



Inspired by Data

# Data Technology

Inspired by Data

---



Ein Unternehmen der



Daten → Digitalisierung → Innovation



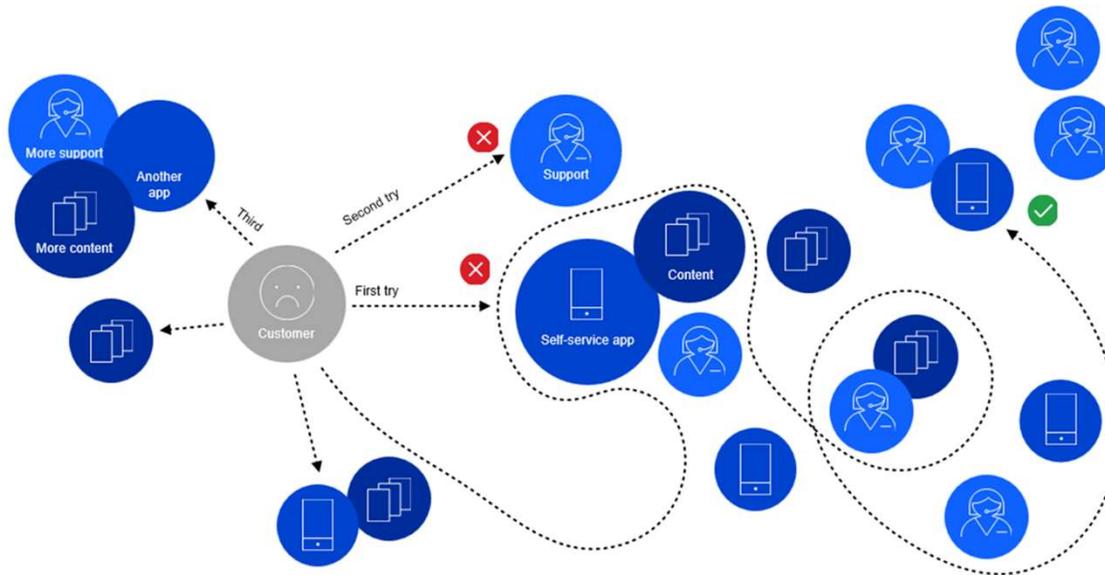
Von der Idee zur Umsetzung eines virtuellen  
Assistenten – eine wertvolle Unterstützung für Ihr Team  
LSZ, 23. April 2024



Ein Unternehmen der



## Motivation – Virtuelle Assistenten Wozu?



- Negative Erlebnisse reduzieren Kundenbindung  
→ Effiziente und rasche Kommunikation macht den Unterschied
- Verringerung personalintensiver repetitiver Berührungspunkte  
→ Effizienzsteigerung durch geleitete Kommunikation
- Personalmangel führt zu weniger kompetentem motiviertem Personal  
→ Assistenzsysteme an Bedeutung gewinnen

Lösung von Service-Fällen erfordert oftmals komplexe Kommunikationsprozesse

Persönlicher Kontakt soll nicht minimiert werden - aber **Menschen** sollen bei zeitintensiven, repetitiven Aufgaben **unterstützt** werden.

## Motivation – Ziel

30%–40%

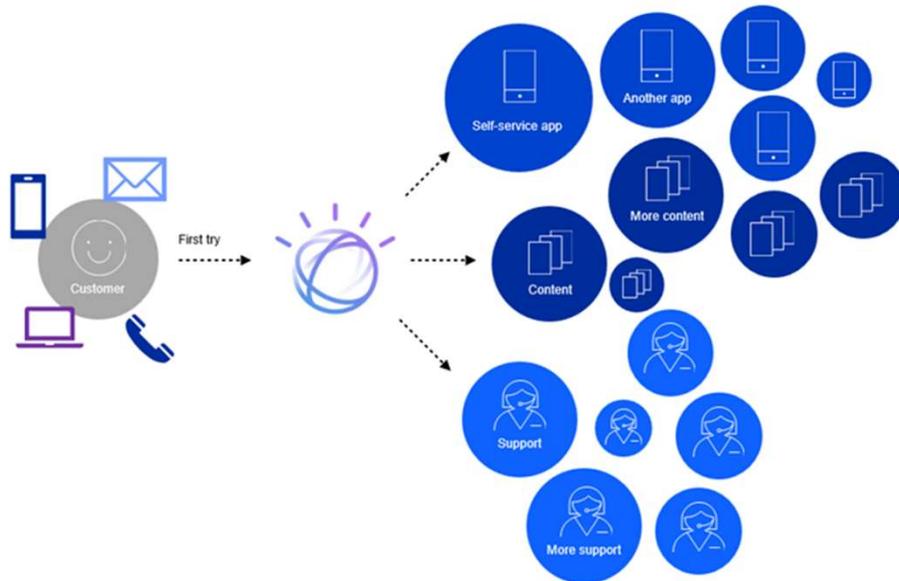
operating cost reduction

15%

increase in NPS

20%

increase in workflow efficiency



Nicht bei komplexen Anwendungsfällen beginnen, sondern einfach starten.

- Fragen / Anliegen können zu 80% bei der ersten Interaktion beantwortet werden
- Assistenten helfen die Frage zu beantworten
  - Bietet der AI-Assistent keine Lösung, wird die Anfrage an den richtigen Mitarbeiter weitergeleitet.
  - Der Mitarbeiter erhält bereits eine qualifizierte Problembeschreibung in Form einer konsolidierten Information aus der Interaktion zwischen dem AI-Assistenten und dem Kunden
- Relativ einfach realisierbare, aber hohe Effizienzsteigerungen  
→ „low hanging fruits“
- Wichtigster Baustein für Umsetzung virtuelle Assistenten sind LLMs

# Was sind Foundation Models bzw. LLMs?



# AI



## Was sind Foundation Models?

- Foundation Models sind große KI-Modelle, die auf umfangreichen Daten trainiert werden
- können für eine Vielzahl von Aufgaben angepasst werden, z.B. Textübersetzung, Bilderkennung...

## Was sind LLMs?

- LLMs sind eine Kategorie von Foundation Models
- Diese sind in der Lage, natürliche Sprache und andere Arten von Inhalten zu "verstehen" bzw. zu generieren

## Können beispielsweise

- Texte zusammenfassen
- Fragen beantworten
- unterschiedliche Sprachen übersetzen

Wo sehen Sie realistische  
Einsatzmöglichkeiten für virtuelle  
Assistenten in Ihrem Unternehmen?

## Use Cases – wobei hilft der Einsatz von FMs oder LLMs?

---

### Funktional

- Zugang zu Support und Services für Kunden und Mitarbeiter im Stil einer Konversation
- Zusammenfassungen / Summaries von Dokumenten und Konversationen
- AI-powered Content Generation (Blog posts, Stellenbeschreibungen, e-mails, Software-Code)
- Unterstützung bei der Automation von einfachen Tasks bis zu komplexen Prozessen und Workflows
- AI-powered Business Intelligence
- Informationssuche und Datenextraktion
- U.v.m.



Ein Unternehmen der



# IBM engagiert sich aktiv darin, KI Lösungen mit Organisationen für ein breites Spektrum umzusetzen

Kundenergebnisse und Analysten Studien

Customer-facing functions and experiences	HR, Finance, and Supply chain functions	IT development and operations	Core business operations
<p><b>Customer service</b> Empower customers to find solutions with easy, compelling experiences 90% customer inquiries handled by an AI assistant<sup>1</sup></p>	<p><b>HR automation</b> Reduce manual work and automate recruiting, sourcing and nurturing job candidates 40% improvement in process productivity<sup>2</sup></p>	<p><b>App modernization, migration</b> Generate code, tune code generation response in real time 60% software development content auto generated by AI<sup>3</sup></p>	<p><b>Threat management</b> Reduce incident response times from hours to minutes or seconds Contain potential threats 8x faster<sup>4</sup></p>
<p><b>Marketing</b> Increase personalization, improve efficiency across the content supply chain Reduce derivative content creation spend<sup>5</sup></p>	<p><b>Supply chain</b> Automate source to pay processes, reduce resource needs and improve cycle times Reduce product origin traceability from days to seconds<sup>6</sup></p>	<p><b>IT automation</b> Reduce automation development, identify deployment issues, avoiding incidents, optimize application demand to supply Reduce mean time to repair (MTTR) by 50%+<sup>7</sup></p>	<p><b>Asset management</b> Optimize critical asset performance and operations while delivering sustainable outcomes Reduce unplanned downtime by 43%<sup>8</sup></p>
<p><b>Content creation</b> Ex. Enhance digital sports viewing with auto-generated spoken AI commentary Scale live viewing experiences cost effectively<sup>9</sup></p>	<p><b>Planning and analysis</b> Make smarter decisions, focus on higher value tasks with automated workflows and AI Process planning data up to 80% faster<sup>10</sup></p>	<p><b>AIOps</b> Assure continuous, cost-effective performance and connectivity across applications Reduce application support tickets by 70%<sup>11</sup></p>	<p><b>Product development</b> Ex. Expedite drug discovery by inferring structure with AI from simple molecular representations Faster and less expensive drug discovery<sup>12</sup></p>
<p><b>Knowledge worker</b> Enable higher value work, improve decision making, and increase productivity Reduce 90% of text reading and analysis work<sup>13</sup></p>	<p><b>Regulatory compliance</b> Support compliance based on requirements / risks, proactively respond to regulatory changes Reduce time spent responding to issues<sup>14</sup></p>	<p><b>Data platform engineering</b> Redesign the approach for data integration using generative AI Reduce data integration time by 30%+<sup>15</sup></p>	<p><b>Environmental intelligence</b> Provide intelligence to proactively plan and manage impact of severe weather and climate Increase manufacturing output by 25%<sup>16</sup></p>

Sources for footnotes: See speaker notes



Was sind Hindernisse auf dem Weg, bzw. Vorbehalte, um AI effektiv für virtuelle Assistenten einzusetzen?

# Was sind typische Hindernisse auf dem Weg, um AI effektiv einzusetzen?

Fundament einer effektiven AI-Wertschöpfung sind die zugrunde liegenden Unternehmensdaten

-  Datenstrategie → fehlende strategische Ausrichtung
-  Data Governance → keine Leitregeln für den Umgang mit Daten
-  Data Management → zu wenig Fokus auf Umsetzung von DG
-  Datenqualität → mangelnde Konsistenz, fehlende Metadaten
-  Datensilos → Daten liegen isoliert
-  Know-how Kosten → Expertise zu erlangen ist ein Kostenfaktor
-  Unzureichende Definition von Use-Cases

Daten → Digitalisierung → Innovation

## Was sind typische Hindernisse, um langfristig von AI zu profitieren?

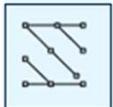
---



- Herausforderungen betreffend der Integration und Skalierung



- AI Governance



- Datenkomplexität



- Hoher Preis



- Limitierte Skills und Erfahrung

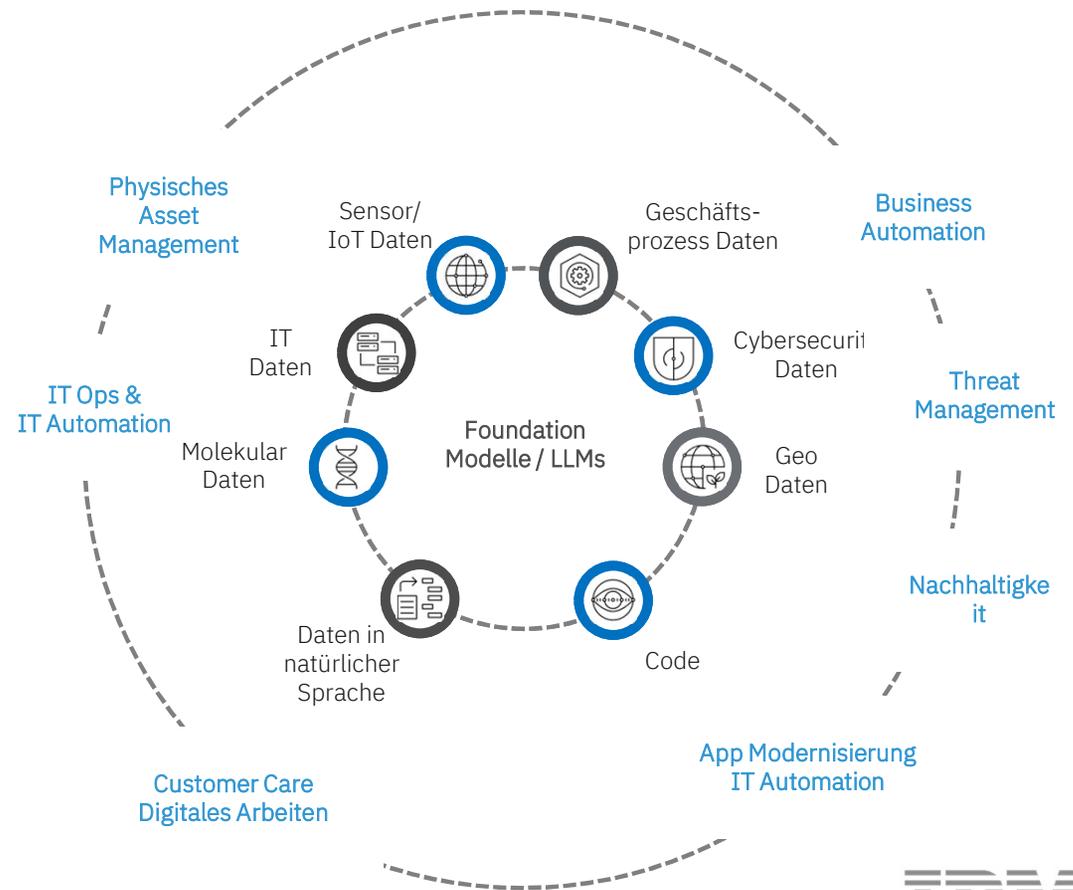


- Fehlende AI Model Development Tools bzw. auch geeigneter Zugang / Skill-Level

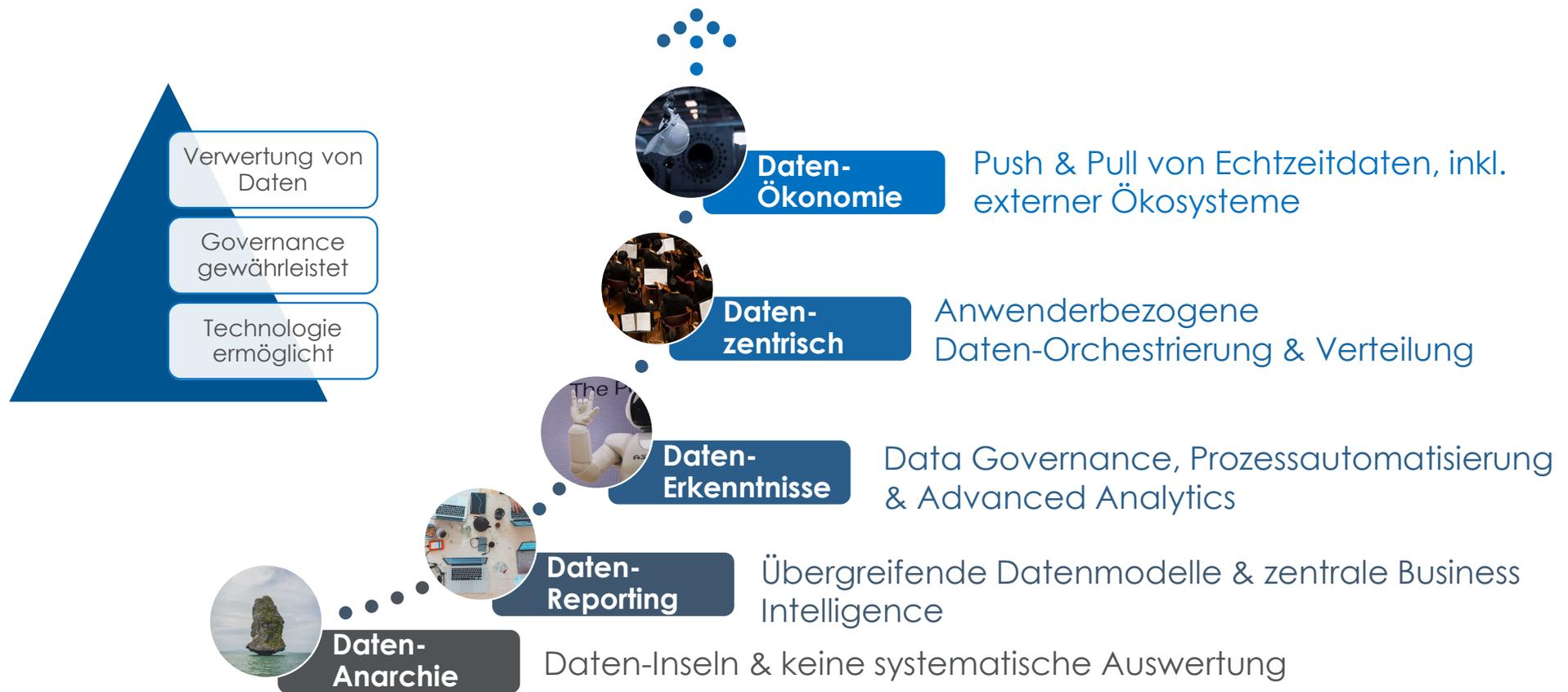
## Wir sind auf einem guten Weg

- ✓ Use-Case genau definiert
- ✓ Herausforderungen wie z.B. Datenintegration, Governance, Datenqualität etc. gemeistert
- ✓ und Foundation Modelle bzw. LLMs werden eingesetzt

→ wir sind auf einem guten Weg



# Reifegradmodell zur Datenwertschöpfung



Wo sehen Sie Ihre Organisation / Abteilung?  
(Reifegrad der Datenwertschöpfung?)

# Schritte zur Datenwertschöpfung



Realitäts-Check  
Wo stehen wir?

Datenzentrische  
Strategie entwickeln

Datenarchitektur  
implementieren

Einsichten aus Daten  
gewinnen

Daten als internes  
Produkt wahrnehmen

# Wie konzipiert & führt man Projekte zur Erstellung von virtuellen Assistenten durch?

## Phase 1 – POC



### Kick-Off Workshop(s)

Zielsetzung & Anforderungen  
Use Case Evaluation



### Proof of Concept

Planung und Umsetzung eines  
konkreten Anwendungsfalles  
"Machbarkeitsstudie"



### Weitere Projekte

Learnings aus dem POC für  
andere erfolgsversprechende  
Anwendungsfälle

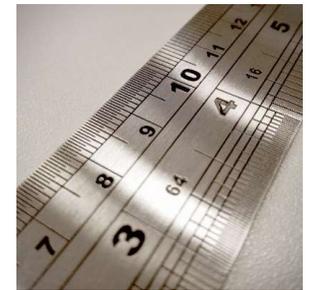
## Phase 2 – MVP



### Minimum Viable Product

Know-how Transfer in den  
Fachbereich und  
Deployment von Ergebnissen

## Phase 3 – Produktion



### Produktivsetzung

Automatisierung von Analyse-  
Prozessen und Deployment

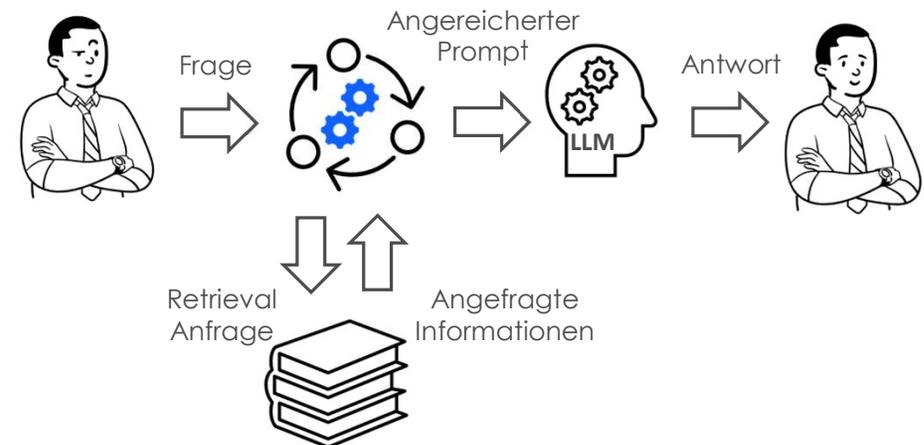
Haben Sie die Erfahrung gemacht, dass LLMs nicht die richtigen / erwünschten Ergebnisse geliefert haben?

# Bitte beachten Sie die Packungsbeilage oder fragen Sie Ihren Consultant

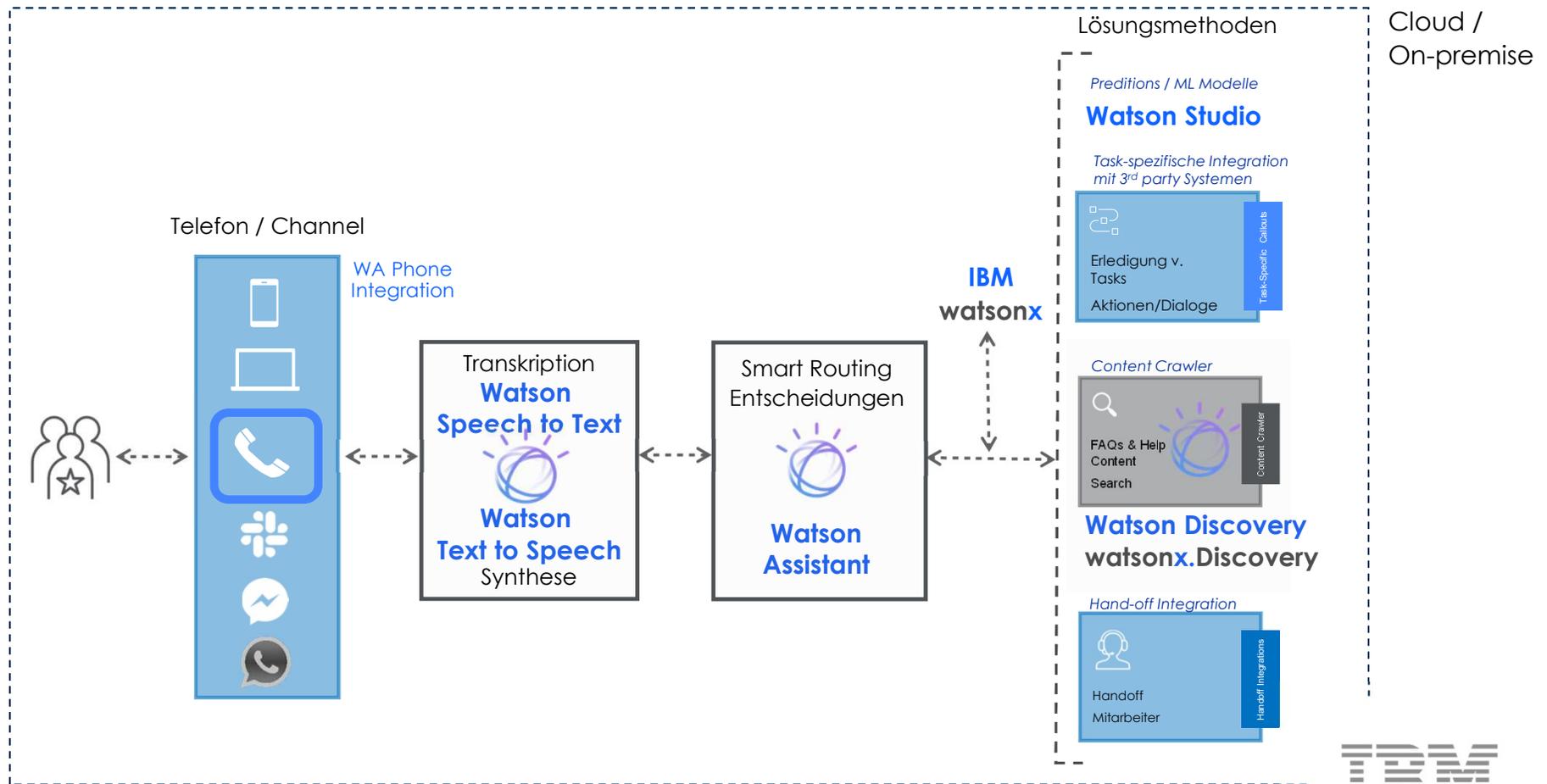
## Disclaimer

- Man darf LLMs nicht unkritisch einsetzen  
→ Warnung vor „Halluzination“
- LLMs können von sich aus nur Informationen aus den Trainingsdaten verlässlich wiedergeben, ansonsten geben sie nur nächstwahrscheinliche Antwort wieder → ungewünschte Resultate
- Zusätzliche Informationen müssen bereitgestellt werden  
→ Fokus auf relevanten Kontext

**RAG (Retrieval Augmented Generation)** ist eine effektive Möglichkeit, um Halluzinationen zu unterbinden.



# Technologieauswahl – Beispiele für „Customer Care“ (IBM Portfolio)



Cloud /  
On-premise

# watsonx

---

Für Unternehmen entwickelte IBM KI- und Datenplattform



## watsonx.ai

Trainieren, prüfen, optimieren und implementieren Sie mühelos Basismodelle und Modelle für maschinelles Lernen. (Unterstützt IBM Models, Open Source Models, 3rd Party bzw. Custom Trained Models).

Inkludiert:

- Foundation Model Library
- Prompt Lab
- Tuning Studio
- Data Science und MLOps



## watsonx.data

Skalieren Sie Analysen und KI-Workloads für all Ihre Daten – und zwar überall mit einem offenen, hybriden und regulierten Datenspeicher.



## watsonx.governance

Verantwortlichkeiten, Transparenz und Erklärbarkeit in Ihren Daten- und KI-Workflows schneller umsetzen mit watsonx.governance.

<https://www.ibm.com/watsonx>



Ein Unternehmen der

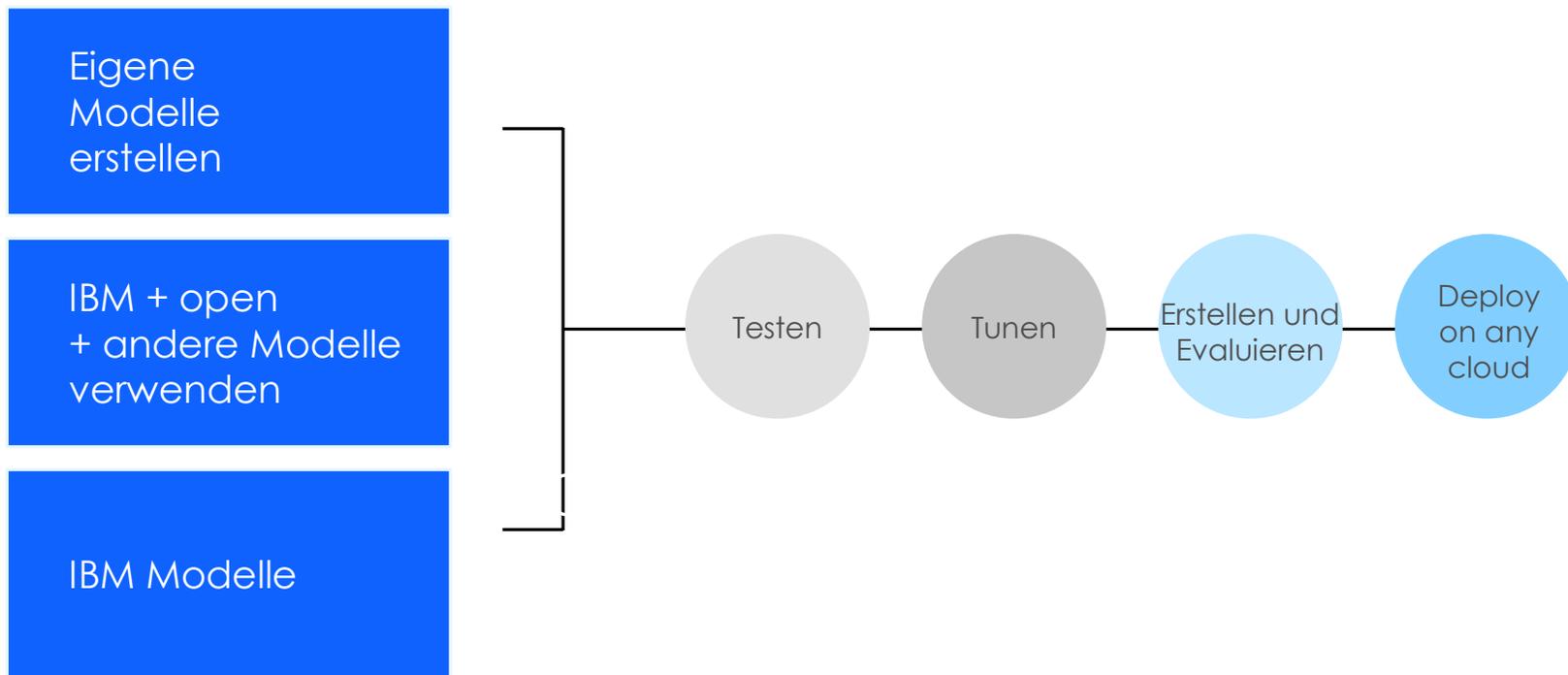
 **SAN**DATA | Die IT-Gruppe

**IBM**  
Silver Partner 20

## watsonx.ai – Unterstützte Modelle

### watsonx.ai

IBM's AI Approach ist offen: multi-model & multi-cloud fähig



# Warum ist die Modell- und Plattform-Auswahl wichtig?

- Unterschiede bei FM Modellen
- Ein Modell wird nicht allen Use Cases gerecht
- “Größer ist nicht immer besser”

Spezialisierte Modelle können general-purpose Modelle “outperformen”  
mit geringeren Infrastruktur Requirements

- Kostenunterschiede
- Möglichkeit, Lösungen auf unterschiedlichen Clouds zu deployen, bewahrt Flexibilität
- Governance wird zusehends wichtiger (Transparenz, Ethik, Vertrauenswürdigkeit usw.)

**Auszug** <https://www.ibm.com/products/watsonx-ai/foundation-models>

**granite-20b-multilingual** (IBM) - Supports Q&A, summarization, classification, generation, extraction, translation and RAG tasks in several languages

**llama-3-8b-instruct** (Meta) - Supports summarization, classification, generation, extraction and translation tasks.

**mixtral-8x7b-instruct** (MistralAI) - Supports Q&A, summarization, classification, generation, extraction, RAG and code generation tasks.

## Fragen und Antworten



Ein Unternehmen der



# Ansprechpartner

---



## Justin Bitterlich

Business Development Manager  
SanData Technology  
Tel.: +49 (0) 151/41485106  
j.bitterlich@sandata.de



## Martin Diermaier

Senior Consultant  
Data Technology  
Tel.: +43 (0) 664 852 04 79  
martin.diermaier@datatechnology.at



## Sabine Riegler

Data & AI Tech. Specialist  
IBM Österreich  
Tel: +43 (0) 664 1212048  
sabine.riegler@at.ibm.com



## Sandra Kral

Software Partner Specialist  
IBM Österreich  
Tel.: +43 (0) 664 / 6185407  
sandra\_kral@at.ibm.com



## Michaela Göschl

Business Development Manager  
TD Synnex  
Tel. : +43 (0) 676/847774431  
michaela.goeschl@tdsynnex.com



Ein Unternehmen der

