

# Der Ulmer Spatz

---

## Die Integration kommunaler Chatbots in die Auskunftstätigkeit von Archiven

---

33. Fortbildungsseminar der BKK

Dr. Jeremias Schmidt | Schrift- und Tondokumentation | Stadtarchiv Ulm

Dominic Eickhoff | Fachberater für Digitalisierung | LWL-Archivamt für Westfalen

26. November 2025

# Was soll das immer mit diesen Chatbots?

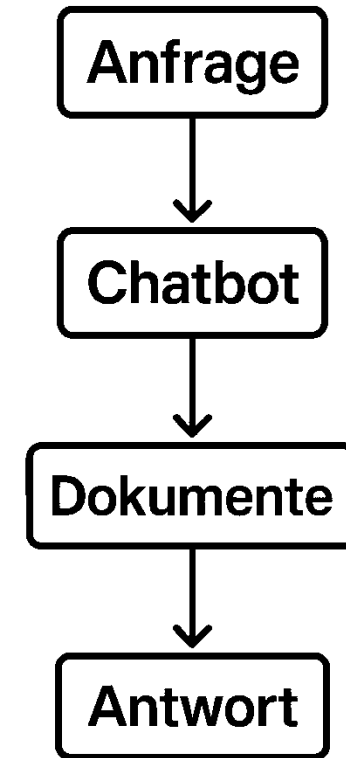
---

- Aktueller KI-Fokus: Texterkennung und Metadatenextraktion
- Generative KI kommt wenig zum Einsatz
- Neu: Chatbots (Large Language Models) als Arbeitsentlastung in der Beauskunftung?
- Welche Daten spielen eine Rolle?



# Chatbot mit RAG – was ist das überhaupt?

- RAG = Retrieval Augmented Generation / „wissensgestützte Textgenerierung“
- Wie ein:e Archivar:in erst nachschlagen, dann antworten.
- Bot durchsucht Wissensdatenbank (Archiv-Publikationen)
- Verwendet nur gefundene Informationen
- Vorteil: Weniger Halluzinationen, nachvollziehbare Informationen, Quellenangaben



# Warum der Ulmer Spatz?

- ✗ Eigener Bot hätte mehrere hundert Euro im Monat gekostet
- Stadt Ulm hatte 2024 bereits einen kommunalen Chatbot
- Ursprüngliche Funktion: Bürger:innenservice
  - Bereitgestellt durch Dienstleister
  - Wann kommt die Müllabfuhr? Wo ist das Bürgerbüro?
- Unsere Idee: Archivdaten per RAG anbinden
- Vorteil: Kosten teilen, Wissen des übergreifenden Chatbots gezielt erweitern
- Bewerbung auf den DIGITAL-Award 2026





# Die Datengrundlage

## 20.000 Seiten Stadtgeschichte

### Publikationen des Stadtarchivs:

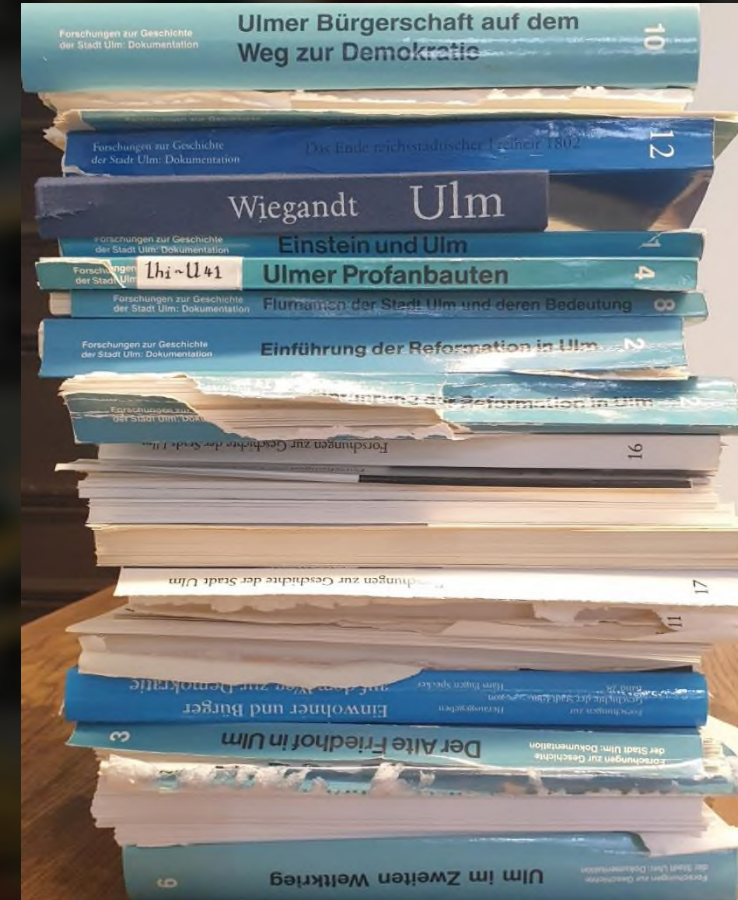
- Überblickswerke & Spezialmonografien
- Biographisches Lexikon
- Stadtchronik (tageweise Ereignisse)

### Herausforderung:

- Großteil der Bücher nicht digital/digitalisiert

### Lösung:

- Digitalisierung via Einzugs-Scanner (mittelmäßige Scans genügen)



Ein Teil der für den Spatz gescannten Publikationen



# Use-Cases: Wofür nutzen wir den Spatz?

## Funktioniert bereits gut:

- ✓ Generische Anfragen zur Stadtgeschichte beantworten
- ✓ Biographische Informationen bereitstellen
- ✓ Social-Media-Content co-kreieren
- ✓ Recherchen zu Themen ab 1945 (größere Informationsdichte)

## In Arbeit:

- 🔧 Verknüpfung mit EAD-XML-Erschließungsdaten (Findbuch.net)
- 🔧 KI-gestützte Stadtführungen



# Beispiel 1 – Generische Anfrage

Probieren Sie selbst:

- ulm.de → Spatz-Symbol (rechts unten)
- Beispielfrage: *"Erzähle mir etwas über Ulm in der napoleonischen Zeit"*

Das funktioniert gut:

- Strukturierte Antworten
- Quellenangaben
- Verweis auf Stadtarchiv für weitere Informationen und Archivgut



Der Ulmer Spatz auf der städtischen Homepage



# Beispiel 2 – Social Media

## Ulmer Finanzkrise der 1990er

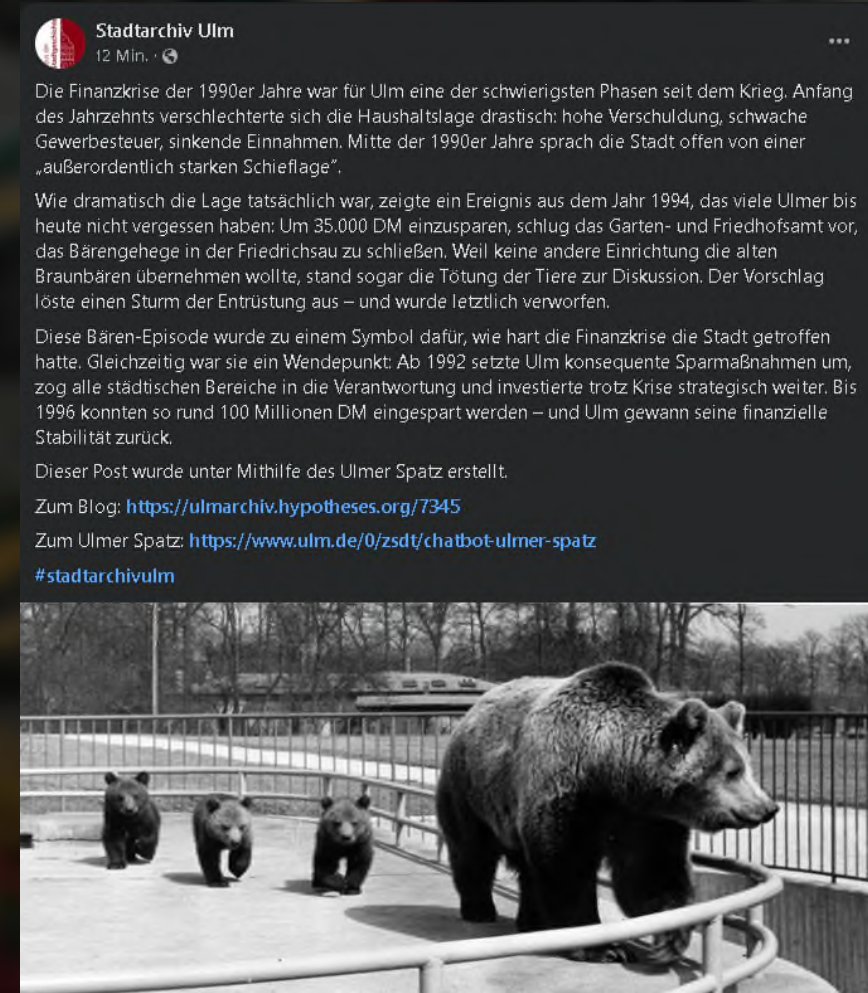
- Kontext: Aktuelle Finanzlage → historische Parallele

## Aufgabe für den Bot:

- *"Erstelle einen Social-Media-Post zur Ulmer Finanzkrise der 1990er"*
- *Ergänzt durch eigene Recherche zu den Ulmer Bären*

## Blogartikel:

- <https://ulmarchiv.hypotheses.org/>



Facebook-Post zur Finanzkrise der 1990er Jahre



# Problem 1 – Spezifische Fragen

*Wann der Bot an seine Grenzen stößt*

Beispiel: Ort der „Alte Pfarrkirche“ (Vorgänger des Ulmer Münsters)

- Fehler: Antworten & Verortungen varriieren
- Korrekt: Außerhalb der Stadtmauern (Alter Friedhof)

Warum?

- Information nur in EINER spezifischen Publikation
- System lädt begrenzte Anzahl Textabschnitte / sog. „Chunks“
- System findet die notw. Stelle oft nicht





# Problem 2 – Halluzinationen

*Wenn der Bot falsch kombiniert*

Pfarrkirchen-Beispiel (Fortsetzung):

- ✗ Bot findet keine eindeutige Antwort
- ✗ Kombiniert verschiedene Infos falsch
- ✗ Platziert Pfarrkirche auf den Weinhof (weil dort frühe Kirchenreste gefunden wurden)
- "Halluzinationen" = Bot erfindet neue Fakten
- Gegenmaßnahme: Optimierung des Systemprompts



Der Weinhof in Ulm – mit Stadtarchiv, aber ohne Kirche



# Problem 3 – Urheberrecht

## Rechtliche Grauzone

- Wem gehören die Rechte?  
Herausgeber:innen oder Autor:innen?

## Das Problem:

- Digitale Zweitverwertung in KI-Systemen
- Besonders komplex bei Sammelwerken

## Lösungsansatz des Archivs:

- Rechtsgutachten in Auftrag
- Autor:innen einzeln anfragen (Zweitlizenz)

## Herausforderung

- Viele Autor:innen verstorben (Erbensuche)

Autoren/Mitarbeiter	Buch/Band
<i>Dr. Günter Dukek</i>	<b>Forschungen zur Geschichte der Stadt Ulm</b> Reihe Dokumentation Herausgegeben vom Stadtarchiv Ulm Band 1 1979 <b>Einstein und Ulm</b> Festakt, Schülerwettbewerb und Ausstellung zum 100. Geburtstag von Albert Einstein Herausgegeben von HANS EUGEN SPECKER
<i>Hubert Fink</i> früher Leiter der Abteilung Neuzeit des Stadtarchivs Ulm	<b>Forschungen zur Geschichte der Stadt Ulm</b> Reihe Dokumentation Herausgegeben vom Stadtarchiv Ulm Band 1 1979 <b>Einstein und Ulm</b> Festakt, Schülerwettbewerb und Ausstellung zum 100. Geburtstag von Albert Einstein Herausgegeben von HANS EUGEN SPECKER  <b>Forschungen zur Geschichte der Stadt Ulm</b> Reihe Dokumentation Herausgegeben vom Stadtarchiv Ulm Band 7 1990 Schriftleitung: Professor Dr. Hans Eugen Specker, Ulm, Stadtarchiv,

Ausschnitt aus der Liste der Autoren  
(Publikationen Stadtarchiv Ulm)

# Lösungsansätze

*Wie wir helfen, das System zu verbessern*

## Metadaten-gestützte Suche

- Tags: "Kirchengeschichte", "Wirtschaft", "Biographie"
- Zeitraum: "Mittelalter", "19. Jahrhundert"
- Orte: "Münster", "Weinhof"

*Vorteil: Bot findet gezielt die richtigen Dokumente*

## Optimierung Systemprompt

- Bei Unsicherheit: Ehrlich zugeben
- Keine Informationen erfinden

```
json
{
  "text": "Die alte Pfarrkirche St. Martin...",
  "metadata": {
    "publication": "Das Ulmer Münster 1377-1890",
    "topics": ["Kirchengeschichte", "Münster"],
    "period": "Mittelalter"
  }
}
```

Beispiel-JSON mit Metadaten



# Strategisches Herangehen

---

- Woher die Kompetenzen nehmen?
- Welchen rechtlichen Rahmen beachten?
- Wie die Finanzierung aufstellen?

# Rechtliche Rahmenbedingungen

- Natürlich die Archivgesetze der Bundesländer/ des Bundes
- DSGVO & BDSG (Bundesdatenschutzgesetz)
- EU AI Act (dt.: KI-Verordnung/ KI-VO): Transparenz & Risikomanagement
- EU Data Act: Fairer Umgang mit Daten
- Text & Data Mining (TDM): DSM-Richtlinie (EU) 2019/790, Art. 3 & 4
- UrhG (Urheberrechtsgesetz)
  - UrhG §44b zu Text und Data Mining
  - UrhG §53 zu Privatkopien
  - UrhG §60c und §60d zu wissenschaftlicher Forschung
- Creative Commons-Lizenzen: Upload & Vervielfältigung mitdenken
- Barrierefreiheit in der IT beachten: BFSG, BGG, BITV 2.0, WCAG



# Die Regeln Ihrer Institutionen nicht vergessen!

- KI-Governance?
  - Compliance-Richtlinien?
  - Datenstrategien?
  - Open Data Policies und Open Data Portale?
  - IT-Sicherheitsrichtlinien?
  - Cloudnutzungsrichtlinien?
  - IT-Abteilung fragen: Welche Tools werden erlaubt/ bereitgestellt?
- Keine Redundanzen erzeugen und/ oder doppelt beschaffen!

# Mögliche Finanzierung

*Für Drittmittel kreativ werden & Kooperationen anstreben:*

- Smart City-Projekte (Überblick auf der Plattform [Urban.digital](#))
- Lokale Stiftungen, Fördervereine und Sparkassen ansprechen
- [Kulturstiftung der Länder](#): Bspw. „[Kultur vermitteln](#)“
- „[Heimat-Ministerium NRW](#)“: Bspw. [Heimat-Scheck](#) für Kleinstprojekte
- [Kulturstiftung des Bundes](#): Bspw. „[TRAFO – Modelle für Kultur im Wandel](#)“
- [Land.Heimat.Innovativ](#) (BMLEH): Bspw. „Unterstützung von Ehrenamt, Kultur, etc.“
- [Bundesministerium des Inneren](#): Bspw. „Digitalisierung des Engagements“
- Deutsche Forschungsgemeinschaft: Bspw. „[Förderinitiative Künstliche Intelligenz](#)“
- [LEADER](#) (EU-Mittel): [Strukturentwicklung ländlicher Räume in NRW](#)
- [ECCCH](#) (EU-Mittel): Datenprojekte in Kulturerbeinstitutionen ([EU Funding Portal](#))



# (Simple) Prompt-Beispiel zur Textgenerierung mit Kontextwissen für einen Chatbot

*"Als erfahrener Archivar sollst du eine hilfreiche Antwortvorlage für einen Archivnutzer erstellen, der bisher nur eine ungenaue Vorstellung von seinem Informationsbedürfnis hat. Nutze die im RAG-System verfügbaren archivischen Erschließungsangaben (bspw. EAD-XML oder Richtext), um eine strukturierte und informative Antwort zu formulieren. Die Antwort sollte den Nutzer durch die verfügbaren Informationen leiten und anhand der im RAG verfügbaren Literatur auch schon mögliche Forschungsfragen aufzeigen, die zu seinem Interesse passen. Die Antwortvorlage soll ca. 15 Sätze haben. Anschließend stellst du das Ergebnis als txt-Datei zum Download zur Verfügung."*

Hinweis: Diese Art von Prompt funktioniert nur mit genügend Kontext, welches dem Sprachmodell innerhalb des aktiven Kontextfensters oder durch eine Wissensdatenbank (s. RAG) zur Verfügung gestellt wird. Allerdings kann es selbst dann noch zu Halluzinationen kommen. Alle Ergebnisse sind immer zu überprüfen.

# Fazit

- Prüfen Sie, ob ihre Kommune an einem KI-Bürger:innen-Service arbeitet.
- Kooperieren Sie mit ihrer Kommune für einen kommunalen Chatbot.
- Es ist nicht nachhaltig, dass jede Dienststelle einer Kommune ein eigenes System beauftragt und auf Dauer betreut.
- Konzentrieren Sie sich stattdessen auf die passende Aufbereitung IHRER spezifischen Daten (Erschließungsdaten, Digitalisate, Publikationen, Handreichungen, Formulare, Satzungen, Gebührenordnungen, etc.).
- Sprechen Sie dafür ihre IT-Abt./ KI- oder Digitalisierungsbeauftragte an.
- Dieser Ansatz spart dem Archiv Zeit & Geld, minimiert Risiken und verbessert vor allem den allgemeinen Bürger\*innen-Service.
- Das Service-Angebot sollte barrierefrei und ohne Medienbrüche sein.



# Weitere Informationen

[KGSt-Guideline 2025](#)

[Archiv.theorie & praxis](#)

[KLA 2018](#)

[iRIGHTS.info](#)

[KI.NRW](#)

[KI-Landkarte Bund](#)

[KI-Campus](#)

[M. Weber \(KrA Reutlingen\)](#)

[R. Voges \(GStAPK\)](#)

[Eickhoff AWK25](#)

[Eickhoff BAT24](#)

Generative KI in Kommunalverwaltungen (Anwendungsfälle LLMs)

Themenheft zu KI im Archiv, 2025 (01)

Handr.: „Das digitale Gedächtnis nachhaltig aufbauen“ (noch ohne KI)

Für aktuelle Informationen rund um Digitalisierung, KI, Recht, etc.

KI-Projektlandkarte, Glossar, Publikationen, Sprechstunden

Bundesministerium für Forschung, Technologie und Raumfahrt

Niederschwellige Lernangebote und Webinare mit Zertifikaten

Github-Repository mit nachnutzbarem Code für einen ChatBot

Github-Repository mit nachnutzbarem Code für einen ChatBot

Präsentation Archivwiss. Kolloquium 2025: „Vom Experiment zur Praxis: Realistische Einsatzszenarien für LLMs in kleinen Archiven“

Tagungsband Berliner Archivtag 2024: „Nutzen von KI im Archiv: Experimentelle Pilotprojekte für nachhaltige Implementierungsstrategien“

# LWL-Archivamt für Westfalen



# Stadtarchiv Ulm

Stadt Ulm

ulm

Bluesky	<a href="https://lwl-archivamt.de">@lwl-archivamt.de</a>
Mastodon	<a href="https://lwlarchivamt@fedihum.org">@lwlarchivamt@fedihum.org</a>
Linkedin	<a href="https://lwl-archivamt">@lwl-archivamt</a>
Blog	<a href="https://blog.lwl-archivamt.de">@blog.lwl-archivamt.de</a>
Threads	<a href="https://lwlarchivamt">@lwlarchivamt</a>
Instagram	<a href="https://lwlarchivamt">@lwlarchivamt</a>
Facebook	<a href="https://LWLArchivamt">@LWLArchivamt</a>

Linkedin	<a href="https://stadtarchiv-ulm">@stadtarchiv-ulm</a>
Instagram	<a href="https://stadtarchiv.ulm">@stadtarchiv.ulm</a>
Facebook	<a href="https://stadtarchiv.ulm">@stadtarchiv.ulm</a>