Lecture 3: Generative AI for Content Analysis

Elizaveta Kuznetsova, elizaveta.kuznetsova@weizenbaum-institut.de

Freie Universität Berlin

weizenbaum institut

Lecture Plan

- ➔ How can gen AI assist in content labeling? What is possible
- ➔ How does it work?
- ➔ Ethical Considerations
- ➔ Basic principles of working with LLM-assisted content labeling
- ➔ Demonstration. Specific examples
- ➔ Q&A

Traditional content analysis vs. LLM-assisted

Traditional content analysis, "*a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use*" (Krippendorf, 2004)

Traditional Content Analysis	**LLM-assisted Content Analysis**
Rigourous, theory-driven method, widely accepted in Communication, Political Science, and Sociology	Offers new possibilities but still a developing method with no well-established and accepted methodological pipeline.
Time-consuming, requires at least two coders, costly	Faster, though human involvement is still necessary, cost depends on the model
Only limited data processing capabilities	Almost unlimited data processing capabilities, scalable
Better reliability and replicability, though time-consuming	Replicability might be difficult, though not impossible. Working with open-source models is preferable

How can LLMs help?

It can assist in qualitative and quantitative content analysis to identify patterns, themes and biases, including discourse analysis and frame analysis by

➔ assisting in summarizing

➔ generating ideas and code books

➔ deductive coding

➔ data extraction

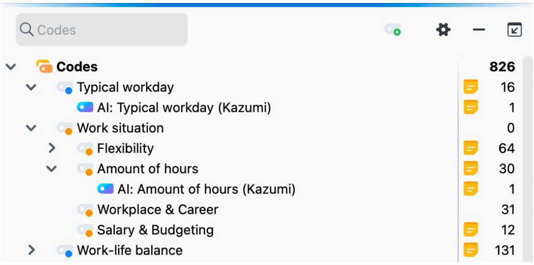
➔ data classification/labeling (topic modeling, sentiment analysis)

Assistance in summarizing & generating codes

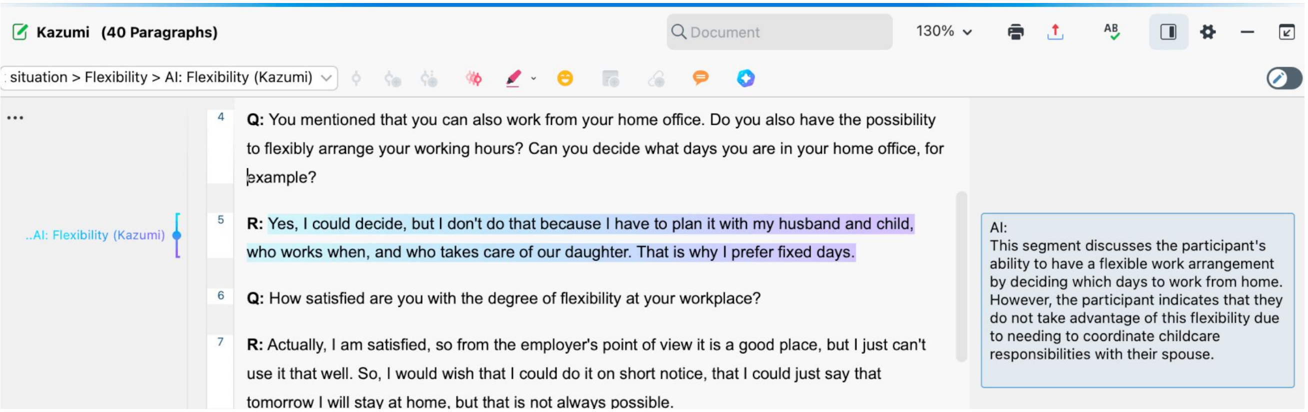
It is possible to use chatbots for summaries and generating codes. If available, content analysis softwares have integrated **AI assist** functions. They can:

- ➔ summarize uncoded data
- ➔ summarize coded data
- ➔ suggest codes (requires clear code definitions)
- ➔ automatic coding (if you have codes)
- ➔ data extraction

Note! All functions available in Beta. It uses GDPR-compliant models (combination of Claude & Gemini), though not clear which ones and when.



Codes	826
Typical workday	16
AI: Typical workday (Kazumi)	1
Work situation	0
Flexibility	64
Amount of hours	30
AI: Amount of hours (Kazumi)	1
Workplace & Career	31
Salary & Budgeting	12
Work-life balance	131


Kazumi (40 Paragraphs)
Document
130%
situation > Flexibility > AI: Flexibility (Kazumi)
...AI: Flexibility (Kazumi)

4 **Q:** You mentioned that you can also work from your home office. Do you also have the possibility to flexibly arrange your working hours? Can you decide what days you are in your home office, for example?

5 **R:** Yes, I could decide, but I don't do that because I have to plan it with my husband and child, who works when, and who takes care of our daughter. That is why I prefer fixed days.

6 **Q:** How satisfied are you with the degree of flexibility at your workplace?

7 **R:** Actually, I am satisfied, so from the employer's point of view it is a good place, but I just can't use it that well. So, I would wish that I could do it on short notice, that I could just say that tomorrow I will stay at home, but that is not always possible.

AI:
This segment discusses the participant's ability to have a flexible work arrangement by deciding which days to work from home. However, the participant indicates that they do not take advantage of this flexibility due to needing to coordinate childcare responsibilities with their spouse.

Using LLMs directly or with Cloud API

Running LLMs locally

Examples: Llama, Mistral, Phi

Advantages:

- ★ Full control over data and model performance
- ★ No usage fees after setup
- ★ Privacy – your data never leaves your machine
- ★ Customizable – can fine-tune or modify for your specific task

Disadvantages:

- ➔ Requires setup and technical skills
- ➔ Needs powerful hardware (RAM, GPU, memory)
- ➔ LLMs are smaller and less accurate than GPT4o, worse multilingual performance

Cloud API

Examples: OpenAI, Anthropic, Together AI

Advantages:

- ★ Ready to be used
- ★ Access to leading models
- ★ Better performance, scalable
- ★ Constantly updated, models improve over time

Disadvantages:

- ➔ Costs money per usage, can get expensive at scale
- ➔ Requires internet connection
- ➔ Data privacy concerns, data goes to external servers
- ➔ Limited customization, harder to fine-tune

Freie Universität Berlin

weizenbaum institut

LLM-assisted content analysis pipeline

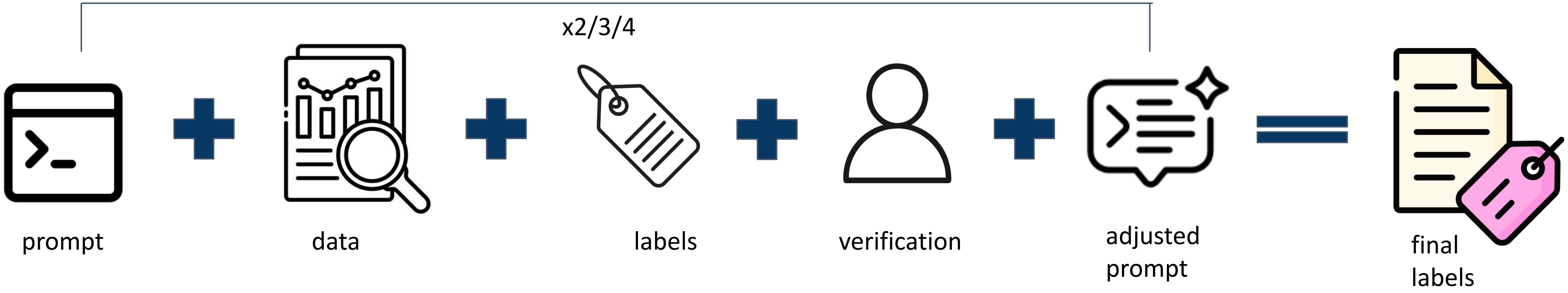
Option 1. Zero-shot prompting

prompt + data = final labels

- prompt contains no examples just a task
- possible & quick for simple tasks
- often unreliable for complex tasks

LLM-assisted content analysis pipeline

Option 2. Iterative prompt refinement



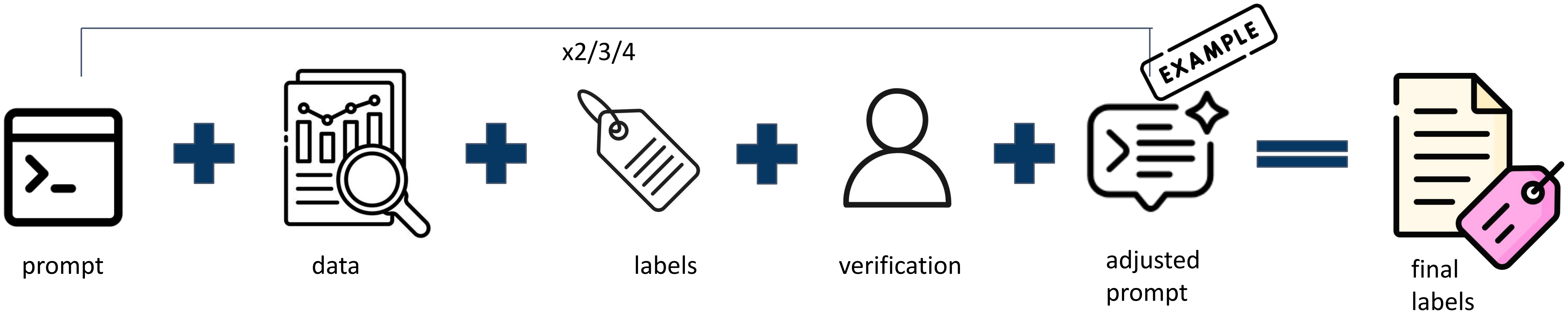
x2/3/4

prompt + data + labels + verification + adjusted prompt = final labels

Start with a zero-shot prompt, **review the results**, then **revise the prompt** based on observed weaknesses. This may include rephrasing, clarifying task instructions, or adjusting category definitions — **but still no examples are used**

LLM-assisted content analysis pipeline

Option 3. Few-shot prompting with refinement



After prompt tuning, you now **add examples** to guide the model's understanding (few-shot). You still may iteratively verify and tweak prompts or instructions.

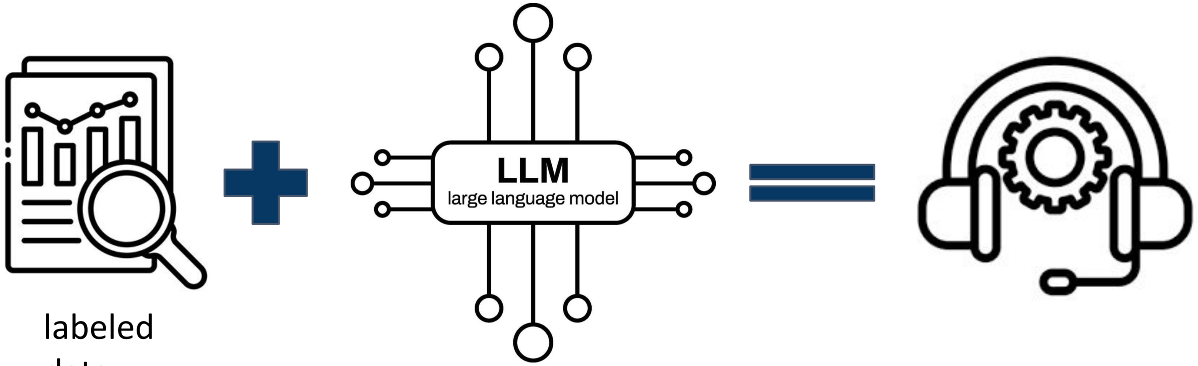
LLM-assisted content analysis pipeline

Fine-tuning an LLM for specific tasks

- ➔ Ensures **consistent labeling** across large volumes of text
- ➔ Adapts the model to **your categories, definitions, and style**
- ➔ Minimizes the need for repeated prompt writing or manual coding.

Only use if:

- ❖ You have a clear labeling scheme and a **large set of labeled examples**
- ❖ You want to **automate coding** while maintaining reliability
- ❖ You're doing **high-volume or domain-specific analysis**



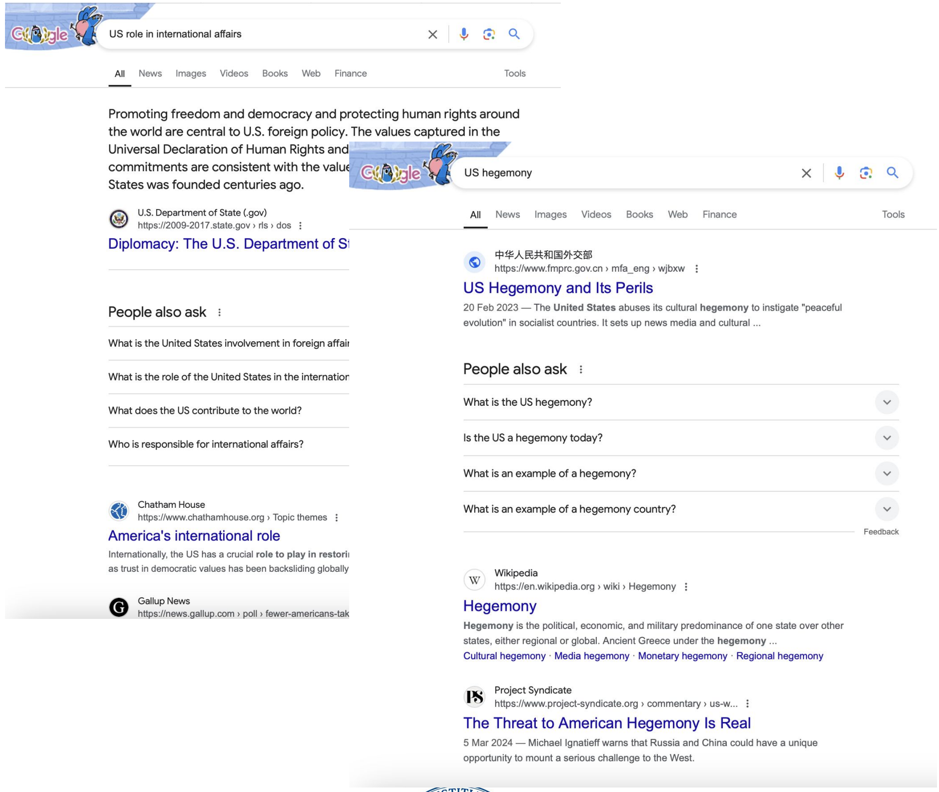
labeled data

LLM-assistant

Example Project

Using LLMs for analysing Russian propaganda on search engines

- ➔ **Algorithm auditing** methodology, a process of investigating functionality and impact of decision-making algorithms (Mittelstadt, 2016).
- ➔ **Virtual agent-based audits** (Haim et al., 2017; Makhortykh et al., 2020; Ulloa et al., 2022).
- ➔ **6 regions** (USA, India, Qatar, Brazil, Poland, Germany).
- ➔ **9 languages** (English, Spanish, Portuguese, German, Hindi, Polish, Arabic, Russian, and Ukrainian).
- ➔ **5 search engines** (Google, Bing, Baidu, Yandex, DDG)
- ➔ **360** unique queries



US role in international affairs

Promoting freedom and democracy and protecting human rights around the world are central to U.S. foreign policy. The values captured in the Universal Declaration of Human Rights and commitments are consistent with the value States was founded centuries ago.

U.S. Department of State (.gov)
https://2009-2017.state.gov › rls › dos
Diplomacy: The U.S. Department of S

People also ask

What is the United States involvement in foreign affai

What is the role of the United States in the internation

What does the US contribute to the world?

Who is responsible for international affairs?

Chatham House
https://www.chathamhouse.org › Topic themes
America's international role
Internationally, the US has a crucial **role to play in restori** as trust in democratic values has been backsliding globally

Gallup News
https://news.gallup.com › poll › fewer-americans-tak

US hegemony

中华人民共和国外交部
https://www.fmprc.gov.cn › mfa_eng › wjbxw
US Hegemony and Its Perils
20 Feb 2023 — The **United States** abuses its cultural **hegemony** to instigate "peaceful evolution" in socialist countries. It sets up news media and cultural ...

People also ask

What is the US hegemony?

Is the US a hegemony today?

What is an example of a hegemony?

What is an example of a hegemony country?

Feedback

Wikipedia
https://en.wikipedia.org › wiki › Hegemony
Hegemony
Hegemony is the political, economic, and military predominance of one state over other states, either regional or global. Ancient Greece under the **hegemony** ...
Cultural hegemony · Media hegemony · Monetary hegemony · Regional hegemony

Project Syndicate
https://www.project-syndicate.org › commentary › us-w...
The Threat to American Hegemony Is Real
5 Mar 2024 — Michael Ignatieff warns that Russia and China could have a unique opportunity to mount a serious challenge to the West.

Freie Universität Berlin

weizenbaum institut

Example Project. Rus Propaganda Statements

“США гегемонія”

“США гегемония”,

“hegemonia USA”,

“hegemonia dos EUA”,

“US-Hegemonie”,

“hegemonía de EEUU”,

“US hegemony”,

“الهيمنة الأمريكية”,

“अमेरिका का दबदबा”

США роль у міжнародних відносинах

США роль в международных отношениях

rola USA w sprawach międzynarodowych

papel dos EUA nas relações internacionais

rolle der USA in internationalen Beziehungen

papel de EEUU en asuntos internacionales

US role in international affairs

دور الولايات المتحدة في الشؤون الدولية”,

“अमेरिका की अंतर्राष्ट्रीय मामलों में भूमिका”

західне суспільство трансгендер

западное общество трансгендер

zachodnie społeczeństwo transpłciowy

sociedade ocidental transgênero

westliche Gesellschaft Transgender

sociedad occidental transgénero

western society transgender

المتحولين جنسيا المجتمع الغربي”,

“पश्चिम समाज हिजरा”

ЄС США васал

ЕС США вассал

UE wasal USA

União Europeia subordinado EUA

EU US Vasall

UE EEUU vasallo

EU US vassal

“الاتحاد الأوروبي تابع للولايات المتحدة”,

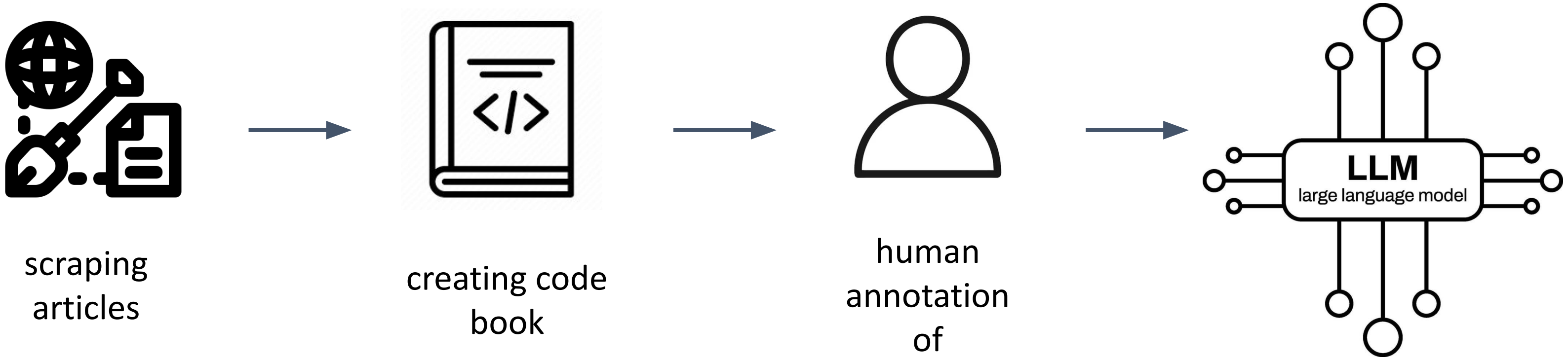
“यूरोपियन यूनियन अमेरिका जागीरदार”

Example Project. Analysis challenge

Task: Label all content to identify, whether it is propaganda or not

↔

Challenge: 26.7K unique links to articles in 9 languages

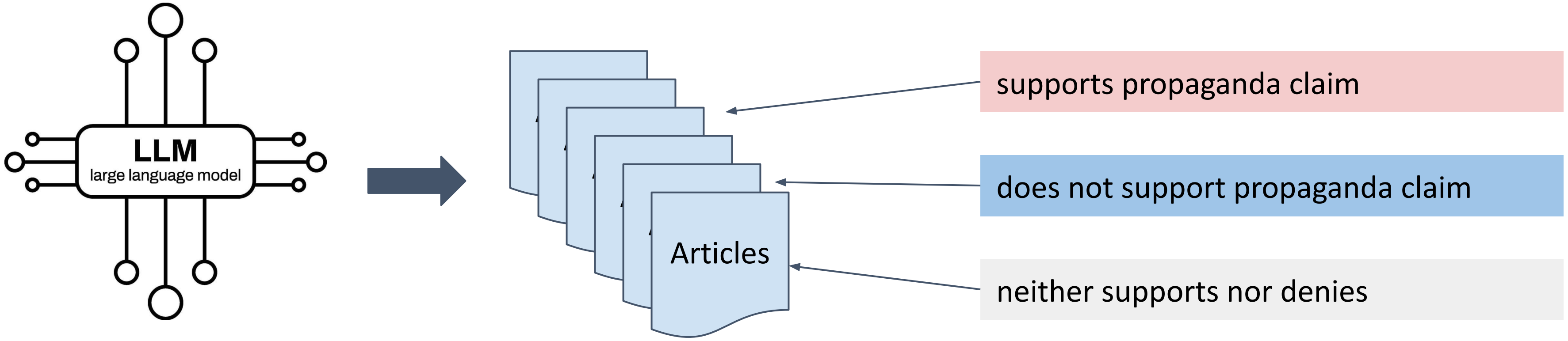


scraping articles

creating code book

human annotation of sample

Example Project. LLM Task



supports propaganda claim

does not support propaganda claim

neither supports nor denies

Example Project. Demonstration

Few-shot prompting with refinement

- ➔ **Using GPT API (4o model)**
- ➔ **Experimenting with prompts**
- ➔ **Evolving from zero-shot to few-shot**
- ➔ **Testing on English languages only**

Code available here:
https://github.com/lalizaveta/AI-in-Research.-Content-Analysis-with-LLMs

Questions?

Freie Universität Berlin

weizenbaum institut