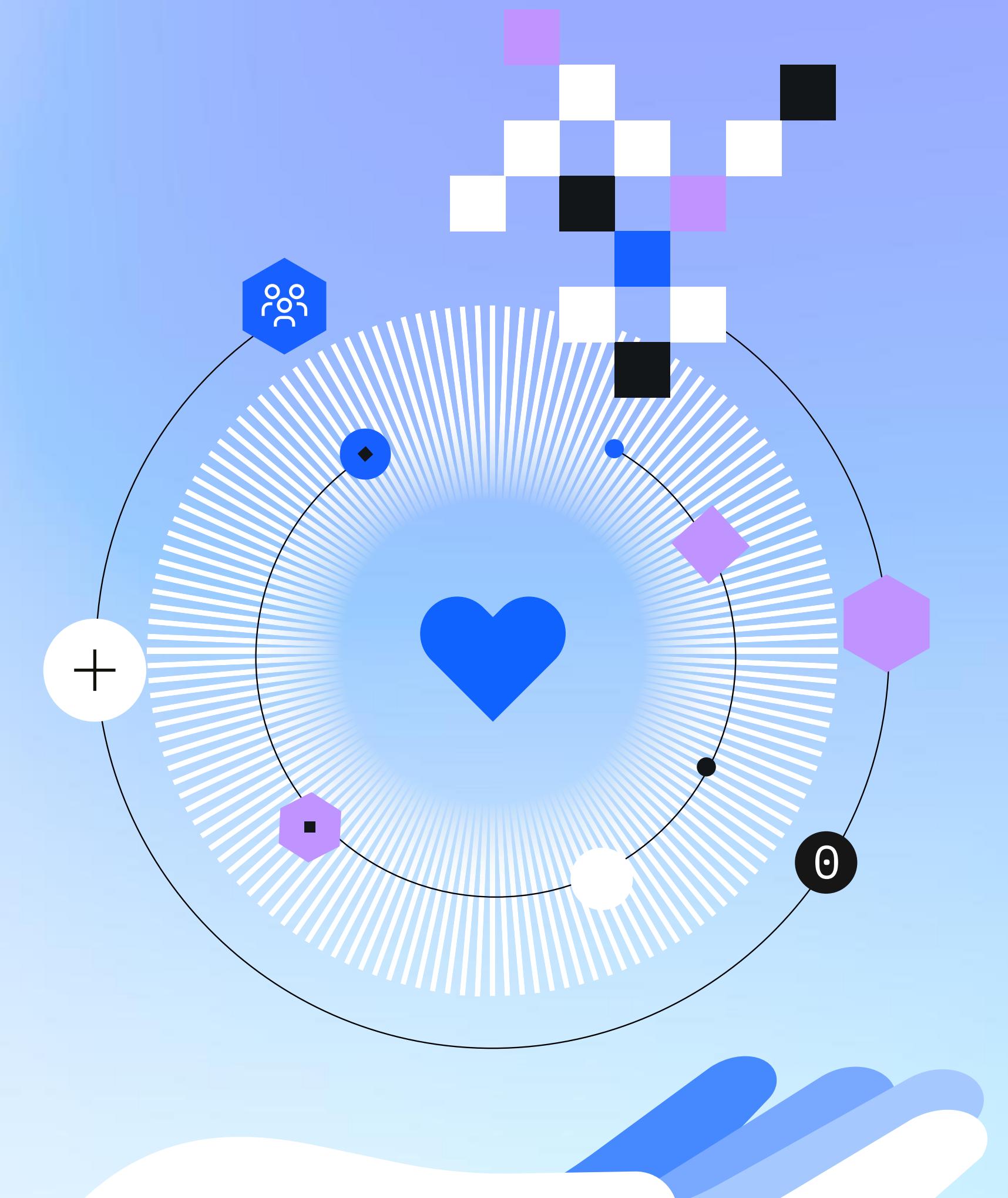


# KI-Chatbots im öffentlichen Sektor

## So finden Sie das passende Modell



# Künstliche Intelligenz – Erwartungen und Möglichkeiten

Die öffentliche Verwaltung steht vor vielfältigen Herausforderungen. Da sind zum einen der zunehmende Fachkräftemangel, komplexer werdende Prozesse und eine sich stetig ändernde Rechtslage. Digitalisierung, Wirtschaftlichkeit und Krisenresilienz sind als Ziele gesetzt. Zum anderen steigen die Erwartungen der Bürger\*innen an die Zugänglichkeit von Informationen, analog zu einer rasanten technologischen Entwicklung. Vor diesem Hintergrund suchen Behörden nach Lösungen, um ihrer Informationspflicht effizient nachkommen zu können.

Schritt zu halten mit den neuen Technologien, die die Erwartungen der Bürger\*innen befeuern, ist zwar eine Herausforderung an sich. Aber die technologischen Möglichkeiten der Künstlichen Intelligenz auszuschöpfen, kann für Behörden zum viel beschworenen Game Changer werden!

So können z. B. KI-Chatbots den Zugang und die Auffindbarkeit von Informationen erleichtern, Bürger\*innen-Anfragen in erster Instanz beantworten und repetitive Aufgaben übernehmen.

Eine Hürde auf dem Weg zum entlastenden Bot: Die Wahl des passenden KI-Modells bei der Vielzahl der Möglichkeiten und dem ähnlich scheinenden Leistungsspektrum der Modelle. Jedes (Sprach-)Modell hat einzigartige Qualitäten, die es entscheidend machen, die jeweiligen Fähigkeiten mit den spezifischen behördlichen Bedürfnissen und rechtlichen Anforderungen in Einklang zu bringen.

[Wir zeigen Ihnen anhand eines Beispiels, wie die Auswahl gelingen kann und Sie durch den Modell-Wirrwarr steuern können.](#)

# Neue Anforderungen von neuen Zielgruppen

Viele öffentliche Institutionen stehen vor ähnlichen Herausforderungen: Sie haben den Auftrag, Bürger\*innen mit teils komplexen Informationen zu versorgen.

## Dezentrales Informationsangebot

Im digitalen Raum stehen diese Inhalte vereinzelt auf Webseiten oder aber als Downloads, oftmals in Form von PDFs bewährter Printprodukte wie Broschüren, Merkblätter oder Formulare, zur Verfügung. Den Ansprüchen an moderne Behördenkommunikation wird beides nicht mehr hinreichend gerecht.

Das Informationsverhalten verändert sich: Bürger\*innen informieren sich größtenteils online und erwarten dabei eine konkrete Antwort auf ihre spezifische Frage. Die passende Antwort möchten sie nicht lange suchen müssen.

## Veränderte Erwartungen

Bürger\*innen möchten auf sie zugeschnittene Informationen und einen individuellen, einfachen und schnellen Zugang zu Inhalten, die sie brauchen. Dieses Informationsverhalten können Print- und statische Onlineprodukte nicht mehr bedienen.

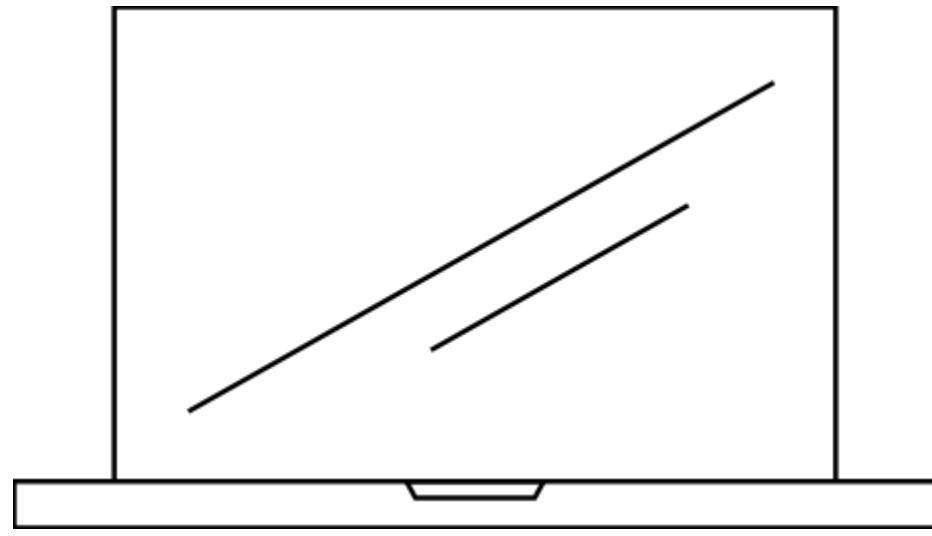
## Kennen Sie das?

- komplexe Themen
- nicht fürs Digitale optimiert
- dezentral zugänglich

# Neue Anforderungen von neuen Zielgruppen

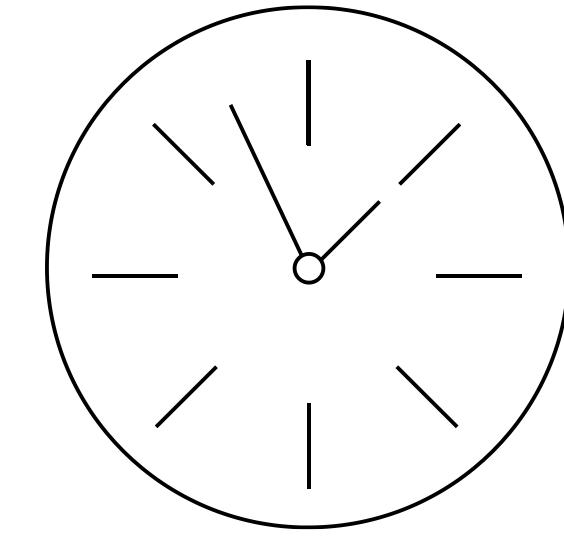
95 %

der Bevölkerung ab 14 Jahren nutzen das Internet.



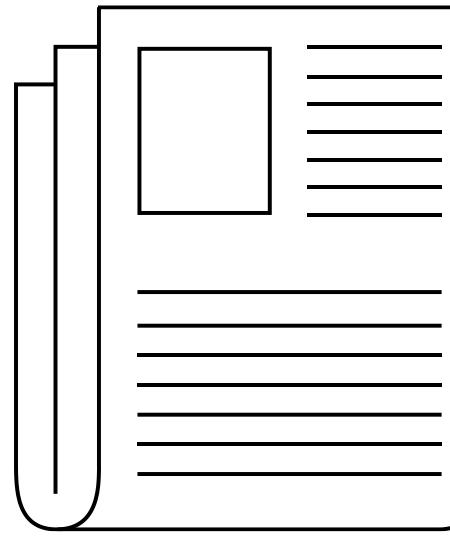
54 Sek

verbringen Menschen durchschnittlich auf einer Website.



55 %

aller Abonnements bei der F.A.Z sind mittlerweile Online-Abos. Andere Zeitungen zeigen einen ähnlichen Trend.



# Neue Anforderungen von neuen Zielgruppen

Telefonische Erreichbarkeit über Bürgerhotlines ist keine hinreichende Option: Der demografische Wandel bedingt einen Fachkräftemangel, was wiederum Bottleneck-Situationen in den öffentlichen Service Centern mit sich bringt.

Der Bedarf nach einem nachhaltigen Kommunikationskonzept für die öffentliche Hand wird auch angesichts wirtschaftlicher Faktoren deutlich: Persönliche Beratung zu gewährleisten, ist nicht nur organisatorisch herausfordernd, sondern auch kostenintensiv.

Darüber hinaus sind Printkosten während der Corona-Pandemie drastisch gestiegen und auf hohem Preisniveau geblieben. Abgesehen vom Kostenanstieg bei Printprodukten geht der Trend eindeutig zur Nutzung digitaler Medien.

Die Lösung: Hotlines und Berater\*innen entlasten, Zugänglichkeit von Informationen verbessern.

Für eine zukunftsfähige und inklusive Kommunikation ist die Digitalisierung von Informationen mit klarem Nutzerfokus unerlässlich.

Wie können Behörden ihrem Anspruch gerecht werden, Bürger\*innen auf einem hohen Niveau individuell zu informieren und zu beraten?

**840.000**

Vollzeitkräfte werden voraussichtlich bis zum Jahr 2030 im öffentlichen Dienst fehlen.

**<30%**

höhere Kosten für Papier im Jahr 2022 im Vergleich zum Vorjahr. Seitdem auf konstant hohem Niveau.

Quellen: Handelsblatt: Im Öffentlichen Dienst werden 840.000 Fachkräfte fehlen, 01/2023; Statista (2022): Entwicklung der Verkaufspreise im Großhandel mit Papier, Pappe, Schreib-, Bürobedarf in Deutschland von Oktober 2021 bis Oktober 2022

So navigieren Sie  
durch die KI-Optionen  
für Ihren Chatbot

# Qualitätsgeprüfte Unterstützung: Der Chatbot als digitaler Bürger\*innen-Berater

Chatbots, die von Künstlicher Intelligenz gesteuert werden, können qualitätsgeprüfte Antworten auf individuelle Fragen liefern. Hinter der menschenähnlichen Konversationsfähigkeit eines KI-Chatbots liegen große Sprachmodelle (Large Language Models, zum Beispiel Chat GPT oder Llama).

Diese werden im Vorfeld mit Milliarden von Daten trainiert. Zur Qualitätssicherung kann ein externer Wissensspeicher an ein Sprachmodell angebunden werden, der nur geprüfte Informationen enthält.

Das Sprachmodell beantwortet Nutzerfragen dann auf Basis des gesicherten Wissensspeichers – nicht aufgrund von anderweitig eruierten Inhalten.

Dieses Verfahren nennt sich „Retrieval Augmented Generation“ (RAG) und sollte gerade bei komplexen Sachverhalten eingesetzt werden.

Wer kann erste Anlaufstelle sein und so Berater\*innen entlasten?

Wer beantwortet initiale Fragen von Bürger\*innen?

Wer kann erste Rückfragen zur Problemeingrenzung stellen?

Wer kann weitere Information zur individuellen Beratung abfragen?

**KI-Chatbots**

# Mit RAG machen Sie Informationen schnell und sicher zugänglich

Retrieval Augmented Generation kombiniert Techniken des Informationsabrufs (Retrieval) mit den Fähigkeiten von Sprachgenerierungsmodellen (Augmentation). Somit können leicht verständliche und korrekte Antworten auf Fragen generiert werden (Generation).

Das bedeutet: Mittels der RAG-Architektur kann ein Chatbot passende Informationen aus Dokumenten extrahieren und in einer Sprache präsentieren, die von Bürger\*innen auch verstanden wird.

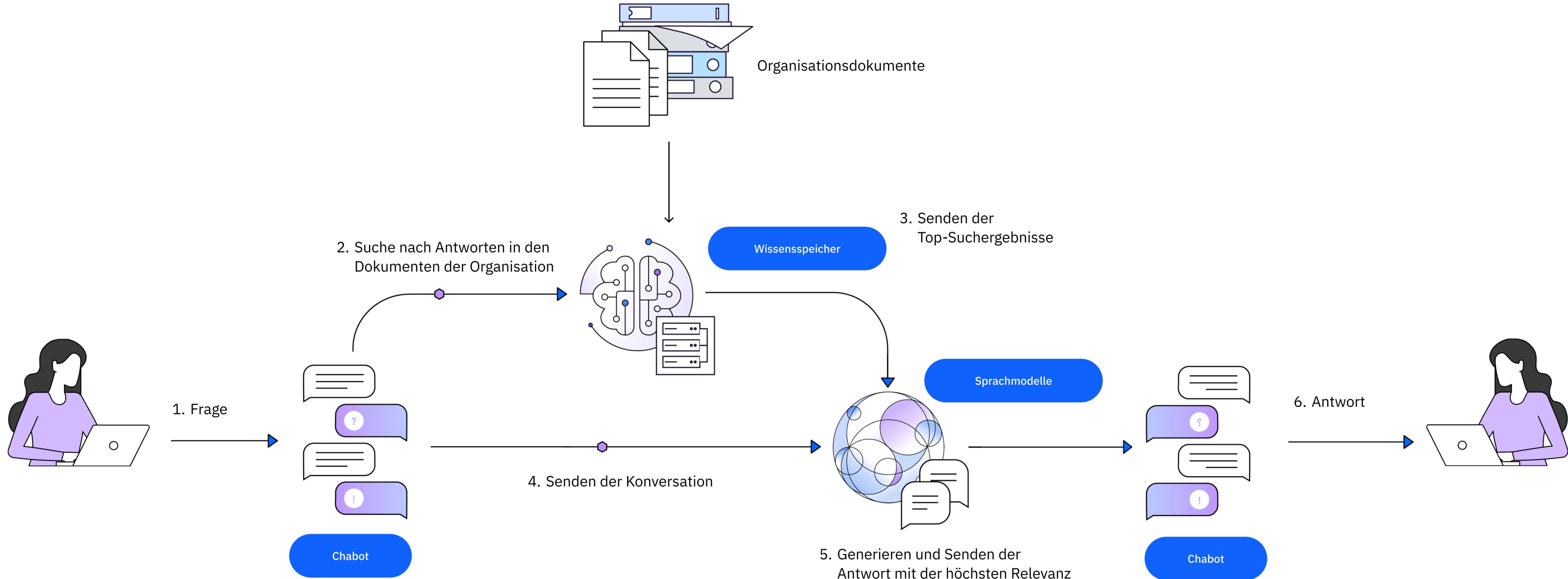
Die Fähigkeit der Sprachmodelle, komplexe Sachverhalte in natürlicher Sprache wiederzugeben, birgt ein großes Potenzial für die öffentliche Verwaltung.

Der RAG-Ablauf verkürzt:

- **Dokumente** (zum Beispiel FAQs, Informationsunterlagen, Checklisten, Merkblätter...) liegen im **Wissensspeicher** ab
- **Sprachmodell** generiert Inhalte aus den Informationen
- **Chatbot** interagiert mit Bürger\*innen, als virtueller Berater, nimmt Fragen an und antwortet



# So funktioniert RAG (Retrieval Augmented Generation)

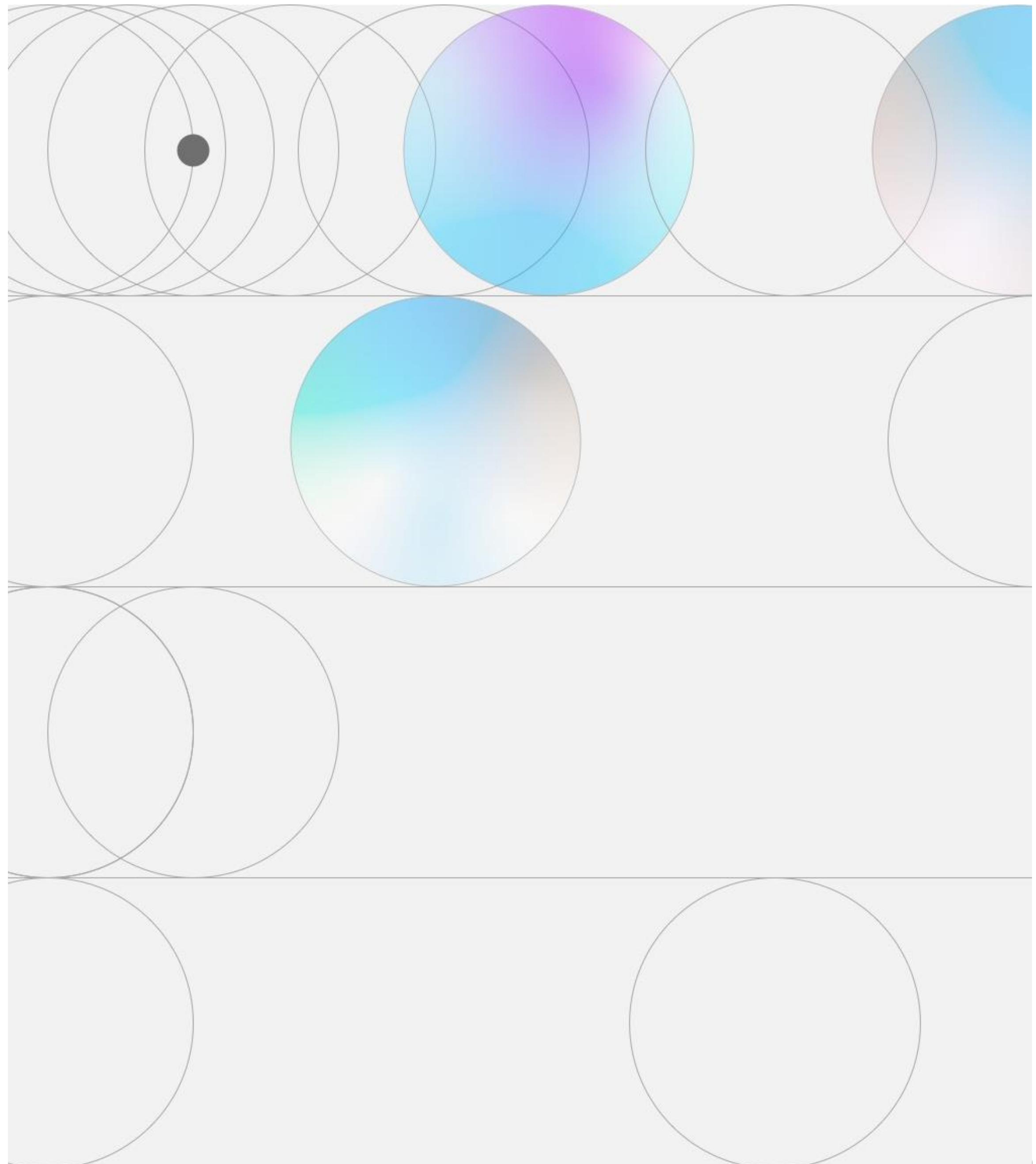


# Large Language Models – in RAG so klug wie der Wissensspeicher

Große Sprachmodelle können erkennen, wie Wörter statistisch zusammenhängen, wissen aber nicht, was sie bedeuten. Um die Modelle zu befähigen, mit Wörtern zu arbeiten, werden diese in Zahlenwerte umgewandelt. Ähnliche Wörter haben ähnliche Zahlenwerte. Das kann auch dazu führen, dass Sprachmodelle teilweise falsche oder unsinnige Antworten liefern.

Durch RAG wird die Qualität der generierten Antworten verbessert – zum einen durch den externen Wissensspeicher. Dort haben die Chatbots Zugriff auf das Expertenwissen. Zum anderen, weil die Einbettung der Zahlenwerte dem Sprachmodell den entsprechenden Kontext liefert. Anders ausgedrückt ruft das RAG-System aus dem Wissensspeicher die Informationen ab, deren Zahlenwerte mit der Eingabe der Nutzenden übereinstimmen oder dieser ähnlich sind.

→ Innerhalb eines RAG-Modells lässt sich die Antwortqualität immer weiter verbessern.



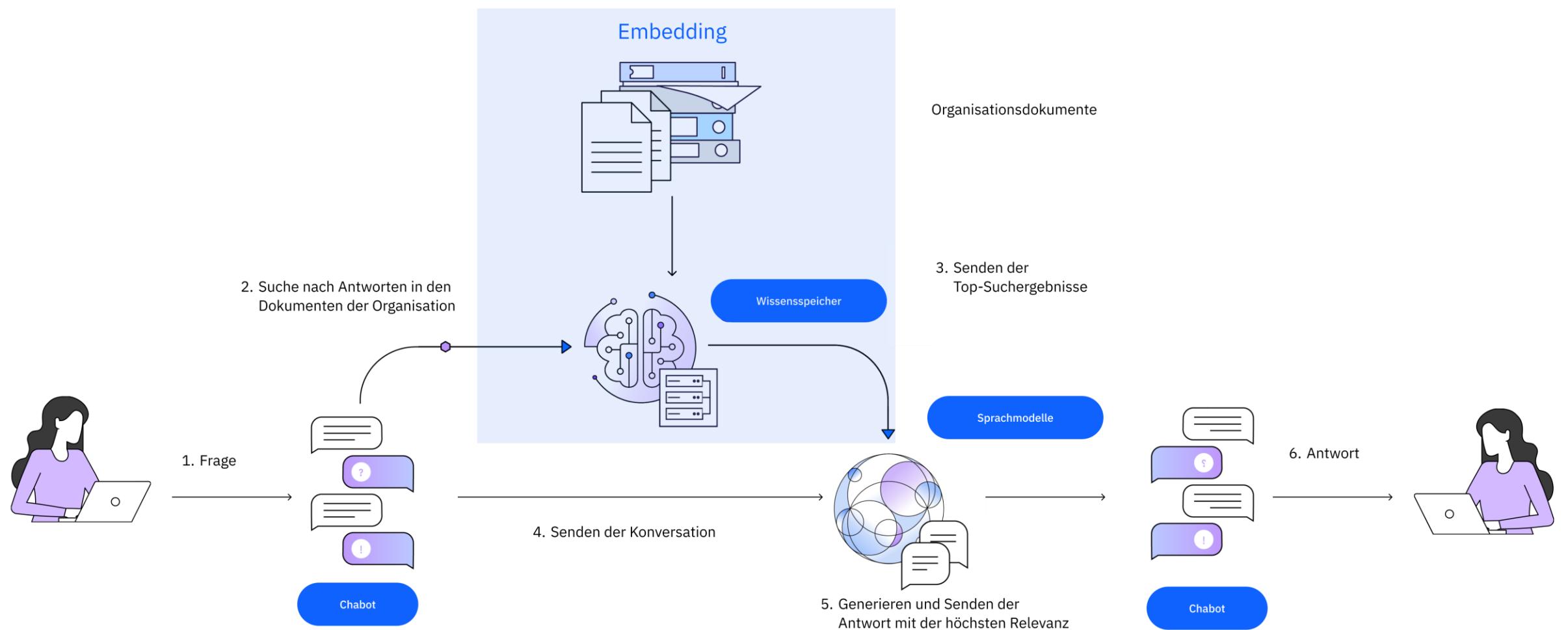
# Embedding-Modelle – semantische Einheiten werden zu Zahlenwerten

Embedding-Modelle sind wie „Zahlenkarten“ für Wörter oder Sätze. Sie wandeln Sprache in Zahlen um, damit Computer sie besser verstehen können. Ähnliche Wörter haben ähnliche Zahlenwerte, was der KI hilft, Bedeutungen zu erkennen.

Im RAG-System werden diese Embeddings gespeichert, um Wissen zu sammeln.

Wenn eine Frage gestellt wird, sucht das System mithilfe der gespeicherten Embeddings nach relevanten Informationen und verwendet sie, um die Antwort zu verbessern.

Es wird also nach Textpassagen gesucht, die numerisch gut zu der Frage passen, also ähnliche Zahlenwerte haben. Durch manuelle und automatisierte Tests ermitteln wir das beste Embedding-Modell. Das gewählte Embedding-Modell kann durch Chunking weiter optimiert werden.



# Chunking – häppchenweise Inhalte für das RAG-System

Beim Chunking werden große Texte in kleinere Abschnitte, sogenannte „Chunks“, aufgeteilt. Diese Chunks werden dann separat verarbeitet durch das Embedding-Modell und als eine Zahlenfolge gespeichert.

Wenn eine Frage gestellt wird, kann das System schnell die passenden Abschnitte finden und für die Antwort verwenden.

So wird es einfacher, gezielt Informationen aus großen Texten zu holen. Für die Sprachmodelle ist es wichtig, dass die richtige Menge (nicht zu viel und nicht zu wenig) an Informationen in den Abschnitten ausgegeben werden.

→ Chunks als besser zu verarbeitende Informationshappen für das RAG-System

IBM® SPSS® Statistics is a comprehensive statistical software solution that is widely used for data preparation and management, advanced data analysis techniques, forecasting models, and data visualizations. It is designed to help not only researchers, data analysts, statisticians, and educators but also business professionals extract actionable insights from their data. Our solution offers a powerful suite of tools that can address any type of analysis, from basic descriptive statistics to complex multivariate methods.

With support for multiple data formats and an intuitive interface, SPSS Statistics transforms data into valuable insights and engaging visualizations through graphs and charts. Its comprehensive features in data preparation, statistical modeling, and reporting enable organizations and data professionals to make informed, data-driven decisions. This helps improve strategies across various fields, including social sciences, government, healthcare, marketing, finance, education, and more.

IBM SPSS Statistics 30.0.0 introduces a comprehensive set of new features and improvements, designed to optimize both data analysis capabilities and user experience. It brings advanced statistical methods, including the introduction of Bland Altman Analysis and new normality tests, making it easier for users to gain deeper insights from their data. In addition to new capabilities, the latest interface upgrades, including dark mode and 4K HD Monitor for Windows, aim to enhance the user experience.

Whether you are conducting complex analyses or managing large datasets, SPSS Statistics 30.0.0 offers performance improvements that can help address your data analysis needs.

Chunk A

Chunk B

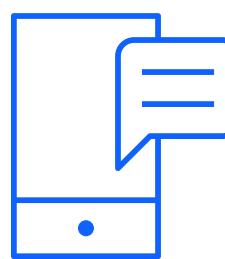
Auf der Zielgeraden zu  
Ihrer Chatbot-Lösung

# Benchmarking: Das passende Setup

Ein gut funktionierender Chatbot besteht aus der Bot-Oberfläche, dem im Hintergrund eingesetzten Sprachmodell und einem angebundenen Wissensspeicher. Auf dem Markt gibt es unzählige Anbieter für Komplett- und Einzellösungen, Sprachmodelle und Kombinationsmöglichkeiten.

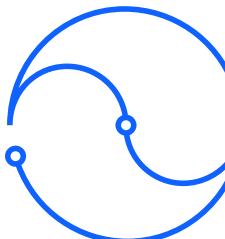
Durch einen Vergleich aller Optionen ermitteln Sie die beste Anwendung.

Wir haben für Kund\*innen einen solchen Vergleich mit einem objektiven Fragebogen durchgeführt, der an ausgewählte Anbieter gesendet wurde. Die abgefragten Kriterien sind gerade im Bereich Datenschutz und Beschaffung speziell an die Bedürfnisse der öffentlichen Verwaltung angepasst. So haben wir nicht nur die beste Lösung gefunden, sondern auch herausgearbeitet, welche Funktionalitäten eine mögliche individualisierte Lösung bereitstellen muss.



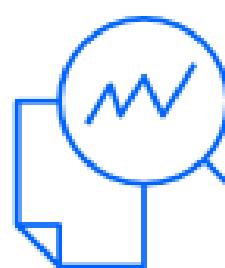
## Chatbot-Oberfläche

Ermöglicht eine benutzerfreundliche, ästhetisch ansprechende und barrierefreie Nutzererfahrung.



## Wissensspeicher und dessen Abfrage

Ermöglicht effiziente, kontextbezogene Abfragen und kontinuierliche Aktualisierungen, birgt jedoch Herausforderungen in Bezug auf Datenqualität und Datenschutz.



## Einsatz großer Sprachmodelle

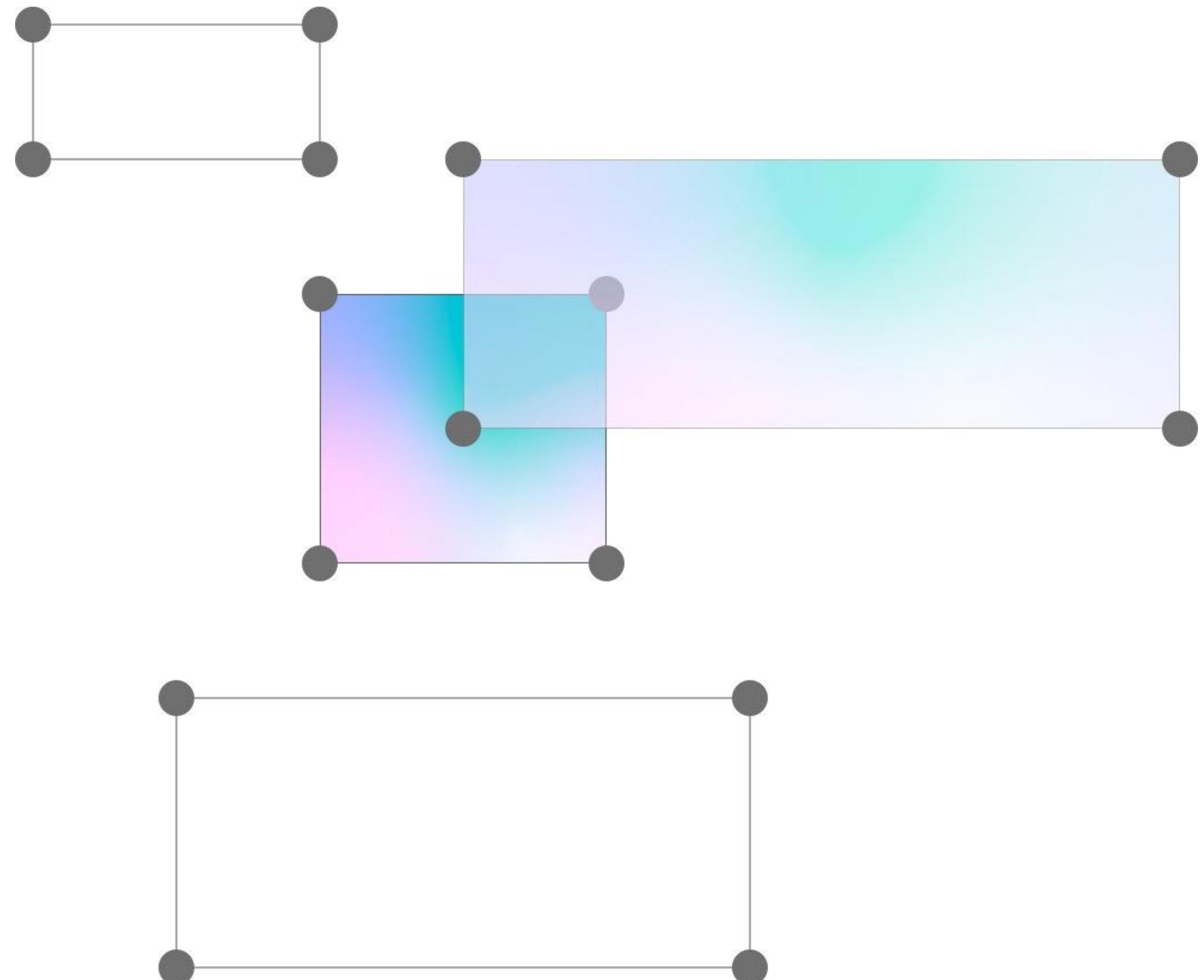
Ermöglicht intuitive und individuelle Interaktionen in natürlicher Sprache.

# Fragenkatalog:

# Der passende Anbieter

In unserem Beispielprojekt haben wir, um für unseren Kunden den passenden Anbieter zu finden, einen Katalog mit 208 relevanten Fragen erstellt und einer Reihe von infrage kommenden Anbietern zur Beantwortung zugeschickt.

Der Fragenkatalog enthält folgende Oberkategorien und darunter liegende Kriterien:



- **Beschaffung:** Gibt es initiale Anschaffungskosten oder nur laufende Kosten?
  - **Sicherheit und Datenschutz:** Wird die DSGVO eingehalten?
  - **Technologische Fähigkeiten** des Sprachmodells und der LLM-Plattform, des Wissensspeichers, des Chats: Ist eine Programmierschnittstelle verfügbar?
  - **Nutzerfreundlichkeit und Support:** Ist die Oberfläche barrierefrei?
  - **Hosting:** Werden lokale Bereitstellungen unterstützt?
  - **Zukunftssicherheit:** Gibt es regelmäßige Updates und Weiterentwicklungen?

# Vergleichsfragen: Die Tests zur Antwortqualität

Auf Basis der Ergebnisse aus dem Fragenkatalog haben wir anschließend verschiedene Sprachmodelle getestet und miteinander verglichen.

Dazu erstellten wir eine Liste von etwa 120 immer gleichen Beispielfragen, die jedes Modell beantwortet, um zuvor festgelegte Anforderungen zu testen.

Die Antworten der Sprachmodelle werden nach verschiedenen Kriterien überprüft.

Diese sind in zehn Bereiche aufgeteilt:

- Frageverständnis
- Richtigkeit, Relevanz und Präzision
- Klarheit und Verständlichkeit
- Konsistenz
- Kohärenz und Kontext
- Persönliche Ansprache und Empathie
- Interaktivität und Engagement
- Antidiskriminierung
- Abgrenzung zu nicht-relevanten Fragen
- Exitstrategie

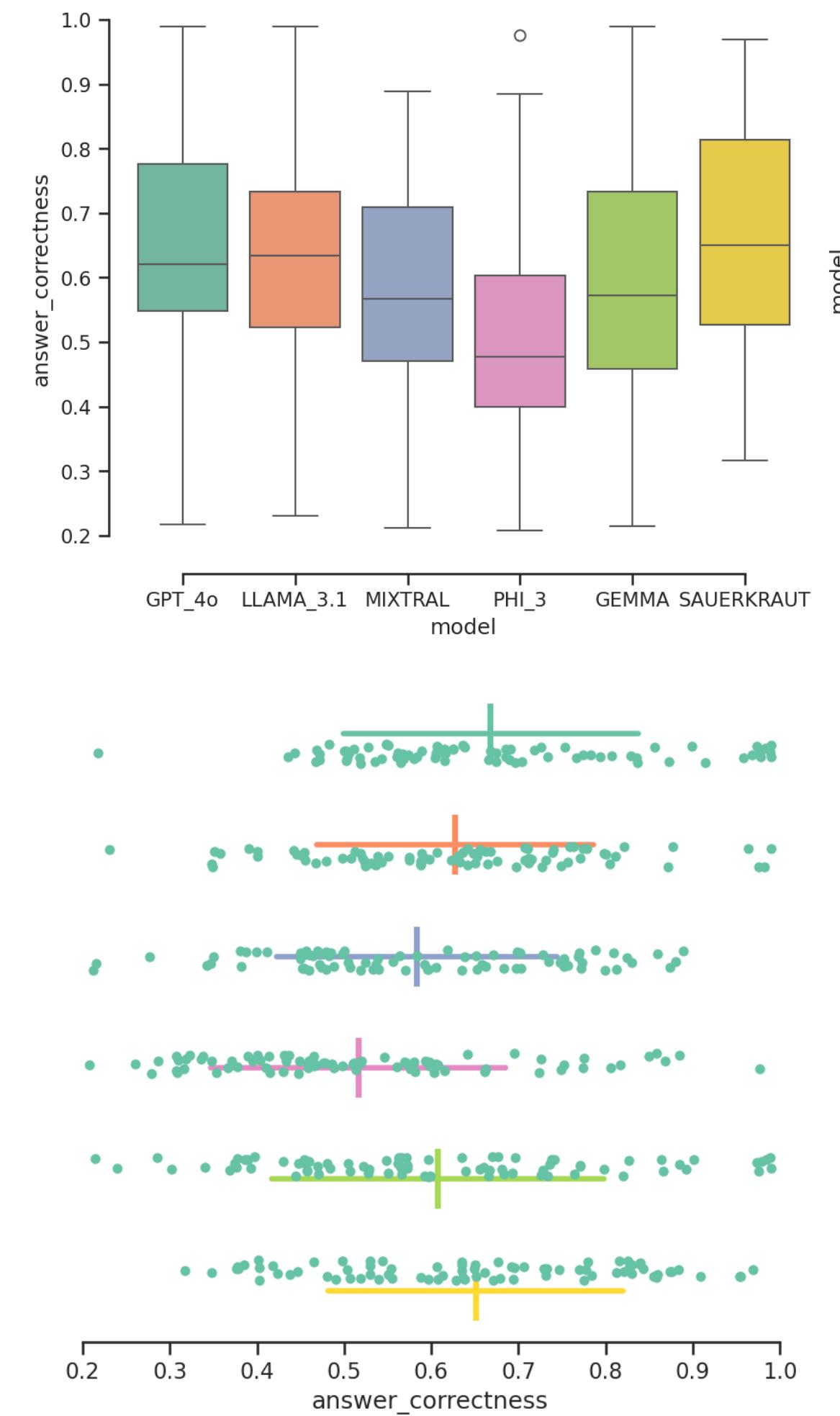
## Antworten auswerten mittels des RAGAS-Scores

Die Antworten der Sprachmodelle werden nach verschiedenen Kriterien überprüft und getestet. Diese Tests laufen teilweise manuell, teilweise automatisiert ab. Für die automatisierten Tests wird der RAGAS-Score genutzt.

RAGAS ist eine Reihe verschiedener Metriken, die bewerten, wie gut die Antworten von KI-generierten Texten eines RAG-Modells sind.

Mit Hilfe der Metriken lässt sich zum Beispiel berechnen, wie gut eine Antwort zu einer Frage passt oder ob eine Antwort alle relevanten Inhalte enthält (Answer Correctness und Faithfulness).

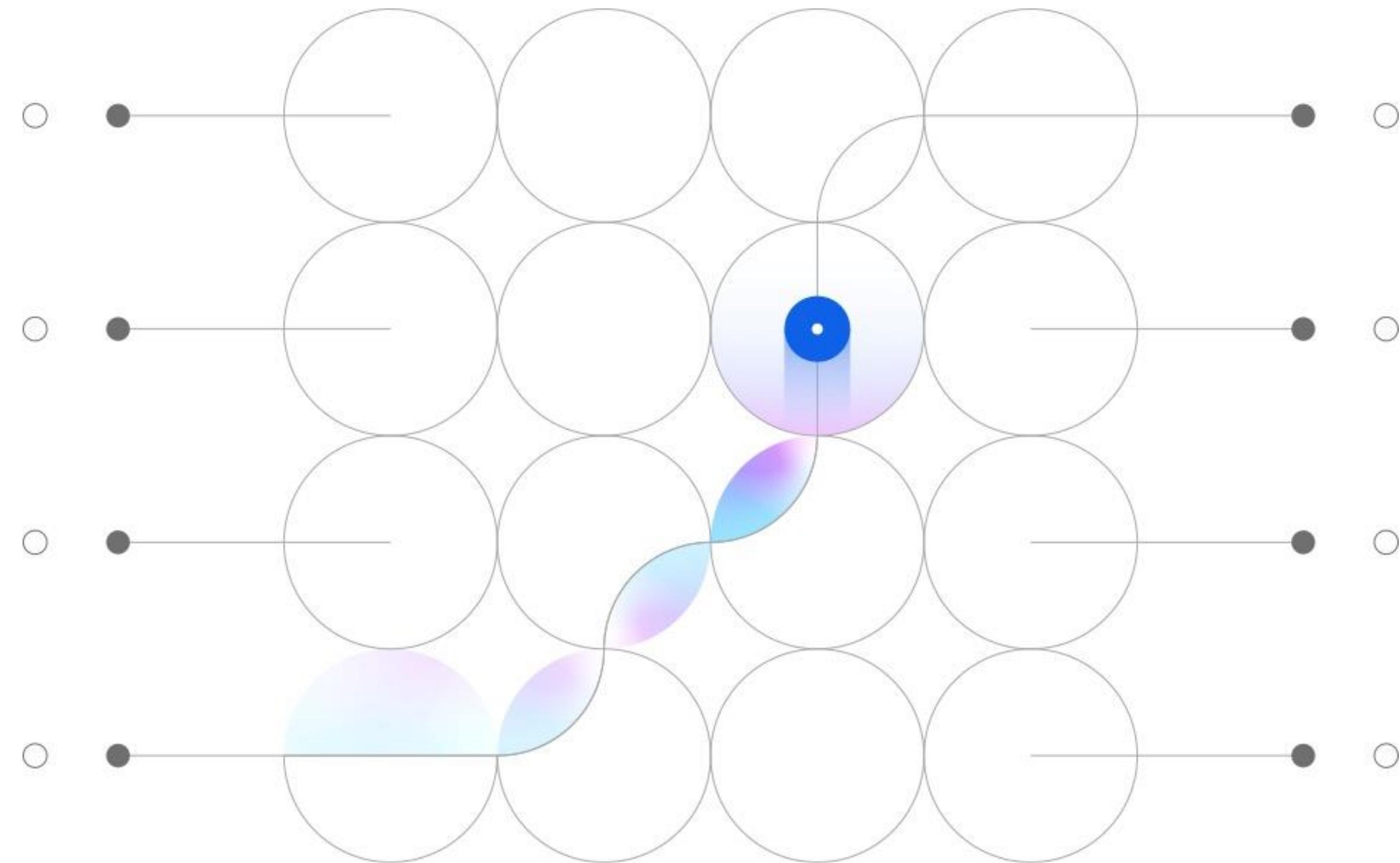
Mithilfe dieser Analyse und der qualitativen Analyse durch Fachexpertinnen und -experten können verschiedene LLMs und verschiedene Entwicklungsstände eines Chatbots miteinander verglichen werden.



Beispiel für die Auswertung zur Answer Correctness und Faithfulness.

## Testergebnis: Unsere Favoriten unter den Sprachmodellen

Die zwei am besten geeigneten Sprachmodelle sind mit dem Stand Q4 2024 nach unserer Auswertung **GPT-4o** und **Llama 3.1**. Der Markt ist jedoch sehr schnelllebig und diese Einschätzung kann sich jederzeit ändern.



Die wichtigsten Überlegungen für die Verwendung im öffentlichen Bereich sind:

### 1. Open Source oder Closed Source?

Bei Open Source Modellen ist der Code frei zugänglich. Das führt einerseits zu einer agileren Entwicklung, andererseits können Schwachstellen im Code so eher ausgenutzt werden.

### 2. Lokales Hosting oder Cloud?

Lokal gehostete Software auf eigener Infrastruktur ist der sicherste Ort – allerdings auch kostspielig. In einer dedizierten Cloud ist der Service isoliert von anderen Organisationen und bietet mehr Sicherheit als in einer öffentlichen Cloud.

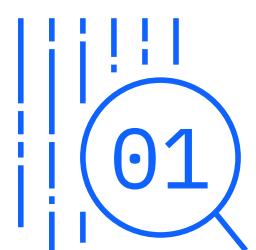
### 3. Daten speichern oder nicht?

Wenn Daten gespeichert werden, müssen Nutzende aufpassen, was sie eingeben. Wenn sensible oder persönliche Daten verarbeitet werden sollen, sollten Daten nicht gespeichert werden.

Technologie ist nichts  
ohne den  
**Faktor Mensch**

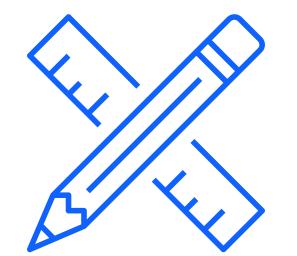


# Der Mensch im Mittelpunkt – Wichtige Überlegungen über das KI-Modell hinaus



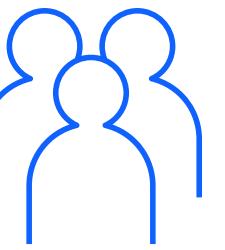
## Daten & Technologie

Verfügen wir über die richtige Informationsarchitektur? Haben wir qualitativ hochwertige, verwaltete, sichere und kontrollierte Daten? Verfügen wir über die richtigen Tools zur Bereitstellung und Feinabstimmung unserer Modelle?



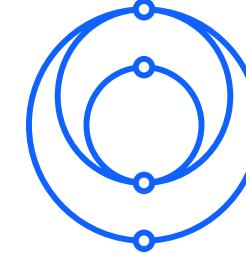
## AI Engineering & Operations

Welche KI-Tools verwenden wir? Wie setzen wir KI ein und verwalten sie?



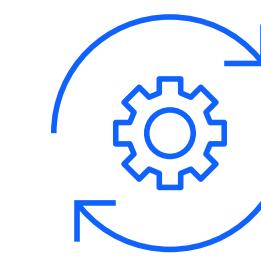
## Menschen

Wie denken die Mitarbeitenden? Wie gewährleisten wir KI-Skills und ermöglichen eine Wachstumsmentalität, die experimentelles Verhalten fördert?



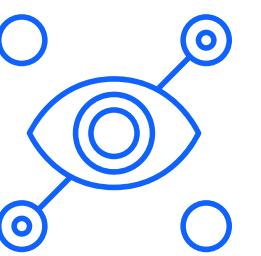
## Veränderungsmanagement

Wie arbeiten und handeln wir? Agil? Wasserfall? Wie fördern wir Vertrauen in und Akzeptanz für KI? Wie komplex ist unsere Organisation?



## AI Operating Model

Wie nutzen wir Generative KI und wenden sie an? Sind Automation und KI in unsere internen Systeme eingebettet und/oder werden sie in unseren B2B-Prozessen eingesetzt?



## Vision & Strategie

Gibt es in meinem Unternehmen eine gemeinsame Vision für Generative KI über alle Geschäftsbereiche hinweg? Arbeiten wir alle abgestimmt auf dieselben Ziele hin?

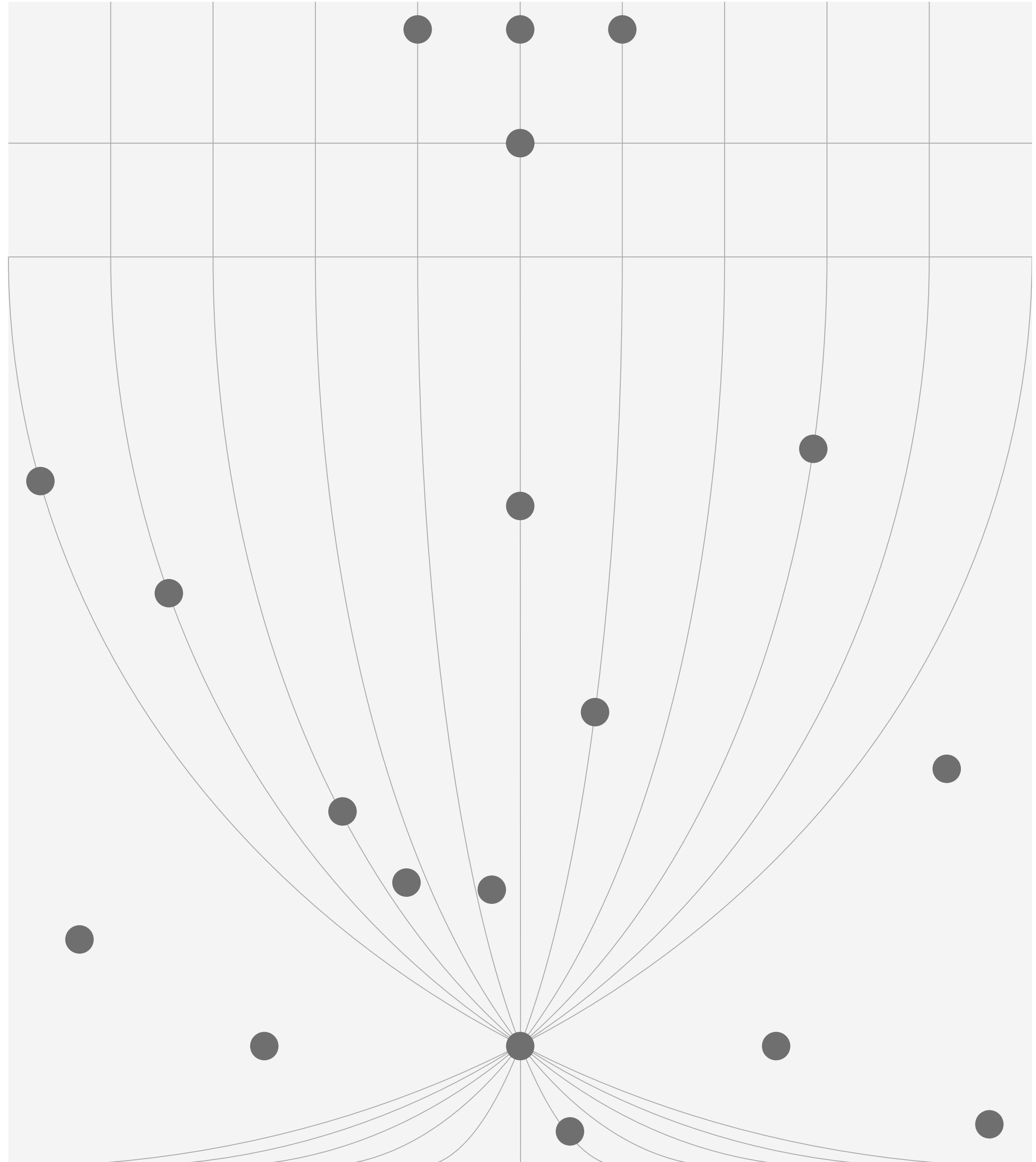
## Fazit: Chatbots sorgen für einen neuen Zugang zu Informationen

Künstliche Intelligenz in Form von Chatbots mit einem Retrieval-Augmented-Generation-System kann den [Herausforderungen des Fachkräftemangels](#) und der komplexen Informationsverarbeitung in der öffentlichen Verwaltung begegnen.

Durch den Einsatz von KI-gestützten Lösungen können Verwaltungen individuelle [Anfragen effizient und verständlich beantworten](#) – so kommunizieren Behörden bürgernah.

Ein strukturierter [Auswahlprozess für geeignete KI-Modelle](#) ist entscheidend, um die spezifischen Anforderungen an [Sicherheit](#), [Datenschutz](#) und [Nutzerfreundlichkeit](#) zu gewährleisten.

Chatbots revolutionieren den Zugang zu Informationen und bieten eine Chance, den Erwartungen der Bürger\*innen gerecht zu werden. Zumal von einer Fortsetzung des Trends auszugehen ist, dass Nutzende Informationen komfortabel suchen und finden möchten. Lösungen im Bereich [Conversational Design](#) werden umso wichtiger, als sie in vielen Fällen die [herkömmliche Informationssuche ersetzen](#).



## Kontakt

Sie möchten mehr erfahren oder haben Fragen?

Unsere Expert\*innen für Künstliche Intelligenz, Content und Conversational Design freuen sich auf Ihre Nachricht!

Lassen Sie uns besprechen, wie Sie heute anfangen können.



Quirin Johannes Koch

Digital Consultant

—

[quirin.koch@ibmix.de](mailto:quirin.koch@ibmix.de)

+49 174 9966899



Sara Stechow

Lead New Business & Partnerships

—

[sara.stechow@ibmix.de](mailto:sara.stechow@ibmix.de)

+49 174 9966948

Team

Maan Al Balkhi  
Expert Lead Data & AI

Eva-Maria Blank  
Practice Lead Content & Communication

Daria Kolesova  
Content Designer

Marie Oelgemöller  
Content Strategist

Mareka Stolle  
Project Lead

## Über IBM iX

Mit unserer Initiative [Staat Digital](#) treiben die Digitalisierung im öffentlichen Sektor voran. Dabei stets im Mittelpunkt: Die Menschen.

Gemeinsam mit unseren Kundinnen und Kunden entwerfen wir nutzendenzentrierte Erlebnisse für heute und die Zukunft. Empathie, Analytics und fortschrittliche Technologien und Methoden sind dabei wichtige Bausteine für nachhaltigen Erfolg.

Mit über 17.000 Expertinnen und Experten in über 50 Studios weltweit zählen IBM Consulting und IBM iX zu den größten Anbietern für digitale Experiences weltweit. Wir sind Consultancy, Digitalagentur, Design-Studio und Tech-Company in einem.

# Let's collaborate

IBM iX

IBM iX Germany GmbH  
Chausseestraße 5  
10115 Berlin

AG Berlin Charlottenburg, HRB 181949 B

Geschäftsführung: Markus Dietrich, Samira Imsirovic-Kaya

[IBM iX Datenschutzerklärung](#)

[ibmix.de](#)

