

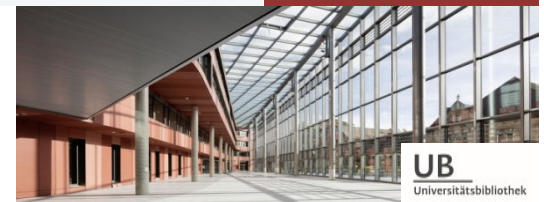
## Kulturerbe mittels KI zielgruppenspezifisch vermitteln

Das Projekt MATE - Maschinell erstellte Begleittexte für Kulturobjekte mittels generativer künstlicher Intelligenz

Dr. Diana Müller, Corinna Berg

Universitätsbibliothek Marburg

5. Juni 2024



# Projektdaten

**Projekt:** MATE - Maschinell erstellte Begleittexte für Kulturobjekte mittels generativer künstlicher Intelligenz

**Laufzeit:** 2023-2024 (6 Monate)


**Antragsteller\*in:** Universitätsbibliothek Marburg, Marburg Center for Digital Culture and Infrastructure

**Förderer:** Hessisches Ministerium für Wissenschaft und Kunst (HMWK)

corvey.ub.uni-marburg.de

## Marburg, Universitätsbibliothek, Ms. 49

Franciscus Galuani de lanua - Johannes Herolt - Theodulus u. a.



**öffnen mit:**  
[DFG-Viewer](#)  
[IIIF-Viewer](#)

**Download:**  
[Digitalisate \(PDF\)](#)  
[Metadaten \(TXT\)](#)

<b>Aufbewahrungsort</b>	Marburg
<b>Einrichtung</b>	Universitätsbibliothek Marburg
<b>Signatur</b>	Ms. 49
<b>Altsignaturen</b>	Mscr. 49 alt D 12 alt
<b>Autoren</b>	Franciscus, Galvanus (um 1348/52) [GND] Herolt, Johannes (-1468) [GND] Nicolaus, de Graetz (-1444) [GND] Gerson, Jean (1363-1429) [GND] Johannes, de Francfordia (1380-1440) [GND] Paul II., Papst (1417-1471) [GND] Theodulus, Italus [GND]
<b>Entstehungszeit</b>	1444-1470
<b>Entstehungsort</b>	Deutschland
<b>Sprache</b>	Latein Deutsch
<b>Katalogbeschreibung</b>	Heyne, Die mittelalterlichen Handschriften der Universitätsbibliothek Marburg, Wiesbaden 2002, S. 131ff.
<b>Medium</b>	Codex
<b>Format</b>	30 x 25 cm (Höhe x Breite)
<b>Beschreibstoff</b>	Papier
<b>Umfang</b>	275 Blätter

Suche

**Gesamter Bestand**

- Signatur
- Aufbewahrungsort
- Entstehungszeit
- Sprachen

**Corveykataloge Severus Campill**

- Ms. 497,1 (Drucke 1803)
- Ms. 497,2 (Handschriften 1803)

<https://corvey.ub.uni-marburg.de/handle/corvey/34>

# Projektziel

Erprobung generativer KI – hier: Large Language Models (LLMs) – für die Erstellung von Beschreibungstexten für Kulturgutobjekte

```
"name": "Franciscus Galvani de Ianua  
- Johannes Herolt - Theodulus u.  
a.",  
"authors": "Franciscus, Galvanus (um  
1348/52), Herolt, Johannes (-1468),  
Nicolaus, de Graetz (-1444), Gerson,  
Jean (1363-1429), Johannes, de  
Francfordia (1380-1440), Paul II.,  
Papst (1417-1471), Theodulus, Italus  
(None)",  
"location": "Universitätsbibliothek  
Marburg",  
"medium": "paper",  
"created": "1444-1470",  
"signature": "Ms. 49",  
"provenance": "Kloster Corvey and  
Campill, Severus",  
"title_alternative": [],  
"language": ["Latein", "Deutsch"],  
"format": "codex",  
"origin": [{"name": "Deutschland", "gnd  
Id": "4011882-4"}]
```



...  
Es handelt sich um einen Codex,  
ein gebundenes Manuskript, das  
zwischen 1444 und 1470 entstand.  
Der Codex enthält Texte der  
Autoren Franciscus Galvanus (um  
1348/52), Johannes Herolt (-1468),  
Nicolaus de Graetz (-1444), Jean  
Gerson (1363-1429), Johannes de  
Francfordia (1380-1440), Papst  
Paul II. (1417-1471) und Theodulus  
Italus.  
...

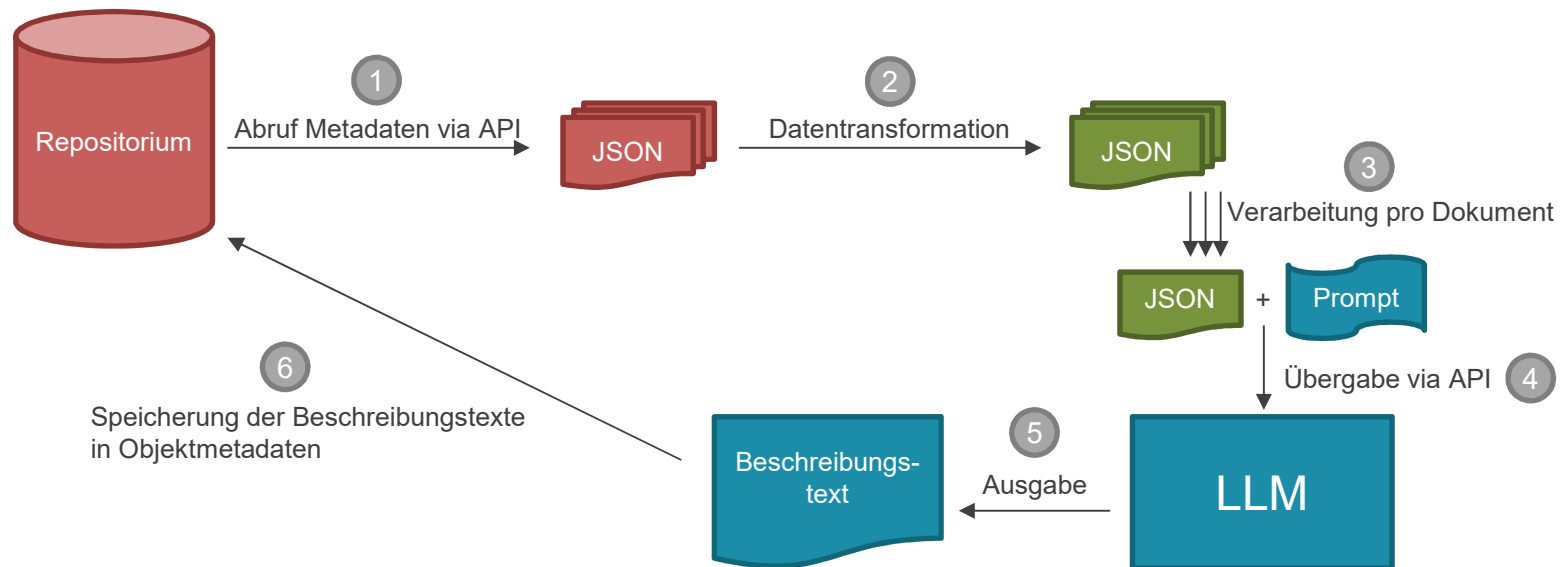


Versionen:

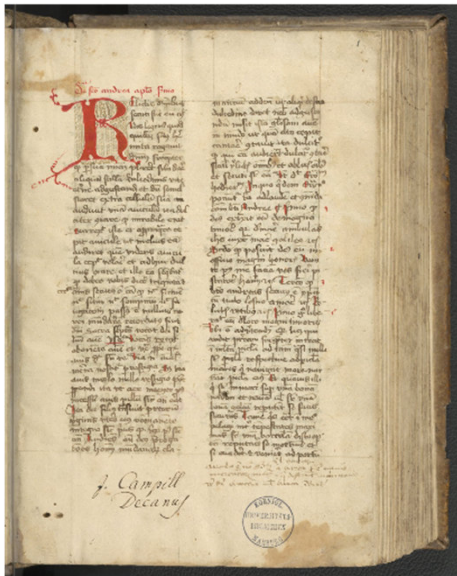
- Für Kinder
- Leichte Sprache
- Übersetzungen
- Audioformat
- ...

allgemeinverständliche Version

# Workflow Texterzeugung



## 2. Metadaten (Auszug)



```
...  
"name": "Franciscus Galvani de Ianua - Johannes  
Herolt - Theodulus u. a.",  
"authors": "Franciscus, Galvanus (um 1348/52),  
Herolt, Johannes (-1468),  
Nicolaus, de Graetz (-1444), Gerson, Jean (1363-  
1429), Johannes, de Francfordia (1380-1440), Paul  
II., Papst (1417-1471), Theodulus, Italus (None)",  
"location": "Universitätsbibliothek Marburg",  
"medium": "paper",  
"created": "1444-1470",  
"signature": "Ms. 49",  
"provenance": "Kloster Corvey and Campill,  
Severus",  
"title_alternative": [],  
"language": ["Latein", "Deutsch"],  
"format": "codex",  
"origin": [{"name": "Deutschland", "gndId": "4011882-  
4"}]  
...
```

## 3./4. Prompting (Beispiel GPT)

### **Prompt instruction:**

As an Historian write an interesting description in German language which is easy to understand for the medieval manuscript by using the metadata [...]

### **Parameter Temperature:**

0.7

### **Model:**

GPT 4

## 5. Ausgabe (Beispiel)

Das Manuskript "Franciscus Galvani de lanua - Johannes Herolt - Theodulus u. a." ist ein **bemerkenswertes** historisches Dokument aus der Mitte des 15. Jahrhunderts (1444-1470). Es wurde von einer Reihe von Autoren verfasst, darunter Franciscus Galvanus, Johannes Herolt, Nicolaus de Graetz, Jean Gerson, Johannes de Francfordia, Papst Paul II. und Theodulus Italus. Dieses Manuskript, das auf Papier geschrieben und **im Codex-Format präsentiert** wurde, befindet sich derzeit in der Universitätsbibliothek Marburg unter der Signatur "Ms. 49".

Das **beeindruckende** Dokument hat seinen Ursprung in Deutschland, **wie aus den Metadaten hervorgeht**. Es wurde sowohl auf Latein als auch auf Deutsch verfasst, was auf die weit verbreitete Verwendung dieser Sprachen in der Wissenschaft und Theologie dieser Zeit hinweist. Die Provenienz des Manuskripts weist auf das Kloster Corvey und **Campill, Severus** hin, was seine historische Bedeutung und seinen geistlichen Kontext unterstreicht.

Das Manuskript ist ein Zeugnis der mittelalterlichen Gelehrsamkeit und Theologie und bietet einen **tiefen Einblick** in die Gedanken und Diskurse jener Zeit. Es repräsentiert eine Epoche, in der das Schreiben und Kopieren von Manuskripten eine wichtige kulturelle Praxis war, die Wissen und Ideen über Generationen hinweg bewahrte und verbreitete. Es ist ein **unverzichtbares Zeugnis** für die Geschichte der Schriftlichkeit und Gelehrsamkeit im mittelalterlichen Europa.



## Fehlermuster (Beispiel): Autor schreibt Buch

- KI-Trainingsdaten: *Rainald Goetz* verfasste das Buch *Abfall für alle*.  
Stuckrad-Barre schrieb das Buch *Noch wach?*
- KI gibt aus: Thomas von Aquin [1225-1274] schrieb die Handschrift Ms. 75 [Entstehungszeit 15. Jh.].
- Richtig wäre: Die Handschrift Ms. 75 entstand im 15. Jahrhundert und enthält einen Text von Thomas von Aquin.

# Herausforderungen

(1) Mittelalterliche Buchkultur = **spezifische Wissensdomäne**, die in Trainingsdaten der LLMs unterrepräsentiert ist (Fehlermuster: Ebenendifferenzierung Werk – Handschrift bzw. Autor – Schreiber, Terminologie).

(2) „Objektbeschreibung“ = **spezifische Textsorte**, die in Trainingsdaten der LLMs unterrepräsentiert ist (Herausforderung: Relevanzgewichtung der Informationen, sprachliche Qualität etc.).

# Texterzeugung mit freien Modellen

- Ausgewählt wurden u.a. folgende Modelle: LLaMA2 (Meta/Microsoft) und Mixtral (Mistral AI)
- Ergebnisse im Vergleich zu GPT 4 (OpenAI):
  - „sachlichere“ Texte (weniger wertende Adjektive), zugleich sprachlich im Deutschen jedoch insgesamt schwächer
  - Halluzination besser zu kontrollieren (Metadaten + Parameter wirksam)
  - Random Seed erlaubt deterministische Ergebnisse

## Stand und Perspektiven

- Vielversprechender Ansatz, um auf automatisierte Weise bei großen Mengen Kulturgutdigitalisaten zusätzliche Präsentationsformen für ein diverses Publikum anzubieten
- Weitere Optimierungsmöglichkeiten:
  - Modell-Feintuning zur Anpassung an die Textgattung „Objektbeschreibung“
  - Nutzung von Retrieval Augmented Generation (RAG) für die Bereitstellung erweiterter Kontextinformationen, insb. auch aus externen Datenquellen (Wissensgraphen etc.)
- Valide allgemeinverständliche Beschreibung ist Voraussetzung, um weitere zielgruppenspezifische Versionen zu erproben

# Kontakt

Dr. Diana Müller  
Universitätsbibliothek Marburg  
[diana.mueller@ub.uni-marburg.de](mailto:diana.mueller@ub.uni-marburg.de)

Corinna Berg  
Marburg Center for Digital Culture and Infrastructure  
[corinna.berg@staff.uni-marburg.de](mailto:corinna.berg@staff.uni-marburg.de)