UNIVERSITY OF MANNHEIM

Representing the Past—
Building and Benchmarking Historical LLMs

Team Project Autumn/Winter 2025

Chair of Data Science in the Economic and Social Sciences

Supervisor: Georg Ahnert

Representing the Past

- **Large language models (LLMs)** are a promising new tool for historical research
  - e.g., Varnum et al. (2024): *Large Language Models based on historical text could offer informative tools for behavioral science*

- LLMs could be used to:
  - Gain a **better understanding about past societies**—e.g., How would people in 1927 have responded to *AI will profoundly change my life in the next 3–5 years* (Stanford AI Index, 2024)?
  - Simulate **counterfactuals**—e.g., *What if* Roosevelt would have taken part in the Potsdam Conference in 1945?
  - Augment **history education**—"talk" to historical figures

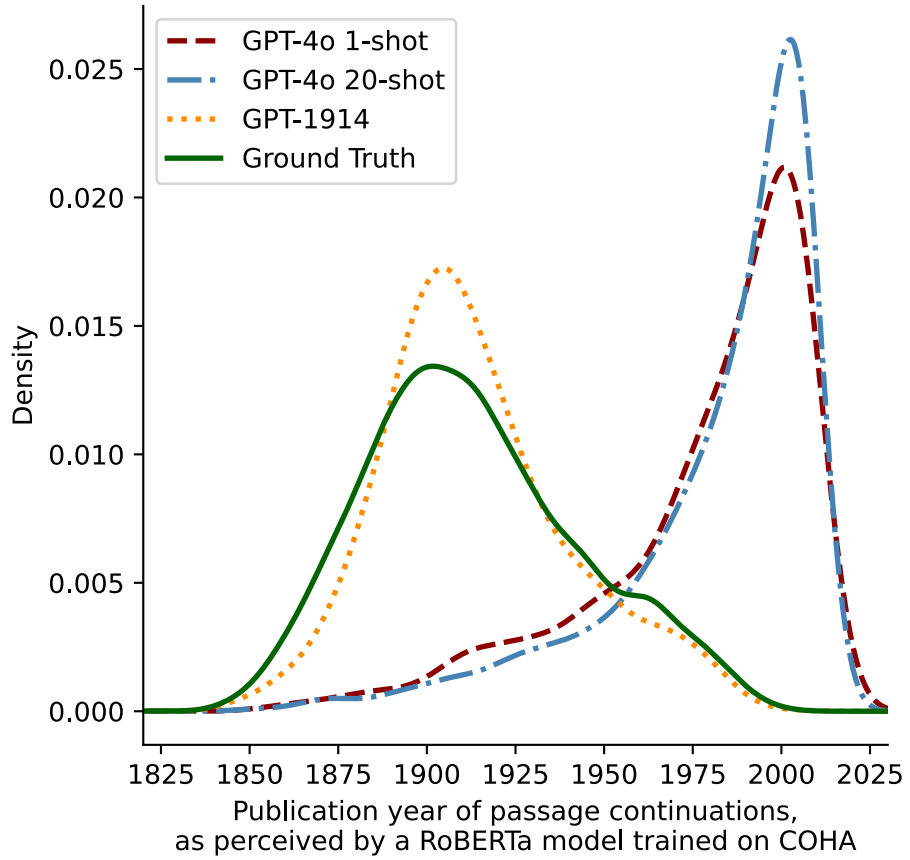
https://www.moma.org/collection/works/88251

Historical LLMs

UNIVERSITY OF MANNHEIM

- Prompting off-the-shelf LLMs does not lead to **historically authentic responses**
  - e.g., Underwood et al. (2025): *Can Language Models Represent the Past without Anachronism?*

- Challenges in building historical LLMs:
  - Off-the-shelf LLMs are mainly trained on **recent text data**
  - **Validation criteria/benchmarks** need to be established
  - Easy to process **historical texts are sparse**
  - Pre-training (& fine-tuning) LLMs is **resource intensive**



Team Project: Representing the Past—Building and Benchmarking Historical LLMs

UNIVERSITY OF MANNHEIM

- Conceptualize benchmarks—What should historical LLMs (*not*) know about?
- Identify & clean historical texts as training data
- Compare design approaches: prompting, pre-training, fine-tuning

- Prerequisites
  - **Strong** Python programming skills
  - Previous experience with LLMs

Team Project: Representing the Past— Building and Benchmarking Historical LLMs

UNIVERSITY OF MANNHEIM

- **Language:** English
- **Duration:** 6 months
- **Participants:** 3–5
- **Prerequisites:** *Strong* Python skills, previous experience with LLMs
- **Process:**
  - Data exploration, coding, coordination with the team
  - Regular meetings with the supervisor to present and discuss your progress
  - 2 presentations in front of the chair: intermediate results, final results
  - Written report to be submitted at the end
- **Contact:** Georg Ahnert – ahnert@uni-mannheim.de