



# HORN & COMPANY

Datengestützte Entscheidungsfindung mit KI-Sprachmodellen

Konzept und zwei Fallbeispiele

Düsseldorf, im März 2024

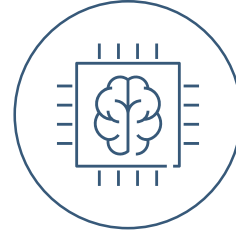
# Generative KI: Ein Booster für die Zugänglichkeit der Datenanalysen für Fachnutzer

## Einleitung



### Benutzer

- Stellt inhaltlich getriebene Anfragen
- Benötigt explorativen Datenzugriff
- Besitzt eher fachliche Expertise



### KI – Sprachmodell

- Generiert passende Datenabfragen
- Erweitert Analysemöglichkeiten
- Übernimmt routinierte Aufgaben



### Aufbereiteter Datenhaushalt

- Harmonisiert mehrere Datenquellen
- Beinhaltet strukt. Informationen
- Passt sich den Anforderungen an



Ermöglichung von komplexen Datenanalysen für Fachexperten ohne umfassende IT-Kenntnisse.



Homogenisierung der Datenbasis: Extraktion und Strukturierung von Informationen aus semi- oder unstrukturierten Daten.



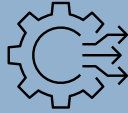
Flexible Anbindung von weiteren internen oder externen Datenquellen sowie BI-Tools.

# Datenaufbereitung: keine reine technische Herausforderung

## Prinzipdarstellung



Problem auf erforderliche Ziel-Analysen herunterbrechen



Passende Zielstruktur für die Datenaufbereitung definieren



Überblick über die relevanten Datenquellen verschaffen

- Welche Anfragen des Benutzers werden erwartet?
- Welche Attribute und Metriken werden für die Analysen benötigt?
- Welche unstrukturierten Daten (Text, Bilder) stehen im Fokus?
- Welche Beziehungen zwischen den Daten sind wichtig?
- In welcher Form soll das Ergebnis aufbereitet werden?

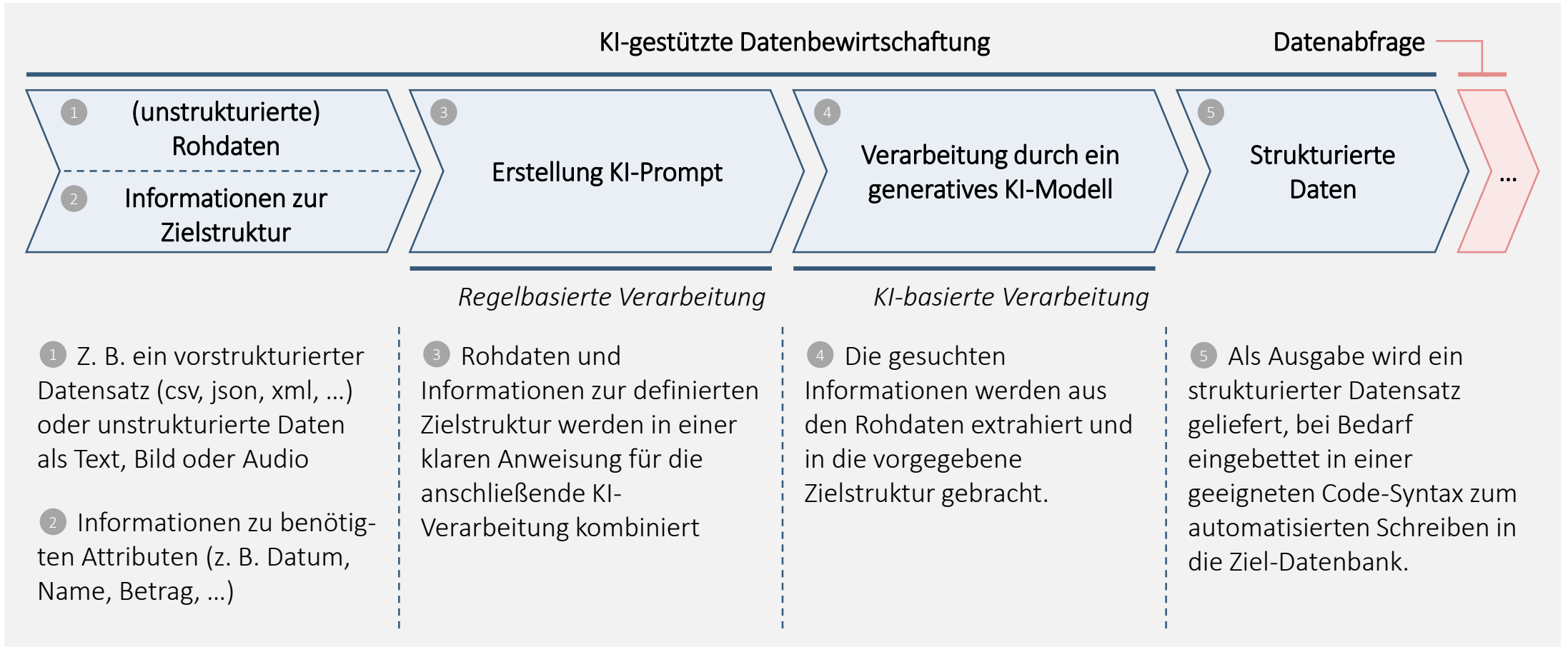
- Welches Datenmodell ist am besten für die gegebene Problemstellung geeignet?
- Welche Aggregationen müssen unterstützt werden?
- Wie häufig müssen die Daten aktualisiert werden?
- Wird eine Historisierung benötigt?

- Welche Datenquellen sind grundsätzlich verfügbar?
- In welchem Format liegen die Daten in den jeweiligen Datenquellen vor?
- Müssen die Informationen in einem zusätzlichen Schritt aus den Rohdaten abgeleitet werden?
- Welche Voraussetzungen für die Nutzung von Daten müssen berücksichtigt werden?

Für eine nachhaltige und ausbaufähige Datenaufbereitung ist eine frühzeitige Auseinandersetzung mit den fachlichen Anforderungen an die späteren Datenanalysen entscheidend.

# Heterogene Quelldaten in eine für Analysezwecke besser zugängliche Form gebracht

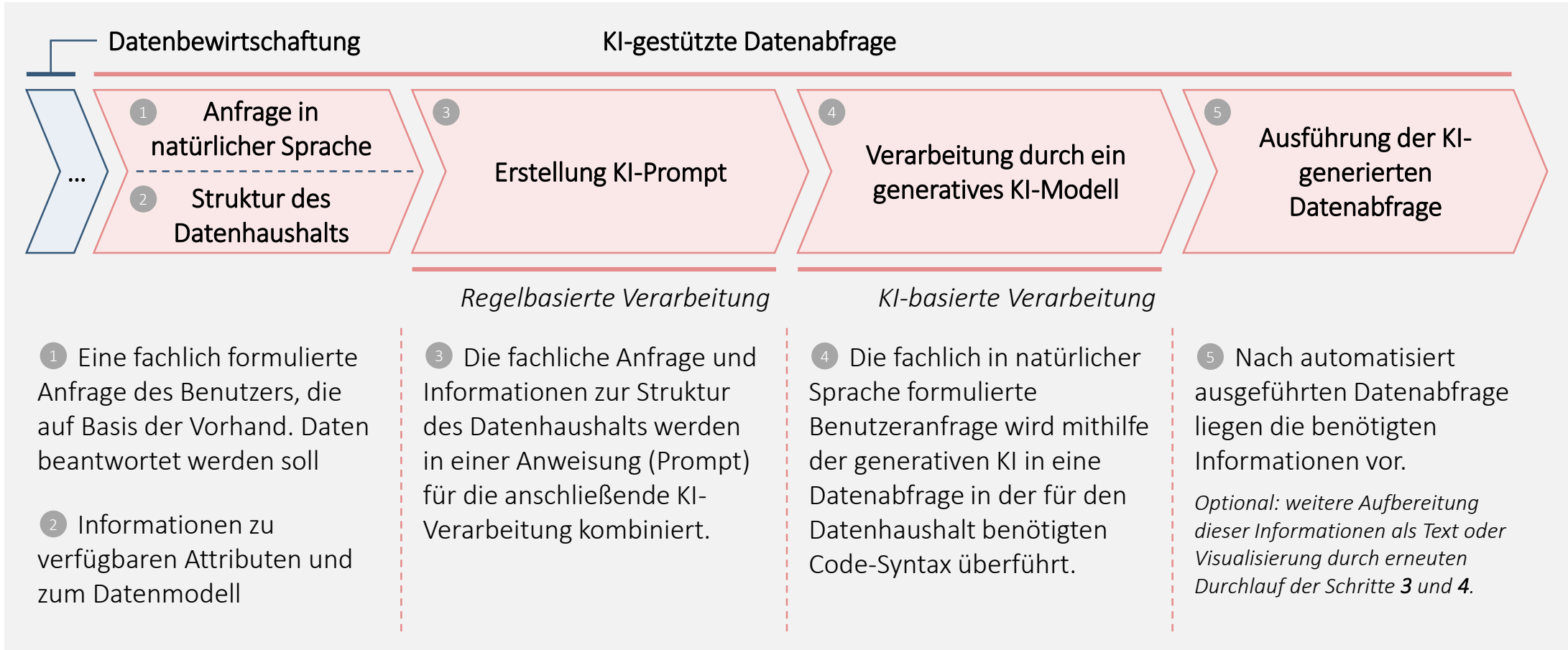
## Prinzipdarstellung



Generative KI hilft dabei, semi- bzw. unstrukturierte Informationen aus unterschiedlichen Quellen an einem Ort und in einer harmonisierten Struktur zusammenzuführen.

# Datenbankabfragen auf Basis von Anweisungen in natürlicher Sprache generiert

## Prinzipdarstellung

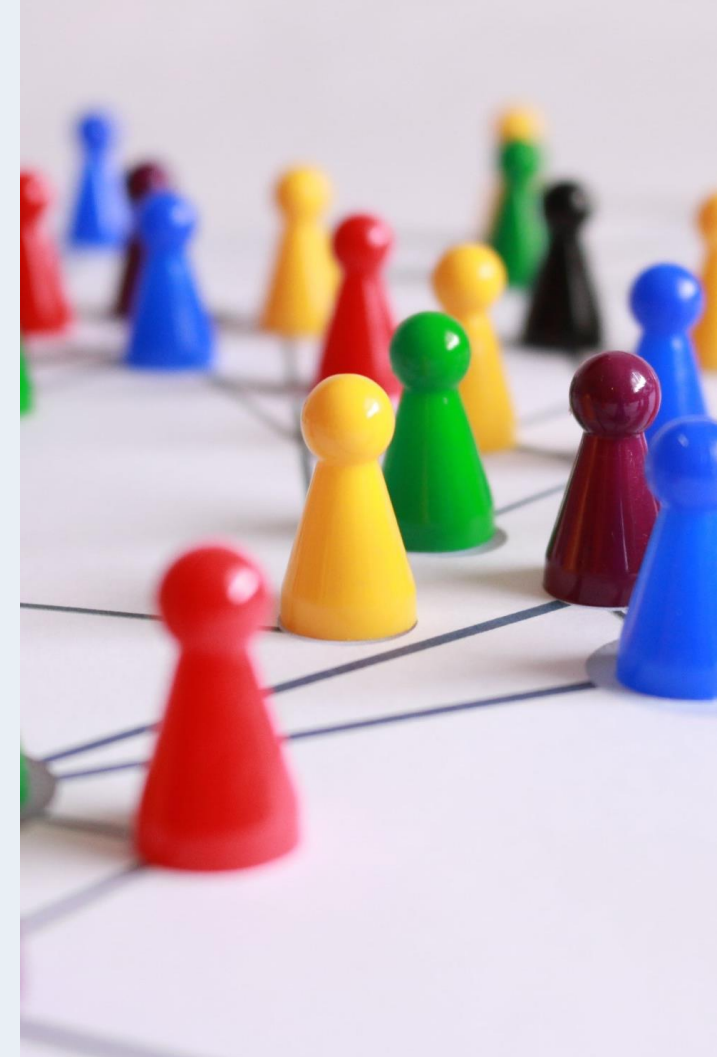


KI-basierte Generierung der Datenabfragen auf Basis der Eingabe in natürlicher Sprache ermöglicht einem deutlich breiteren Anwenderkreis, selbständig Datenanalysen durchzuführen.

# Mit KI werden komplexe Datenanalysen einem breiteren Benutzerkreis zugänglich

## *Demo am Beispiel von Unternehmensnetzwerken*

- In diesem Fallbeispiel nutzen wir frei verfügbare Daten über Unternehmen und ihre Beziehungen zueinander
- Diese Beziehungen spiegeln wesentliche Elemente der Geschäftstätigkeit der Unternehmen wieder und sind im Laufe der Zeit Änderungen unterworfen – die Unternehmen bilden somit Knoten in einem komplexen dynamischen Netzwerk
- Analyse von solchen Netzwerkdaten kann in der Praxis aus diversen Fragestellungen heraus resultieren, darunter ...
- ... laufende Überprüfungen von Kundendaten bei Banken im Rahmen des KYC-Prozesses, die wiederum durch bestimmte Änderungen von Kundendaten ausgelöst werden
- ... Identifizierung der Gruppe verbundener Kunden (GvK), die nach CRR als ein einziges Risiko behandelt werden sollen
- ... Optimierungsprobleme (z. B. Optimierung der Lieferkette), die mit einer umfassenden Analyse der Netzwerkdaten zu den Unternehmen einhergehen
- In unserer Demo haben wir mit den Daten von LEI-Datenbank und OffeneRegister weitestgehend automatisiert einen harmonisierten Datenhaushalt aufgebaut ...
- ... und zeigen, wie komplexe Analysen dieser Daten sich ganz automatisiert mit fachlichen Anfragen in natürlicher Sprache durchführen lassen



# Die angedachten Analysen bestimmen die Wahl des Datenmodells

## Harmonisierung der Daten

### Quelle 1: LEI Register



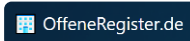
#### Inhalte

- LEI-ID, wenige Stammdaten
- Beziehungsdaten mit Fokus auf Besitzverhältnisse

#### Struktur

- Einheitliche Struktur
- Unterschiedliche Vollständigkeit

### Quelle 2: OffeneRegister



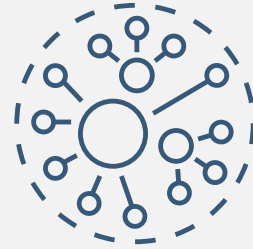
#### Inhalte

- Umfangreiche Stammdaten aus dem Handelsregister

#### Struktur

- Unterschiedlich strukt. Datensätze
- Unterschiedliche Vollständigkeit

### Ziel: Graph-DB auf Basis von Neo4j



#### Inhalte

- Alle Datenattribute aus LEI
- Alle Attribute aus OffeneRegister
- Abgeleitete Beziehungsdaten zwischen den Entitäten aus verschiedenen Quellen

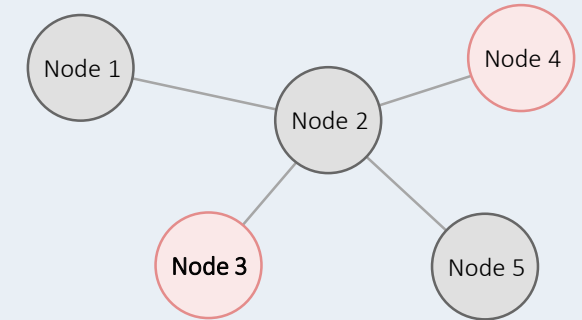
#### Struktur

- Einheitliche Struktur
- Für Netzwerkanalyse gut geeignetes Graph-Datenmodell

### Netzwerkvisualisierung in Neo4j Bloom



```
1 MATCH <...>
2 WHERE <...>
3 RETURN <...>
```



Node Properties	Node 3
<id>	3
Property 1	<Value 1>
Property 2	<Value 2>
Property 3	<Value 3>
...	...

Graphdatenbanken sind besonders geeignet, um komplexe Beziehungen zwischen Datenpunkten zu speichern.

# Explorative Datenanalysen: mit KI auch ohne besondere Datenanalyse-Skills möglich

## Überblick Prozess

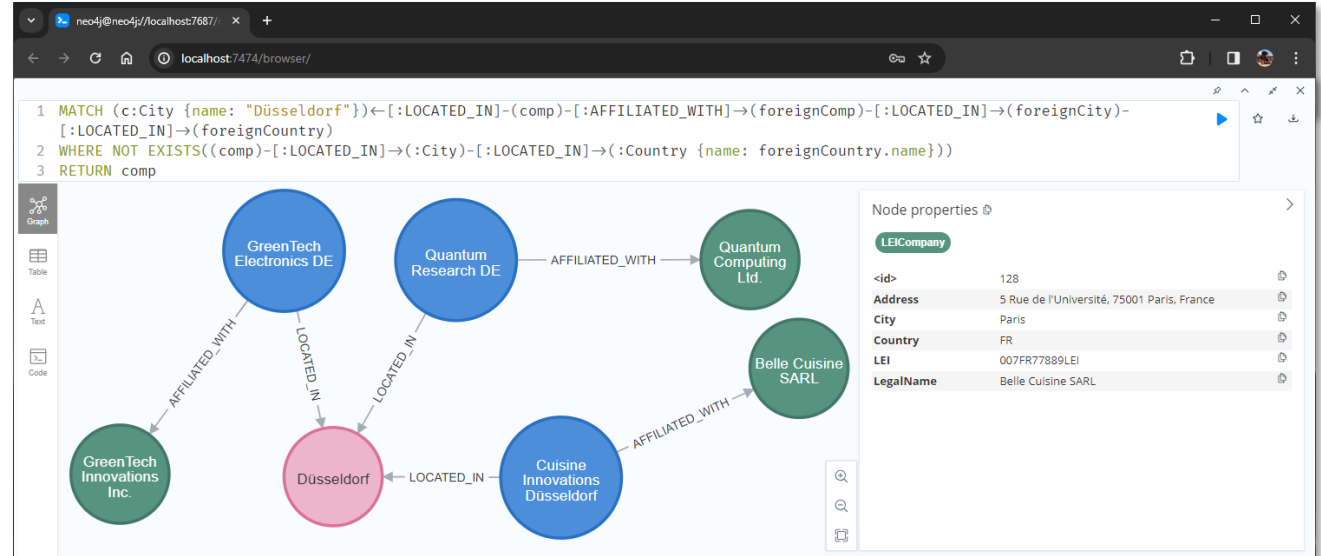
Fachlich formulierte Benutzeranfrage

Ergebnis inkl. Visualisierung



Finde Tech-Firmen im Raum Düsseldorf und visualisiere das globale Unternehmensnetzwerk, in welchem diese eingebettet sind.

Geben Sie hier Ihre Anfrage ein...



### Erstellung LLM-Prompt

```

My Neo4j-DB: The DB has nodes LEICompany, ORCompany, <...>.
LEICompany: Each LEICompany has a Legal Entity Identifier (LEI), <...>
ORCompany: Each ORCompany has a company number (company_number), <...>
    
```

### Generierung Neo4j Query

```

MATCH (c:City {name: "Düsseldorf"})<-[:LOCATED_IN]-(comp)-[:AFFILIATED_WITH]-(foreignComp) <...>
    
```

### Datenabfrage

```

[{'comp': Node('ORCompany', Address='Berliner Allee 42, 40212 Düsseldorf', SubsidiaryName='Belle Cuisine SARL', company_number='DUS103', name='Cuisine Innovations <...>')}]
    
```

Verarbeitung: Regelbasiert (Python)

Verarbeitung: KI (ChatGPT-4)

Verarbeitung: Regelbasiert (Neo4j)

Durch KI-basierte Generierung der Datenabfragen auf Basis der Eingabe in natürlicher Sprache werden datengestützte Entscheidungen einem deutlich breiteren Anwenderkreis ermöglicht.

1. Die Entwicklung des Prototypen erfolgte auf Basis von Originaldaten aus LEI und Offenerregister.de. Für den Showcase in dieser Unterlage wurden die Firmennamen und Orte zwecks Anonymisierung durch synthetische Namen ersetzt.



# KI kann die bisher mühsame Analyse von Textdokumenten revolutionieren

## *Demo am Beispiel von Versicherungspolicen*

- Die umfangreichen Versicherungsscheine beinhalten eine Vielzahl von Informationen, welche aufgrund des hohen Aufwands nicht strukturiert erfasst werden.
- Im vorliegenden Beispiel besteht das primäre Ziel in der effizienten Extraktion und Analyse relevanter Informationen aus einer großen Menge von Dokumenten.
- Insgesamt soll der Evaluationsprozesses beschleunigt werden, um Entscheidungsträgern einen schnelleren Zugriff auf fundierte Informationen zu ermöglichen.
- Die Aufgabe besteht im Wesentlichen aus zwei Teilen: Extraktion von Informationen aus gegebenen Quellen sowie ihre Interpretation inkl. Antwortgenerierung.
- Für die Extraktion von Informationen kommen insbesondere die Stärken von abfragebasierten Modellen in Frage...
- ... wohingegen für die Interpretation sowie eine konkrete Antwortgenerierung sich der Einsatz von KI-Sprachmodellen (Large Language Models: LLMs) anbietet
- Eine Technik, die diese beiden Komponenten kombiniert, nennt sich Retrieval Augmented Generation (RAG) und bildet die Basis für die vorliegende Demo
- Auf diese Weise gewonnenen Informationen sollen in einer strukturierten Form für vielfache Verwendungen in den späteren Analysen abgelegt werden.
- Dauerhafte Speicherung der Daten in einer möglichst universal wiederverwendbarer Form ist ein wichtiger Hebel mit Blick auf die Kosteneffizienz der KI-basierten Ansätze.

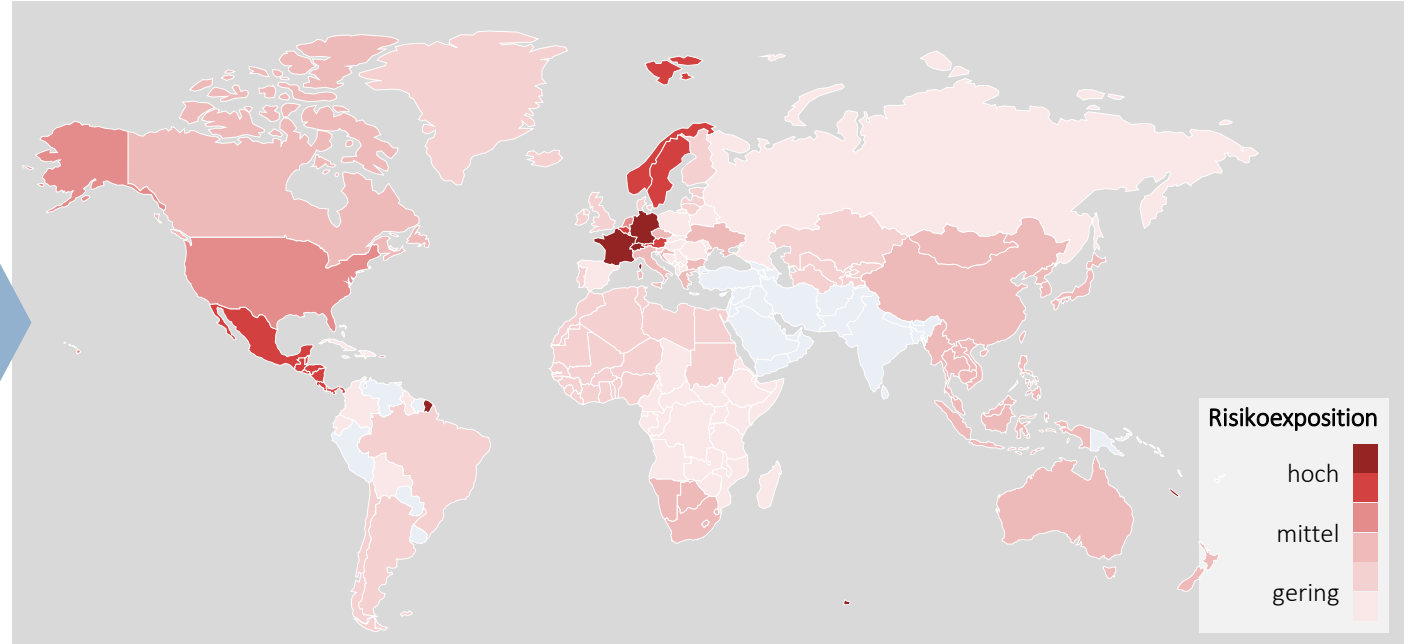


# Geschäftskritische Analysen werden durch Einsatz von KI massiv beschleunigt

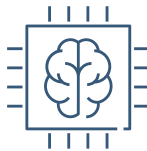
## Prinzipdarstellung

**Aufgabe:** Analysiere die vorliegenden Versicherungspolicen und visualisiere die geografische Verteilung der Risikoexposition auf der Weltkarte.

**Verfahren:** Extraktion von Adressen und Versicherungssummen direkt aus den Versicherungspolicen liefert die für die Visualisierung benötigten nach regionalem Merkmal aggregierbaren Daten.



Versicherungspolicen



LLM



Datenbank









BI-Tool

Die aus Versicherungsscheinen extrahierten Daten werden in strukturierter Form gespeichert – sie stehen anschließend für vielfältige Analysezwecke zur Verfügung.

# Einsatz von LLM in der Praxis: es sind verschiedene Hürden zu überwinden

## Überblick Herausforderungen und Lösungsansätze

Herausforderung	Risiko	Mitigation
 <b>Datenschutz &amp; Sicherheit</b>	Sicherstellung der Einhaltung von verpflichtenden Vorschriften und Bestimmungen (z. B. DSGVO).	<ul style="list-style-type: none"> <li>Nutzung von LLM, die von in der EU betriebenen Cloud-Computing-Plattformen bereitgestellt werden (z.B. Azure)</li> </ul>
 <b>Halluzinationen &amp; Fehlinformationen</b>	Die Sprachmodelle neigen manchmal dazu, falsche oder irreführende Informationen zu generieren.	<ul style="list-style-type: none"> <li>Kombinierung von LLMs mit externen Datenabfragen zur Verbesserung der Antwortgenauigkeit und Kontextrelevanz</li> </ul>
 <b>Kosten &amp; Ressourcen</b>	LLMs erfordern erhebliche Rechenleistung und Ressourcen, was in der Nutzung mit Kosten verbunden ist.	<ul style="list-style-type: none"> <li>Entwicklung intelligenter Algorithmen um die LLMs herum</li> <li>Verwendung kleinerer, spezifischerer Open-Source-Modelle</li> </ul>
 <b>Anpassung an spezif. Domänen</b>	Generische Sprachmodelle müssen für bestimmte Fachgebiete angepasst werden, um optimale Ergebnisse zu erzielen.	<ul style="list-style-type: none"> <li>Re-training bestehender Modelle mit sehr wenigen Trainingsdaten</li> </ul>
 <b>Verständnis von Kontexten</b>	Komplexere Kontexte können Sprachmodelle an ihre Grenzen bringen und falsche Interpretationen hervorrufen.	<ul style="list-style-type: none"> <li>Cleveres Prompt-Engineering, verbessertes Long-Term Memory und Konversationsverfolgung</li> </ul>
 <b>Ethik &amp; Verantwortung</b>	Autonom generierte Inhalte, die unter ethischen Gesichtspunkten potentiell problematisch sein können, müssen kontrolliert werden.	<ul style="list-style-type: none"> <li>Umsetzung der ethischen Aufsicht und der Inhaltsfilter im Rahmen der EU ACT</li> </ul>

Die Ausgestaltung der Lösung im Detail variiert je nach Anwendungsfall – unsere umfangreiche Projekterfahrung hilft uns dabei, den jeweils passenden Ansatz treffsicher zu wählen.

# In Zukunft sind weitestgehend automatisierte Entscheidungsfindungen denkbar

## Ausbaustufen

Zunehmender Grad der Automatisierung

### Stufe III

- Verknüpfung von LLMs mit Code-Datenbanken
- LLM-basierte Unterstützung der Entwicklung logischer Schritte auf Basis der verfügbaren Werkzeuge
- Automatisierte Verwendung der weiterentwickelten Logik mithilfe eines Code-Ausführungstools
- Generierung von Rückfragen zur Beschaffung von weiteren für die Durchführung von Analysen benötigten Informationen
- Ablösung von herkömmlichen BI-Tools

### Stufe II

- Nutzung von LLMs zur Ermittlung von relevanten Fragestellung im Kontext der formulierten Aufgabe
- Unterstützung durch regelbasierte Routinen wie z. B. Speichern und Abrufen von Informationen aus der Datenbank
- Aktives Kostenmanagement durch Abwägen zwischen KI-unterstützten Ad-Hoc Anfragen und Wiederverwertung der persistierten Ergebnisse

### Stufe I

- Erledigung von wenigen, im Vorfeld definierten Aufgaben
- Unterstützung durch vordefinierte regelbasierte Verarbeitung
- Erweiterung der Funktionalität erfordert manuelle Programmierung durch Experten

Der Übergang zu vollautomatisierten Tools signalisiert eine wegweisende Ära in der Datenanalyse, in der KI-gesteuerte Prozesse die Effizienz und Präzision entscheidend verbessern.

# Wir stehen Ihnen jederzeit für Rückfragen zur Verfügung

*Autorenteam und Ansprechpartner*

Jan-Henrik  
Fischer



*Geschäftsführender Partner*

jan-henrik.fischer@horn-company.de

Mobil: +49 162 2726 060

Dr. Sergej  
Hermann

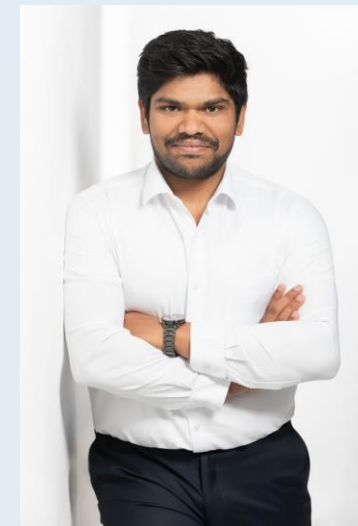


*Principal*

sergej.hermann@horn-company.de

Mobil: +49 162 2726 059

Anoj Winston  
Gladius



*Senior Associate*

anoj-winston.gladius@horn-company.de

Mobil: +49 162 2627 365

# HORN & COMPANY

Internationale Top-Management-Beratung

DÜSSELDORF | BERLIN | FRANKFURT | HAMBURG | KÖLN | MÜNCHEN | NEW YORK | SINGAPUR | WIEN