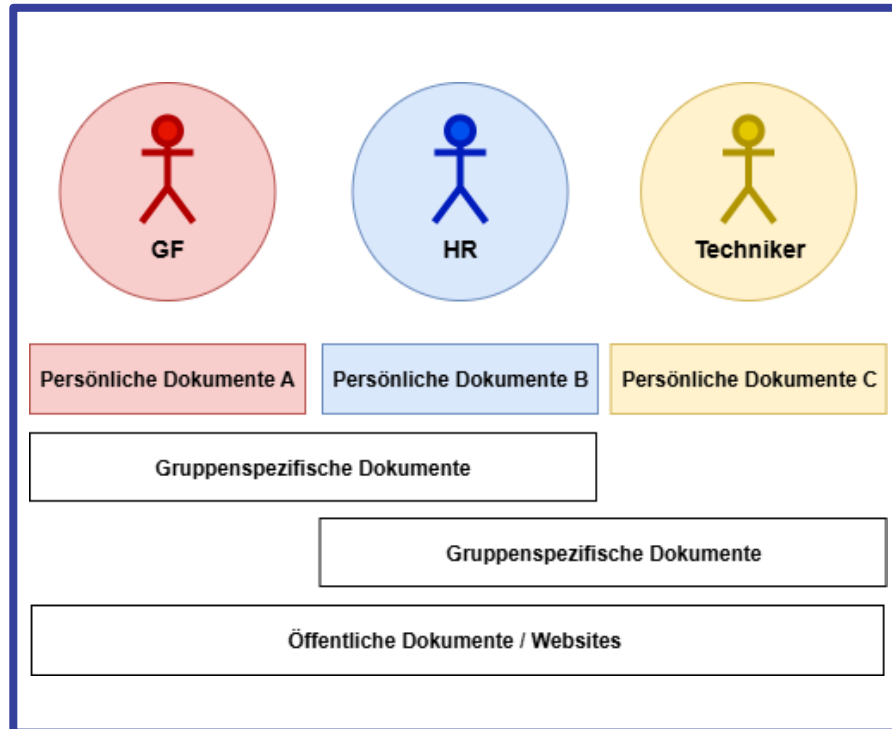
- Öffentliche APIs, Datenbanken, Web-Content
- Aktuelle Informationen, aber weniger Kontrolle
- Datenschutzrechtliche Herausforderungen

INDIVIDUELLES  
ARBEITSMATERIAL**Strukturiert:** Datenbanken, CSV, JSON**Unstrukturiert:** PDFs, Word, E-Mails

**Aber: Das Ergebnis kann nur so gut sein wie die Qualität der Daten!**

# Zugriffskontrollen

## Prinzip der **minimalen Berechtigung**



## Zugriffskontrollen

- Nutzer- und Rechtemanagement in RAG technisch realisierbar
- Individuelle Wissensdatenbanken

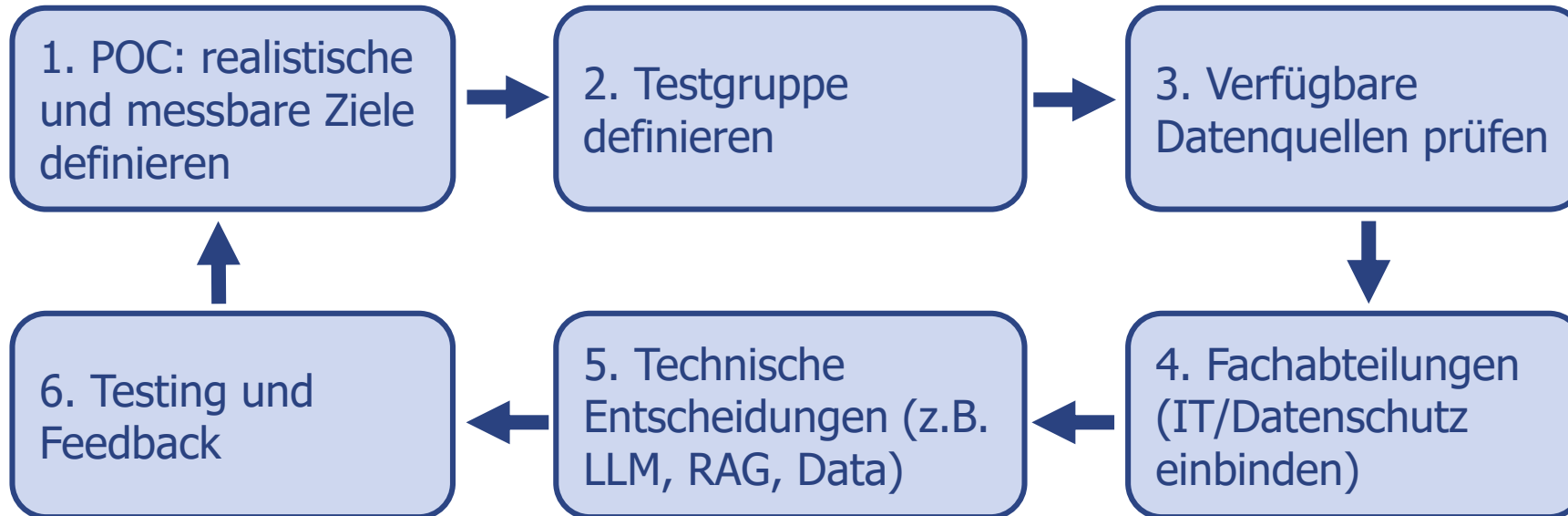
## Berechtigung ist nur so gut wie die gelebte Berechtigungsqualität

- Pflege des Berechtigungskonzepts?
- Geteilte Accounts?
- Sensible Dokumente im geteilten Informationsspeicher?

→ **Regelmäßige Kontrolle der Zugangsberechtigungen erforderlich**

# ***RAG Projekt erfolgreich realisieren***

**MVP-Ansatz:** Testen mit kleinen Datensatz und **messbaren Kriterien**

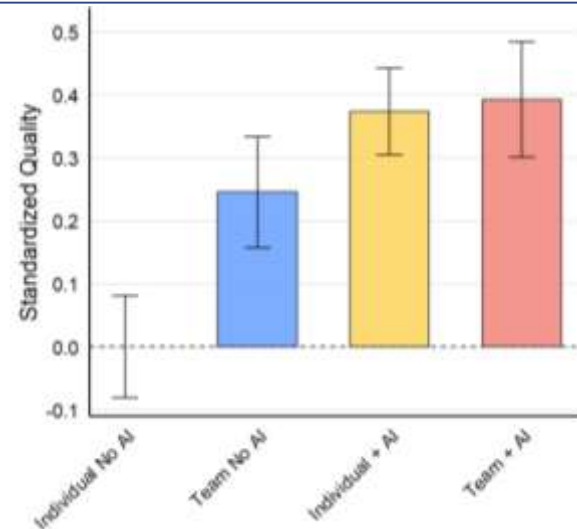


- Mögliche Fehlerquellen: **Datenqualität**, zu ähnliche Dokumente, Chunk-Größe-Dilemma, zu komplexe Fragestellung, implizites Wissen
- **Datensouveräne Lösung** wählen, um Anpassungen zu ermöglichen
- **Schulung** der Mitarbeiter (KI-Verordnung)

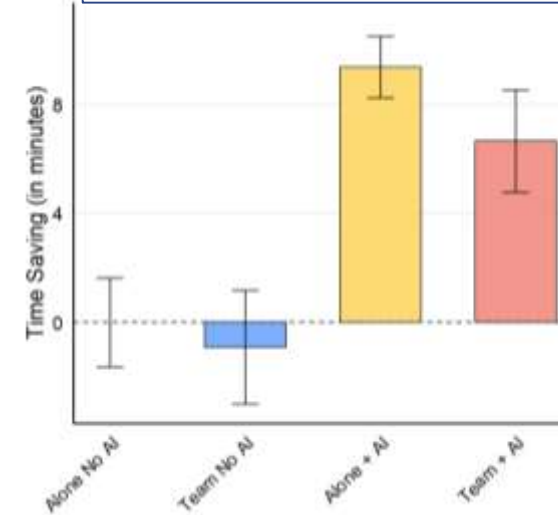
## Gitarrensolo

# Lohnt sich der Aufwand?

Qualität der erarbeiteten Lösung



Realisierte Zeitersparnis



- Jede Frage die durch KI schneller gelöst werden kann, erzielt 10 min Zeitersparnis (ca. 5 – 8 €)
- Ø Fragenanzahl pro Jahr: Low (<100), Medium (100-500) oder High (>500)

→ **Niedrige Nutzungshürde vs. hohe Compliance-Anforderung**

*Bridge*

# ***Erste Gedanken***



- ↑ Kontrolle über ergänzte Daten
- ↑ bestenfalls weniger Halluzinationen
- ? Qualität der ergänzten Daten
- ? Was passiert bei Widersprüchen?
- ggf. Maßnahmen nötig



- ↑ Mehr Transparenz bei ergänzten Daten möglich
- ↑ bestenfalls Quellenangaben in Ausgabe
- ! Erklärbarkeit/Transparenz bleiben unvollständig

Bridge

# Erste Datenschutz-Einordnungen



↑ Mandantentrennung

↑ Rechte-/Rollenkonzept

? Einsatz datenschutzfreundlicher Modelle möglich?

! Gezielte Angriffe auf RAG-Systeme beachten!



↑ Mandantentrennung

! Verwendung noch zweckkonform?

! bedroht durch Verkettung von ergänzten Daten mit denen aus dem LLM

Bridge

## Erste Datenschutz-Einordnungen



! Informationen im LLM verringern sich durch RAG-Methode nicht, aber:

? Einsatz datenschutzfreundlicher Modelle möglich?

↑ Löschen etc. in ergänzten Daten (nur dort!)

! ein rechtswidrig trainiertes Modell bleibt ein solches – auch im RAG-System!

? Einsatz der RAG-Methode kann (!) die Risiken für Betroffene eindämmen

! Einzelfallprüfung nötig!

*Outro*

***Ausblick***

## RAG: Rockt KI damit den Datenschutz?



*Outro****Ausblick***

- Gute Potentiale erkennbar, insbesondere in Bezug auf Richtigkeit, Transparenz und Integrität/Vertraulichkeit
- Anpassungen von KI-Modellen (insb. für RAG-Einsatz trainierte) und KI-Systemen können hilfreich sein
- Wenn die Risiko-Eindämmung festgestellt werden kann, wird datenschutzkonformer KI-Einsatz eher möglich
- Einige Aspekte zum Teil auch auf andere Ansätze übertragbar (z.B. Model Context Protocol)

*Zugabe*

## ***Offene Diskussion***



*Save the date*

**Do., 11. September, 18 Uhr**  
**ULD, Holstenstr. 98, Kiel**

**Frag' für 'ne Freundin**

Austausch rund um KI und Datenschutz

