AI (also) for Humanities Scholars:

The CAS NLP at UniBe

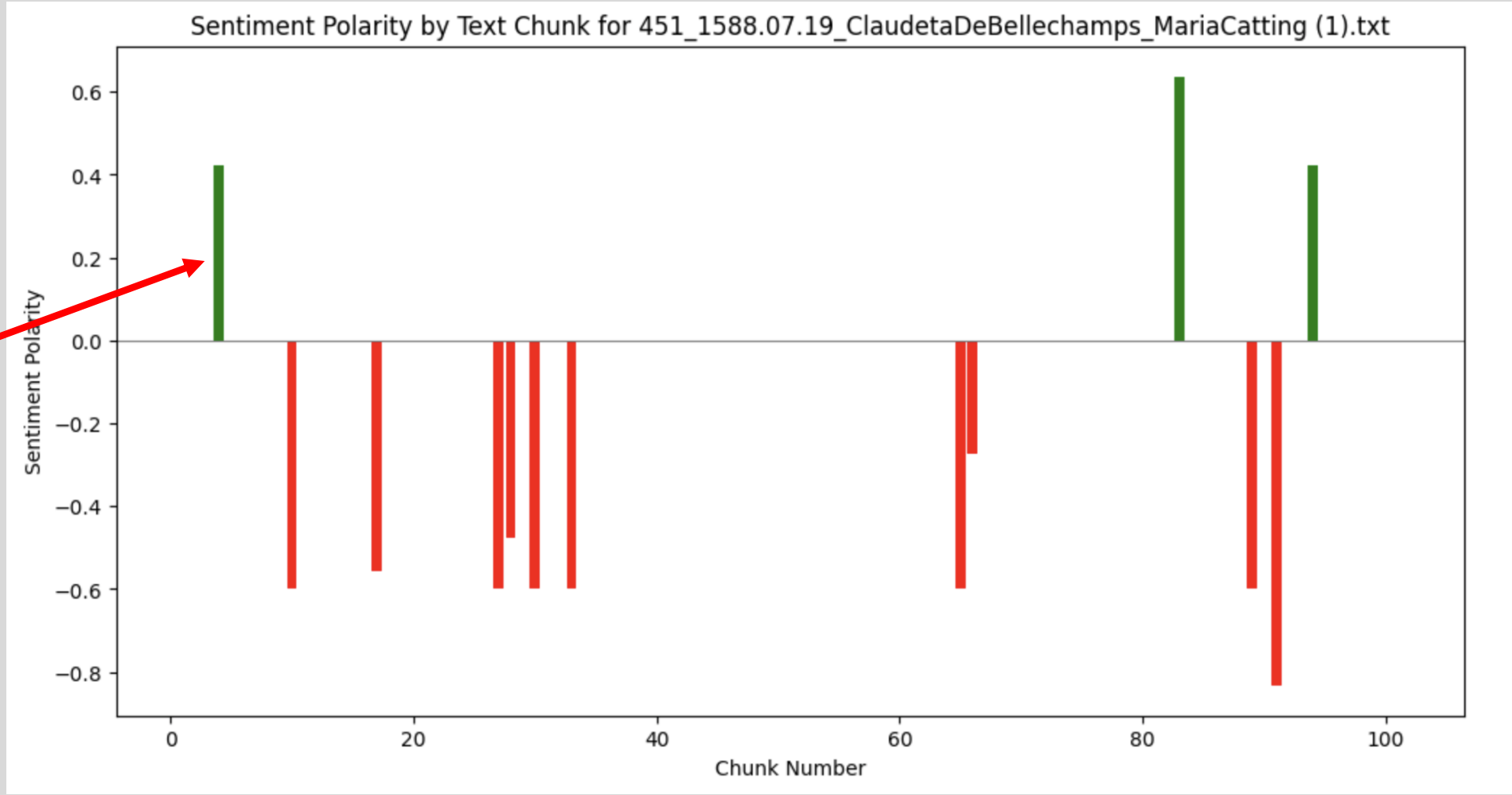
Witch Hunt in Switzerland

- Switzerland in the center of a «witch-hunt-zone»
- Approx. 10’000 witch trials in Switzerland
- 400 witch trial pages available for Bern

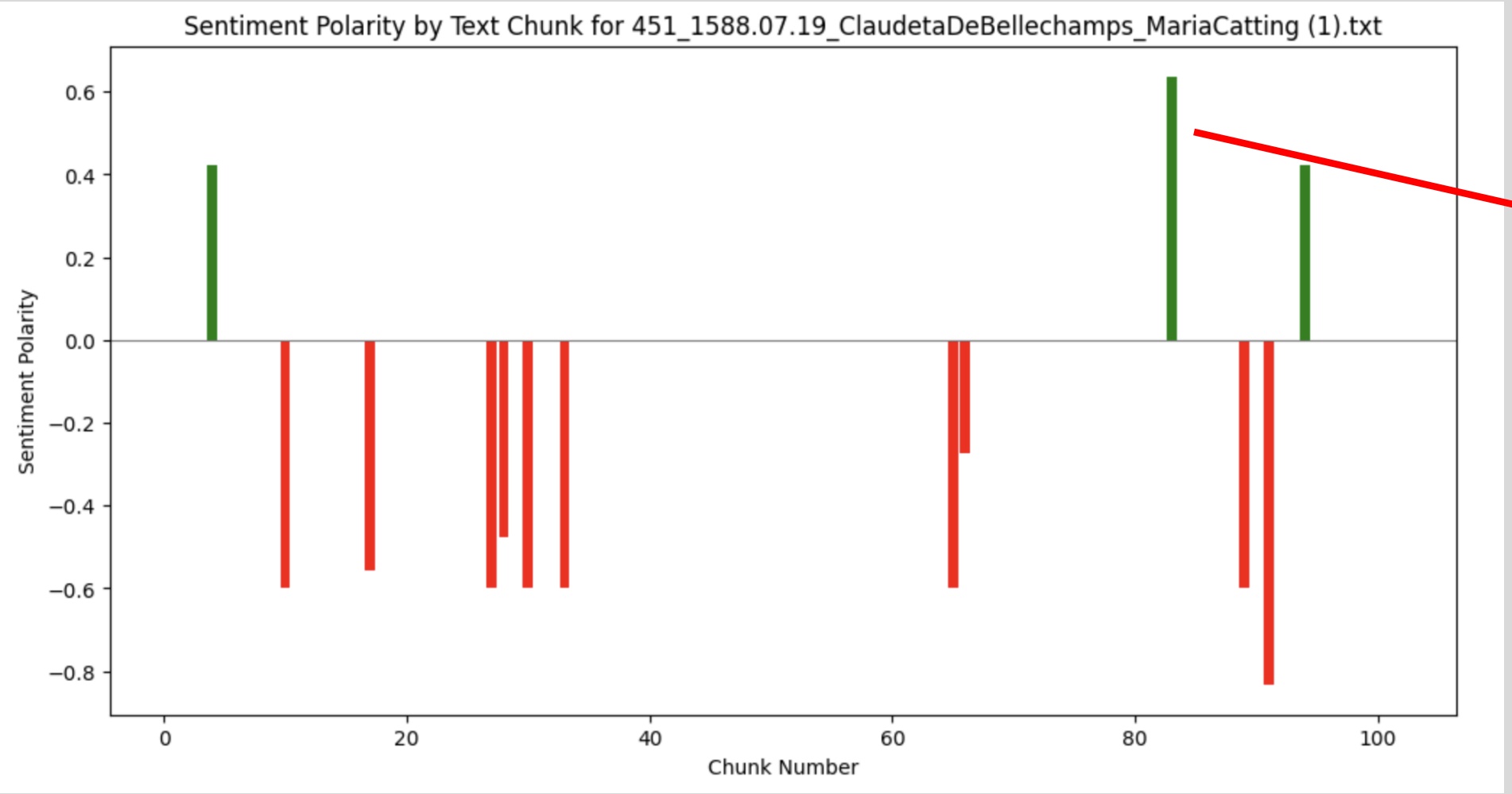
Sentiment Analysis

Emotions in Witch Trial Papers

[…] frommenn, fürnemmen fürsichtigen Ersamēn, vnnd wÿsenn Herren, Niclauss zur khindens ~~beid~~ des Rhats, Bilger Steineggers […]

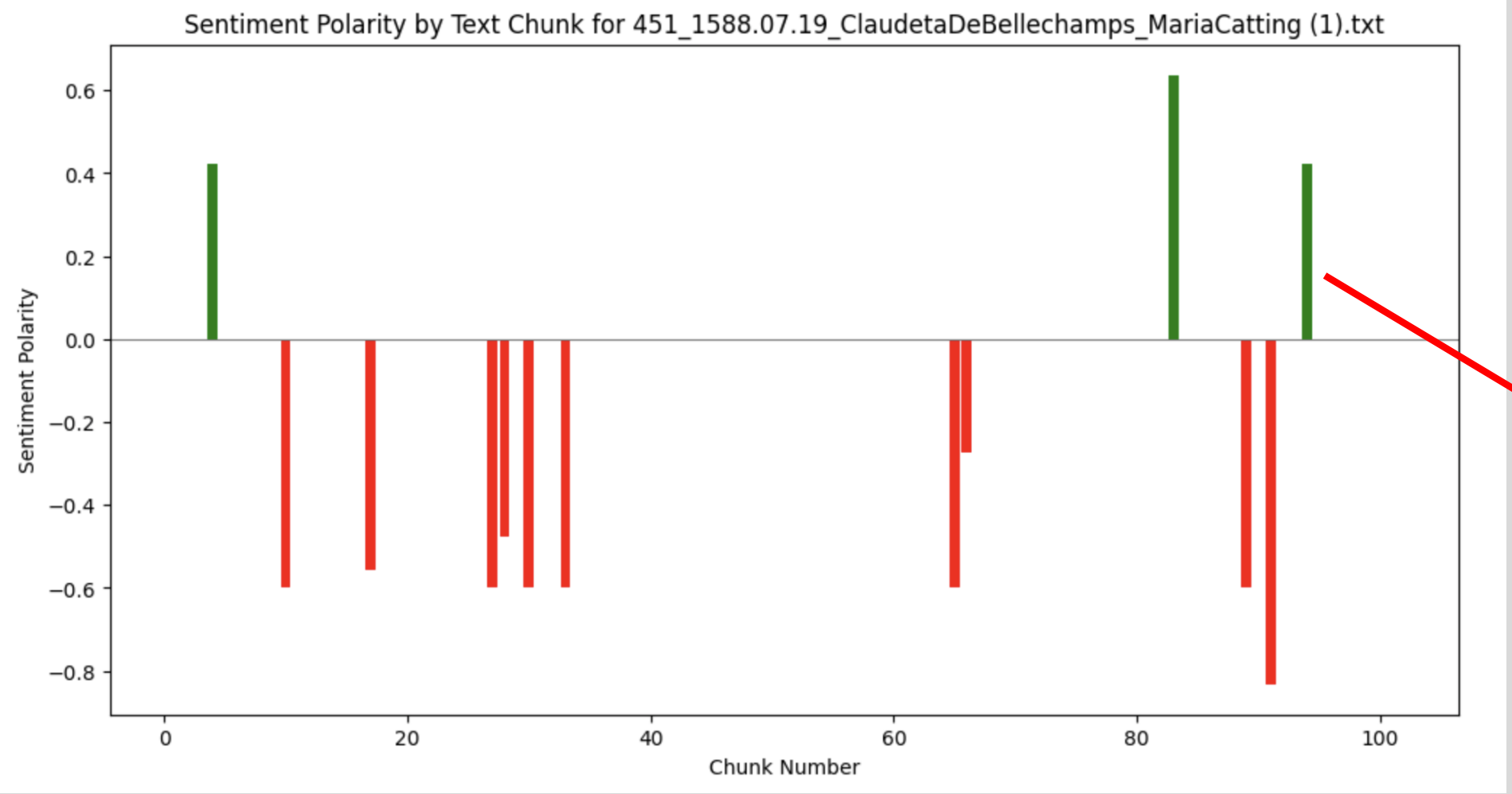


Emotions in Witch Trial Papers



[…] Endtlich hatt sÿ ouch bekhendt, vnnd verjächen wie das sÿ vff ein zÿtt, zu der vorgemeltenn Claudeta de Bellechamps, gan Serniaulx inn ir Huss gangenn, vnnd ira Salb abgeforderet, (ein armes kindt oder Meittlin, So sÿ vonn Armutt wägenn, zu ira vffgenommen, vnnd erzogenn, wölches anfüssen, vbell erforenn gwässen, vnnd grossenn Schmertzen daran gelittenn) darmit zesalbenn, vnnd zeheillenn […]

Emotions in Witch Trial Papers



[…] vnnd alls sÿ söllichs gesächenn, sÿe sÿ vbell darab erschrockenn, vnnd hiemit die vbrige Salb (darmitt niemandts wÿtter dardurch geschediget wurde) glÿch angentz inn den See geworffen etc. wöllichs ira aber von hertzenn, vnnd Rüwlich leÿdt sÿe etc. […]

How? Why? What? NLP!

What is it?

Natural **L**anguage **P**rocessing
Field at the intersection between computer science and linguistics
uses machine learning to process and interpret text and data

What can it do?

Analyse Text at <u>Scale</u>:
Big Data**:** massive textual datasets (archives, literature, social media etc.)
Uncover Broad Trends**:** (patterns, shifts, and large-scale phenomena that are difficult manually)

<u>Enhance</u> Qualitative Insights:
Structure & "Objectivity"**:** Extract key themes, entities, sentiment, and relationships systematically.

<u>Unlock</u> New Research Questions:
Data-Driven Discovery: Generate novel hypotheses and explore previously intractable questions.

NLP: Speaker identification

Challenge: Speaker identification from text transcripts

Humans: context, word choice, stylistic recognition, dialogue flow…
→ intuitive, but subjective and tedious

ML models: statistical patterns: word frequencies, n-grams, text embeddings…
→ fast, consistent, but inscrutable

Case Study: Dimension 20 "Fantasy High" (Fig & Fabian)

Trained on 2 Mio Tokens of podcast transcript
Testing different models:

Model	Accuracy
Logistic Regression	73%
LSTM	74%
DistilBERT (Fine-tuned)	82%

NLP: Fragment reconstruction

Challenge: Ancient Greek inscriptions are often damaged, fragmented, and lack clear dating or origin. Traditional manual methods are slow and subjective.

NLP approach: Ithaca

Trained on the largest dataset of digitized ancient Greek inscriptions.

Restores damaged text fragments with 62% accuracy (improving historian accuracy to **72%** when used collaboratively!).

Attributes inscriptions to their original location (71% accuracy) and dates them within 30 years.

Redates key historical texts, challenging established historical interpretations.

Assael et al., 'Restoring and Attributing Ancient Texts Using Deep Neural Networks' (2022), 10.1038/s41586-022-04448-z

How? Why? What? NLP!

- https://www.unibe.ch/continuing_education_programs/cas_natural_language_processing/index_eng.html

We are looking forward to seeing you there!