Universität Bremen

ZeMKI Zentrum für Medien-, Kommunikations- und Informationsforschung

{klaus} for content analysis

1.7.2025, GöAID

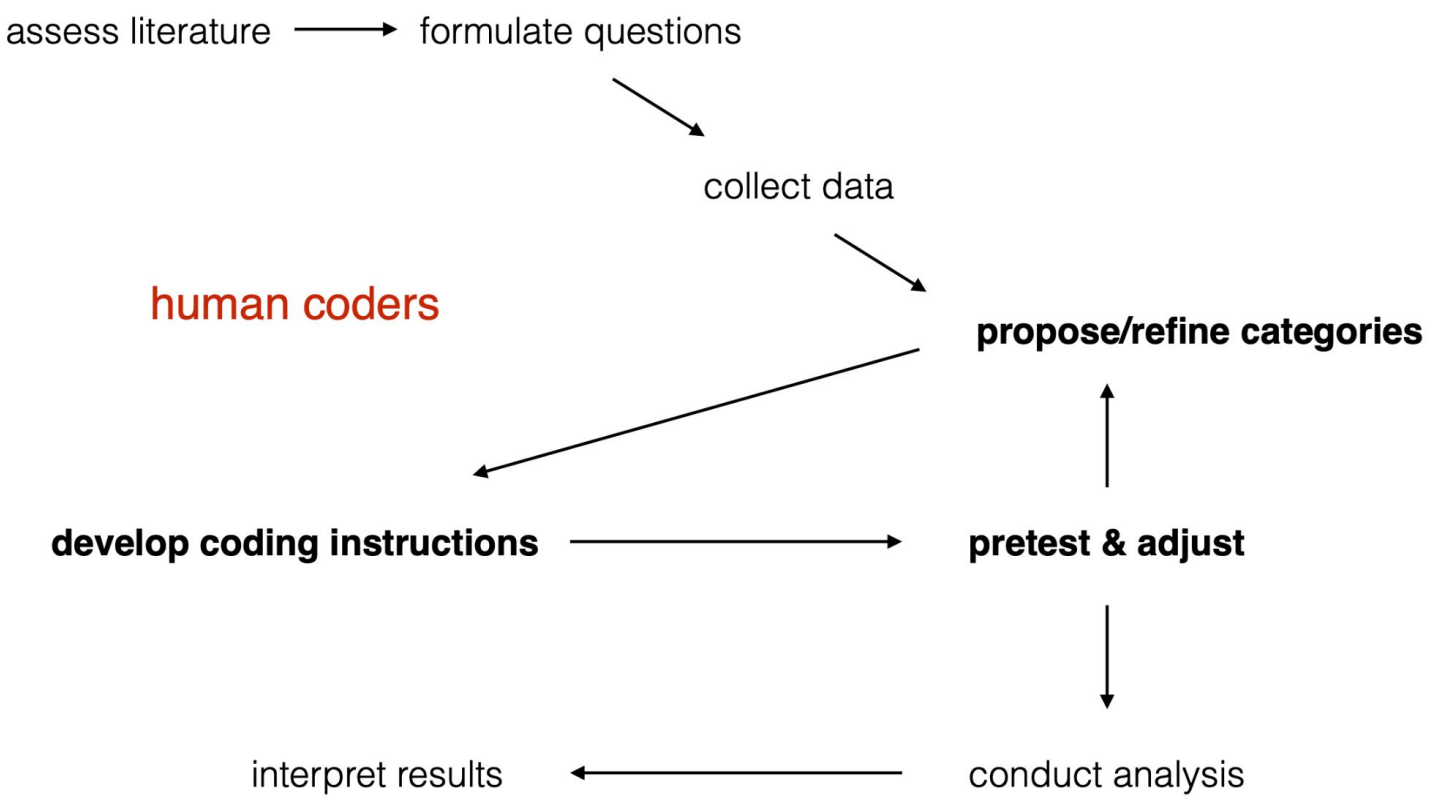
Cornelius Puschmann, U Bremen

What is content analysis?

“Content analysis is a research technique for making replicable and valid inferences from texts or other meaningful materials to the context of their use.”

(Krippendorff, 2004)

The CA workflow



Fairness and Election Coverage. Presidential Race, 1996

v1. Story identification number ____

v2. Story day (month and date) ____

v3. Newspaper name and circulation rank ____

General Story Characteristics

v.4. Story prominence (FP=2; SP1=1; Inside=0)

v5. Story origin

1=newspaper's own reporter 2=newspaper's state bureau
3=newspaper's D.C. bureau 4=AP ____
5=other bureau/newspapers

v6. AP unique (if v5=4)

0=duplicate story 1=unique story, blank=no AP ____

v7. Partisan side sourcing effort

1=only Clinton or Dole 2=both Clinton and Dole
3=only Perot 4=Perot and Clinton or Dole ____
5=Perot and Clinton and Dole

v8. Primary story source

1=debate 2=speech/rally
3=document 4=poll ____
5=interview 6=other media interview
7=ad 8=other

(Riffe et al., 2005, p.138)

Sample coding categories

id	body	author	author_fullnam	timestamp	type	url	image_url
DJjiV1Usu2u	I was in London today to meet Foreign Ministers of the Weimar+ format – France, Germany, Italy, Poland, Spain and hosted by the United Kingdom. Together, we stressed that Russia must agree a 30-day ceasefire with Ukraine immediately to pave the way for talks on lasting peace. Putin should stop playing games and get serious about engaging in peace talks.	kajakallas	Kaja Kallas	2025-05-12 13:3	photo	https://www.insta	https://scontent-(
DJb9qe-MUWq	It was important to be in Lviv today, to mark Europe Day with Ukraine and its people. Our presence here with Foreign Ministers from across Europe underlines our unity and unwavering solidarity with Ukraine. We came with very important deliverables: 🔹 We made €1 billion available for Ukraine's defence industry. This funding will directly support Ukrainian defence companies and secure additional military aid over the coming months. On Monday, we will organise a major forum in Brussels to link up Ukraine's defence industry with ours. 🔹 We endorsed the Special Tribunal for the Crime of Aggression against Ukraine That means that nobody can be left unpunished for the crimes committed, not even the leaders who have decided to send all these soldiers here to commit war crimes and all the atrocities that we have recognised here. Every inch of Russia's war has been documented. It leaves no room for doubt. And no room for impunity.	kajakallas	Kaja Kallas	2025-05-09 14:5	photo	https://www.insta	https://scontent-(

O	P	Q	R	S
Presence of KK	**Presence of Others**	**Visible Interaction with Other People**	**National/European Symbolism**	**Emotional Appeal**
Yes	Yes	Yes	European	Determin…
Yes	Yes	Yes	European	Determin…
Yes	Yes	Yes	European	Empathy
Yes	Yes	Yes	European	Hope/Opti…

A light-weight R package for LLM-based content analysis

Modern large language models (LLMs) offer considerable advantages for standardized content analysis. `klaus` facilitates use of both proprietary and open source LLMs by offering a simple interface through which to serve data and apply categorization. Presently the package supports the proprietary APIs of OpenAI, Anthropic and Google, as well as for local use via ollama (through tidyllm). In addition, for academic research, it is also possible to use the non-commercial ChatAI API service provided by GWDG or Blablador provided by the Forschungszentrum Jülich.

klaus is largely a convenience wrapper for the SAIA API and several commercial APIs via {tidyllm}

chatai	Interact with the Chat AI Completion Endpoint
chatai_models	List Available Chat AI Models
code_content	Code Text Content Using an API and a Codebook
parse_codebook	Parse a Codebook into JSON

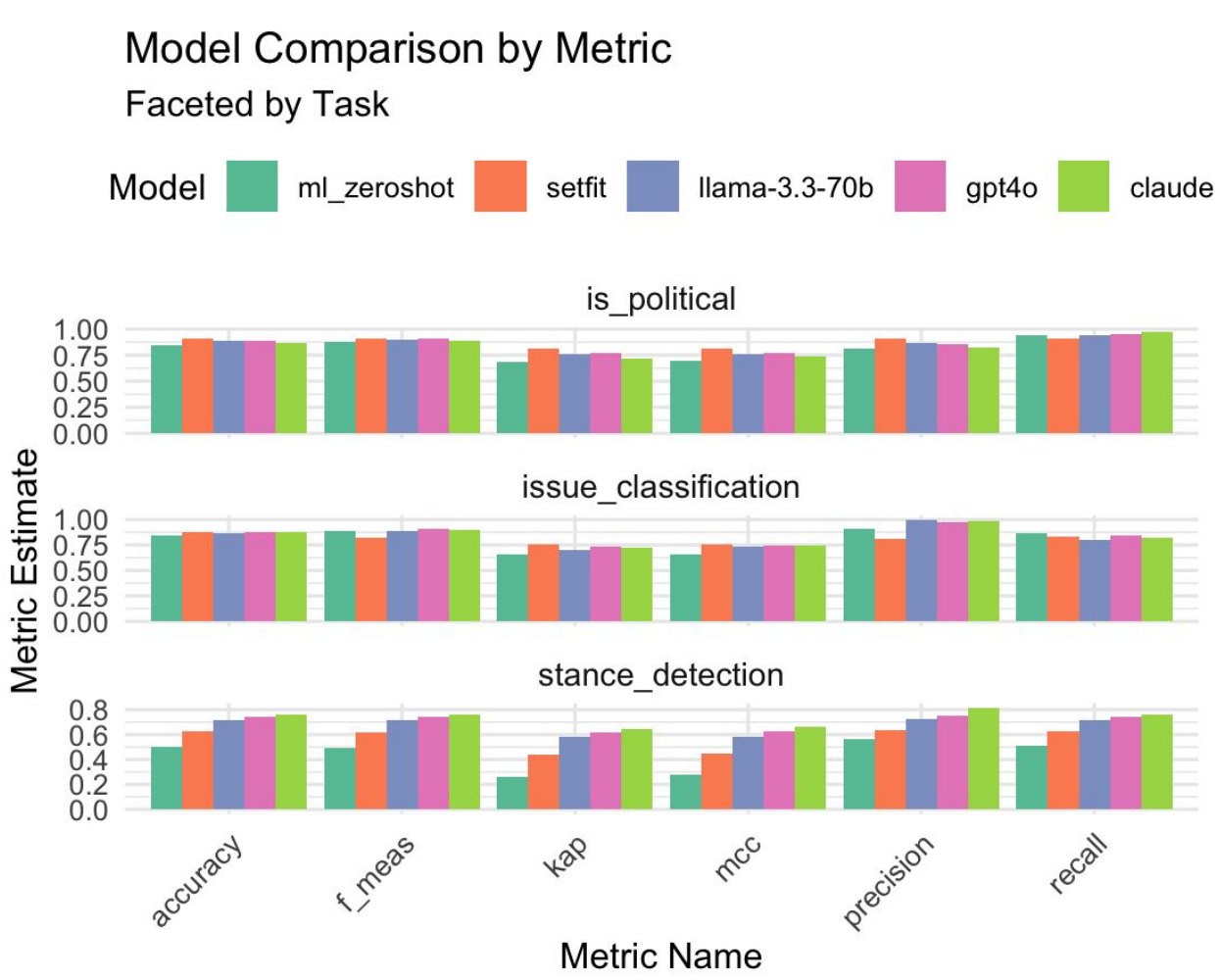
codebook:

category *(character)*	label *(character)*	instructions *(character)*
sentiment	positive	Code this if the sentiment of the tweet is positive
sentiment	negative	Code this if the sentiment of the tweet is negative
sentiment	neutral	Code this if the sentiment of the tweet is neutral

```
data_to_code <- readr::read_csv("data_sentiment.csv")

general_instructions <- "You are a highly accurate and consistent text
classification model that specializes in analyzing English-language Twitter posts.
Your task is to determine the sentiment of the tweet reproduced below. You must
strictly follow the classification rules without deviation. Do not return any
additional information outside the classification scheme. Use JSON."

formatting_instructions <-  "Always return a single JSON object with the category
name as the key for each coded text. The value should be an object containing a
'label' key and a single value among multiple options. Each JSON object should have
the following structure:"

codebook <- readr::read_csv("codebook_sentiment.csv")

coded_data_chatai <- code_content(data_to_code, general_instructions,
formatting_instructions, codebook, provider = "chatai", model =
"llama-3.3-70b-instruct")
```

```
Coding data with claude (4 rows)
Iterating over content...
Coding data with claude / claude-3-7-sonnet-20250219...
Processed text 1 of 4
Coding data with claude / claude-3-7-sonnet-20250219...
Processed text 2 of 4
Coding data with claude / claude-3-7-sonnet-20250219...
Processed text 3 of 4
Coding data with claude / claude-3-7-sonnet-20250219...
Processed text 4 of 4
Parsing JSON responses (4 rows).
Done. Joined results for 4 parsed rows.
```

	user	text	category	label
1	peter	It is terrible what is happening to this country	sentiment	negative
2	john	I love Donald Trump	sentiment	positive
3	mary	The stock market has falled dramatically because of t...	sentiment	negative
4	mike	The situation has been calm following the recent midt...	sentiment	neutral


Model Comparison by Metric
Faceted by Task
Model
ml_zeroshot
setfit
llama-3.3-70b
gpt4o
claude
is_political
issue_classification
stance_detection
Metric Estimate
1.00
0.75
0.50
0.25
0.00
0.8
0.6
0.4
0.2
0.0
accuracy
f_meas
kap
mcc
precision
recall
Metric Name

ZeMKI Universität Bremen

thanks for listening -- any questions?