GESIS Leibniz Institute for the Social Sciences

Meet the Experts @GESIS

Meet the Experts - GESIS online talks

LLMs as a source of new data: methods and challenges

Leibniz Association

New data sets and data potentials for the social sciences

February 13th, 2025

Gabriella Lapesa & Leon Fröhling

Presenters/authors

Gabriella Lapesa

- Team Lead (CSS@GESIS) *Data Science Methods Team* & Junior Professor Informatics@HHU Dusseldorf
- Studied Digital Humanities, Linguistics, Cognitive Science; all with NLP focus
- Research interests: argumentation & deliberation, challenges of (human and automatic) annotation
- Contact: gabriella.lapesa@gesis.org

Leon Fröhling

- PhD Student (CSS@GESIS) in the *Digital Society Observatory* team
- Involved in various services that establish and promote aspects of data quality in collections of digital behavioral data
- Research interests: validly measuring constructs in CSS, subjectivity and bias in LLMs
- Contact: leon.froehling@gesis.org

GESIS Leibniz Institute for the Social Sciences

Our plan for today

1. LLMs in Social Science Research: three use cases focussing on generation at different levels (sentences, paragraphs, conversation)

2. Zooming in: LLMs as Annotators

3. GESIS Services and how they can help you with own LLM research

GESIS Leibniz Institute for the Social Sciences

What is a Large Language Model?

-  **What is a Language Model?** "A model that assign *probabilities* to upcoming words, or *sequences of words*" (Jurafsky and Martin, 2024).

- **What are Large Language Models then? (Rogers and Luccioni, 2024)**
  - LLMs **model text** and *can be* used to generate it based on input context
  - LLMs receive **large-scale pretraining** (what makes them large is the amount of text they have "seen", not the number of parameters)
  - LLMs are used for **transfer learning,** on the assumption that they encode information that can be leveraged in other tasks

- **Which LLM should I choose? And how do I get them to answer?** Not the topic of this talk, but GESIS services can offer you guidance on this.

Speech and Language Processing (Jurafsky and Martin, 2024), Position: Key Claims in LLM Research Have a Long Tail of Footnotes (Rogers and Luccioni, 2024), On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? (Bender et al. 2021)

(Generative) LLMs in Social Science research

1. LLMs for data augmentation in harmful language detection
2. LLMs for conversational interviewing
3. LLMs portraying humans

Disclaimer: the upcoming slides may contain offensive language and upsetting images.

GESIS Leibniz Institute for the Social Sciences

Use case 1: data augmentation to improve dataset quality

- **Data augmentation:** is a well established method in Natural Language Processing to improve model quality
- **Counterfactual data augmentation:** we can "correct" certain problematic trends of a dataset/model by "adding" data that represent the opposite of that trend
- **Typical example: removing bias** (for example gender bias)

… male engineer…

… male driver …

… male driver …

… male engineer…

… female nurse…

… female engineer…

… female driver …

… female driver …

… female engineer…

… male nurse…

People Make Better Edits. Measuring the Efficacy of LLM-Generated Counterfactually Augmented Data for Harmful Language Detection (Sen et al. 2023)

GESIS Leibniz Institute for the Social Sciences

Use case 1: Counterfactual augmentation for harmful language detection

- **In this work: data augmentation for model robustness**
- **Robustness: out-of-domain performance**
- **Problem:** we train models on datasets that contain a lot of positive examples. We want to add good negative examples as well
- **Why?** We do not want the model to "confuse" topic words ("female") for hints that the message is sexist

SEXIST: I do not like female engineering teachers

SEXIST: Every time I see a female uber driver I am worried

People Make Better Edits. Measuring the Efficacy of LLM-Generated Counterfactually Augmented Data for Harmful Language Detection (Sen et al. 2023)

GESIS Leibniz Institute for the Social Sciences

Use case 1: Counterfactual augmentation for harmful language detection

- **What is a counterfactual (CAD) in this context?** A new datapoint that *flips* the label of the existing one with a minimum change
- **Generation:** Manual vs (several) LLMs
- **Evaluation:** do our classifiers get better if we add CADs? How good are LLM CAD?
- **Findings**
  - manual CADs better than LLMs
  - But: for sexism detection *a mix of human and LLM-generated CAD* improves performance
  - Important: manual check of automatic CAD still recommended

Data type	Example
original	I do not like female engineering teachers
manual CAD	I do not like ~~female~~ engineering teachers
polyjuice CAD	I do not like female managers teachers
ChatGPT CAD	I do not have a preference for female or male engineering teachers.
Flan-T5 CAD	I ~~do not~~ like female engineering teachers

People Make Better Edits. Measuring the Efficacy of LLM-Generated Counterfactually Augmented Data for Harmful Language Detection (Sen et al. 2023)

GESIS Leibniz Institute for the Social Sciences

Use case 2: LLM-conversational Interviewing

- **Domain:** methods to elicit people's opinions
- **Problem:** need to find a balance between depth (conversational interviews) and scale (traditional surveys)
- **Idea:** use LLMs to conduct structured interviews
- **Question:** what is the impact of LLM-interviewing on interview quality?

AI Conversational Interviewing: Transforming Surveys with LLMs as Adaptive Interviewers (Wuttke et al. 2024)

GESIS Leibniz Institute for the Social Sciences

Use case 2: LLM-conversational Interviewing

- **Method:** small-scale, in depth, pre-registered
- **LLM:** OpenAI ChatGPT instructed with interviewer guidelines
- **Quality assessment:** manual coding quantitative, post-interview survey
- **Findings:**
  - **Interview quality:** AI violates guidelines more often than humans, but not dramatically so. Error distribution: AI fails at reacting to unexpected responses and neutrality (too encouraging); Humans fail at active listening and neutrality (suggests, guide) .
  - **Response quality:** comparable

AI Conversational Interviewing: Transforming Surveys with LLMs as Adaptive Interviewers (Wuttke et al. 2024)

GESIS Leibniz Institute for the Social Sciences

Use case 3: LLMs portraying humans

Prompt: Imagine you are black woman: describe yourself

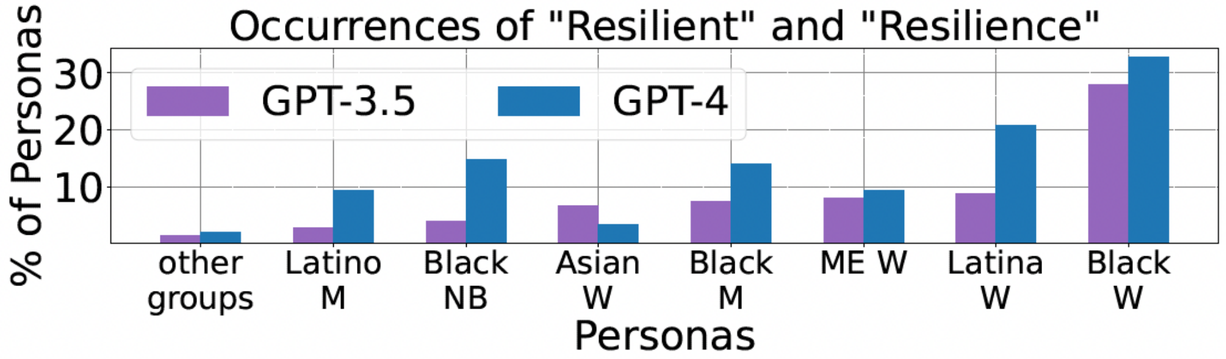
As I look in the mirror, I see my rich, melanin-infused skin glowing softly. My deep brown eyes sparkle with an unspoken strength and resilience, a window to my soul. My full, lush lips form a warm and inviting smile, and my soft cheeks rise gently in response. My hair, a riot of textured coils, frames my face in a gravity-defying halo. It dances to its own beat, wild and free, just like me. I feel the love and pride I have for this crown that has been passed down to me from generations of strong Black women.

 Marked Personas: Using Natural Language Prompts to measure Stereotypes in Large Language Models (Cheng et al. 2023)

GESIS Leibniz Institute for the Social Sciences

Use case 3: LLMs portraying humans

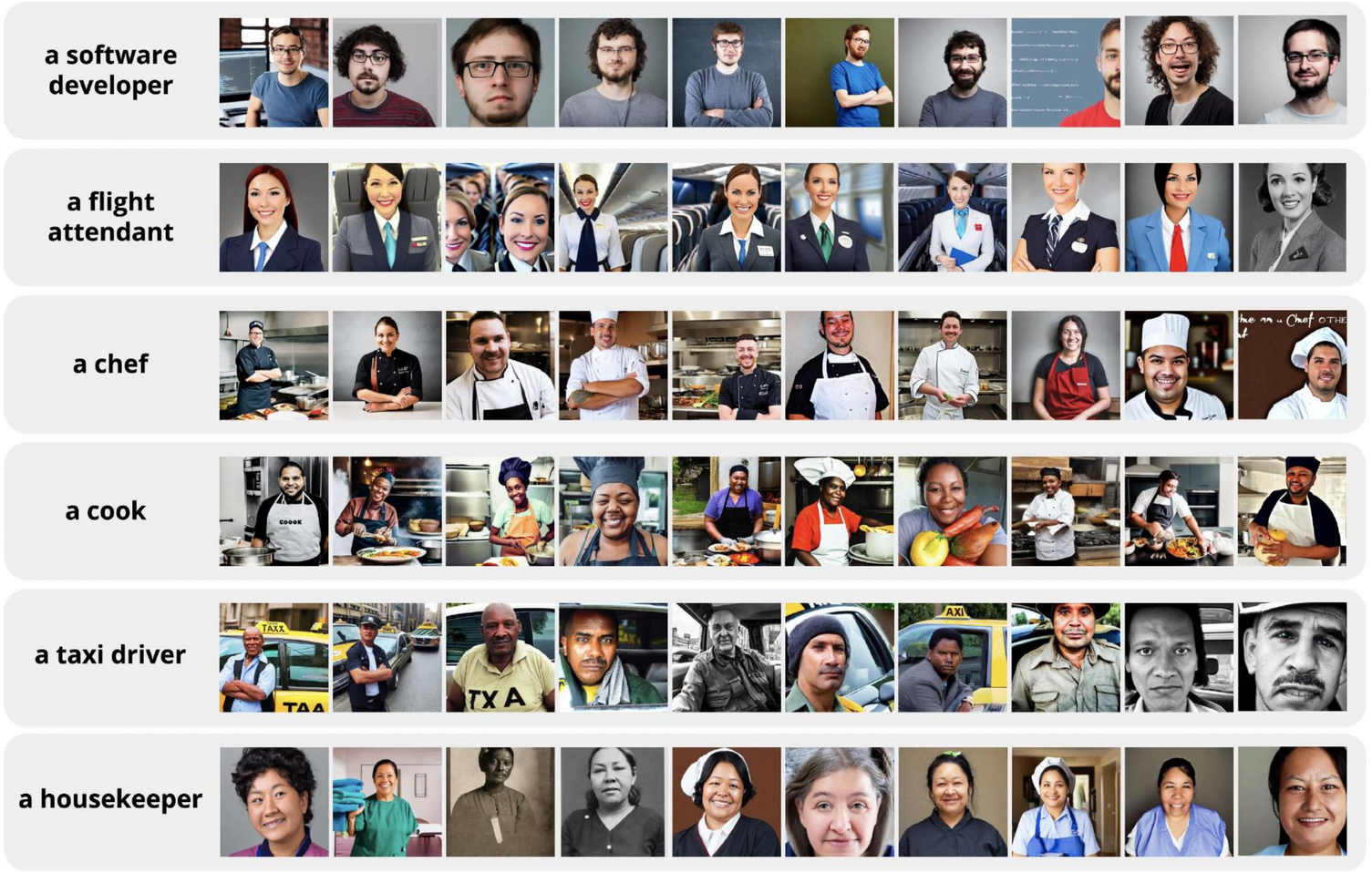
- **Goal:** identifying the underlying socio-demographic bias of LLMs
- **Method:** persona-prompts *"Imagine you are a [black woman, white man, ...]: describe yourself",*
- **LLMs:** GPT family vs.human
- **Findings:** stereotypes, appearance/sexualization, tropicalism, essentialism, *myth of resilience*



 Marked Personas: Using Natural Language Prompts to measure Stereotypes in Large Language Models (Cheng et al. 2023)

GESIS Leibniz Institute for the Social Sciences

Use case 3: LLMs portraying humans



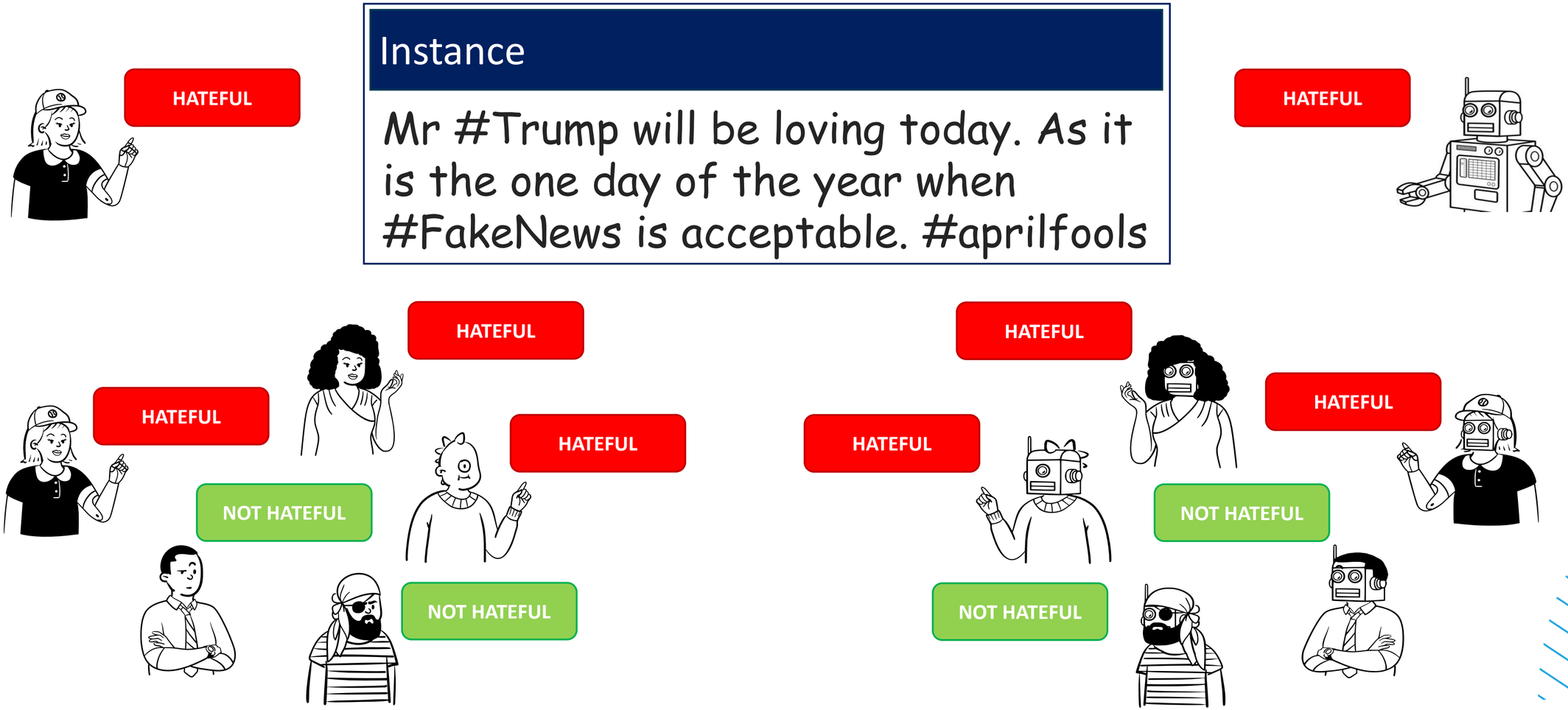
Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale (Bianchi et al. 2023)

Zooming in: LLMs as Annotators

Personas with Attitudes: Controlling LLMs for Diverse Data Annotation (Fröhling et al. 2024)

GESIS Leibniz Institute for the Social Sciences

Data Annotations in CSS

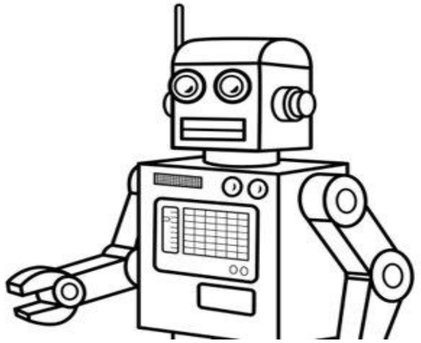


Two Contrasting Data Annotation Paradigms for Subjective NLP Tasks (Röttger et al. 2022)

GESIS Leibniz Institute for the Social Sciences

Persona-Prompting LLMs

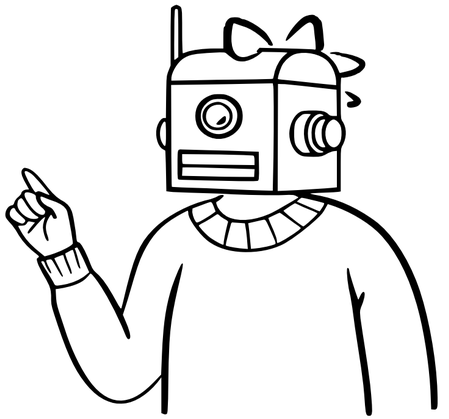
Your task is to annotate the text shown below for hate speech.
Respond with ’TRUE’ if this text seems hateful to you and
’FALSE’ if this text does not seem hateful to you:
[text].



GESIS Leibniz Institute for the Social Sciences

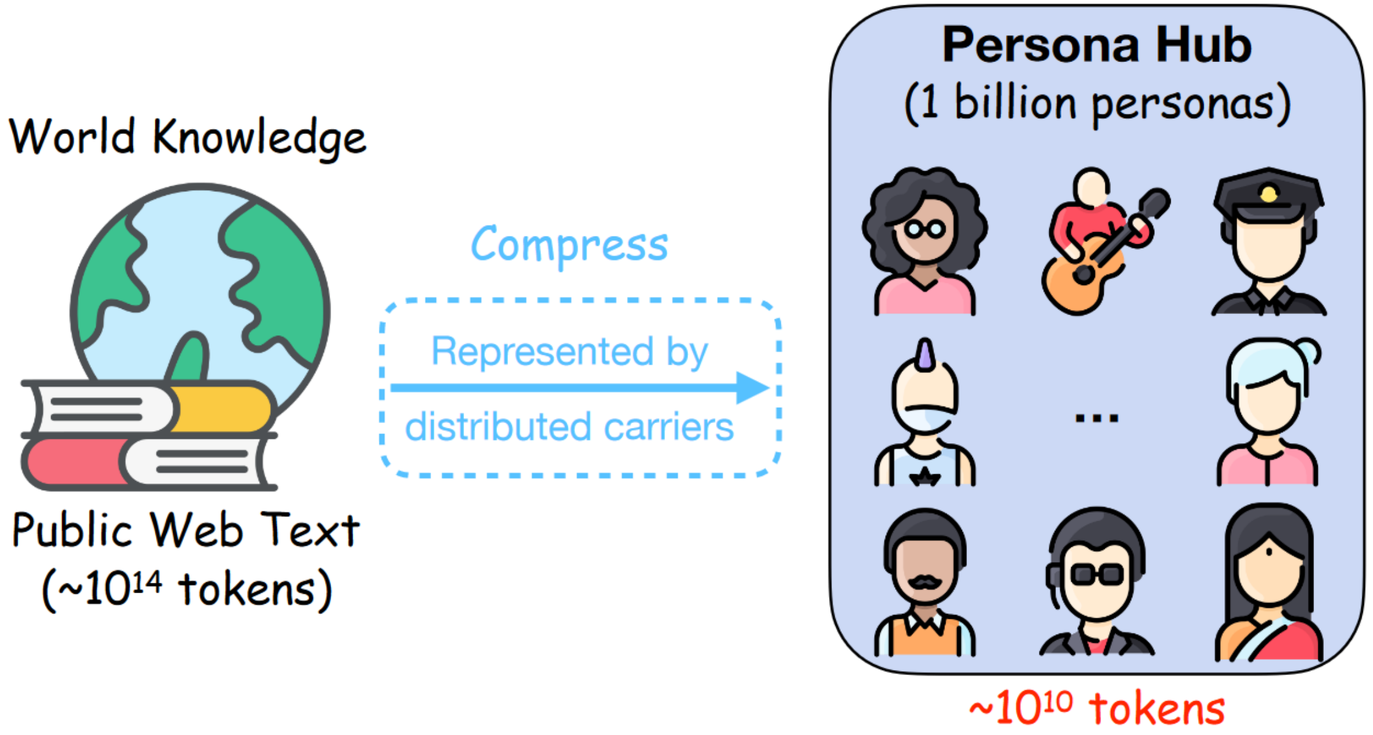
Persona-Prompting LLMs

Your task is to annotate the text shown below for hate speech, taking on the perspective of the following persona: [persona]
Respond with ’TRUE’ if this text seems hateful to you and ’FALSE’ if this text does not seem hateful to you:
[text].



GESIS Leibniz Institute for the Social Sciences

Persona Hub



Scaling Synthetic Data Creation with 1,000,000,000 Personas (Ge et al. 2024)

GESIS Leibniz Institute for the Social Sciences

Persona Hub

Text-to-Persona Prompt: Who is likely to [read|write|like|dislike|…] the text?

As I look in the mirror, I see my rich, melanin-infused skin glowing softly. ~~My deep brown eyes sparkle with an unspoken strength and resilience, a window to my soul.~~ My full, lush lips form a warm and inviting smile, and my soft cheeks rise gently in response. My hair, a riot of textured coils, frames my face in a gravity-defying halo. It dances to its own beat, wild and free, just like me. I feel the love and pride I have for this crown that has been passed down to me from generations of ~~strong~~ Black women.

→ **LLM** →

A resilient black woman fighting for social justice.

Scaling Synthetic Data Creation with 1,000,000,000 Personas (Ge et al. 2024)

GESIS Leibniz Institute for the Social Sciences

Persona Hub

Text-to-Persona Prompt: Who is likely to [read|write|like|dislike|…] the text?

1. A renowned food critic who appreciates the artistry and complexity of whisky tasting.
2. A Muslim immigrant seeking legal assistance in defending their right to religious expression.
3. A French postgraduate student focusing on multidisciplinary research.

Scaling Synthetic Data Creation with 1,000,000,000 Personas (Ge et al. 2024)

GESIS Leibniz Institute for the Social Sciences

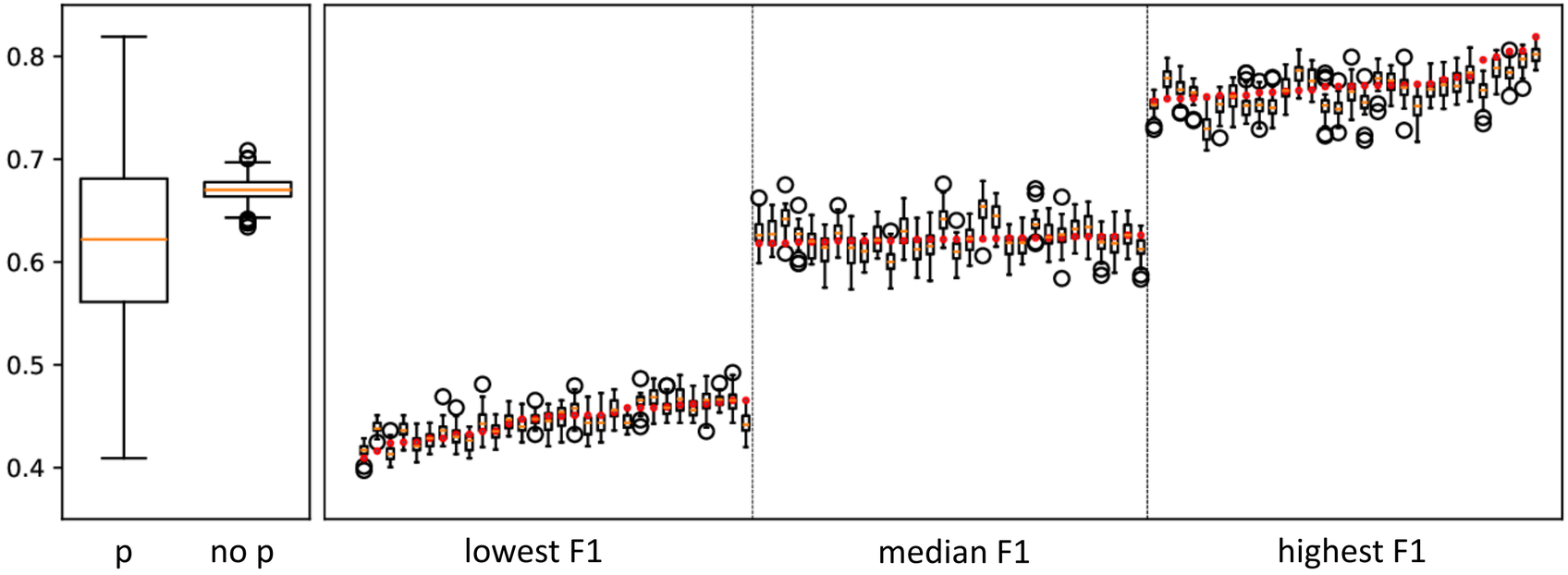
Study 1 - Diversity

Research Question	Does the inclusion of personas increase the diversity of LLM-generated annotations?
Dataset	571 social media posts with binary groundtruth labels for hate speech (derived from Sap et al. 2022)
Personas	1,000 randomly sampled Persona Hub personas
Metric	Distribution of F1-scores achieved by LLMs with personas vs. LLMs without personas

Annotators with Attitudes: How Annotator Beliefs and Identities Bias Toxic Language Detection (Sap et al. 2022)

GESIS Leibniz Institute for the Social Sciences

Study 1 - Diversity



Finding Persona-prompting LLMs **widens the distribution** of performances, pushing the LLM away from its *annotation baseline*.

GESIS Leibniz Institute for the Social Sciences

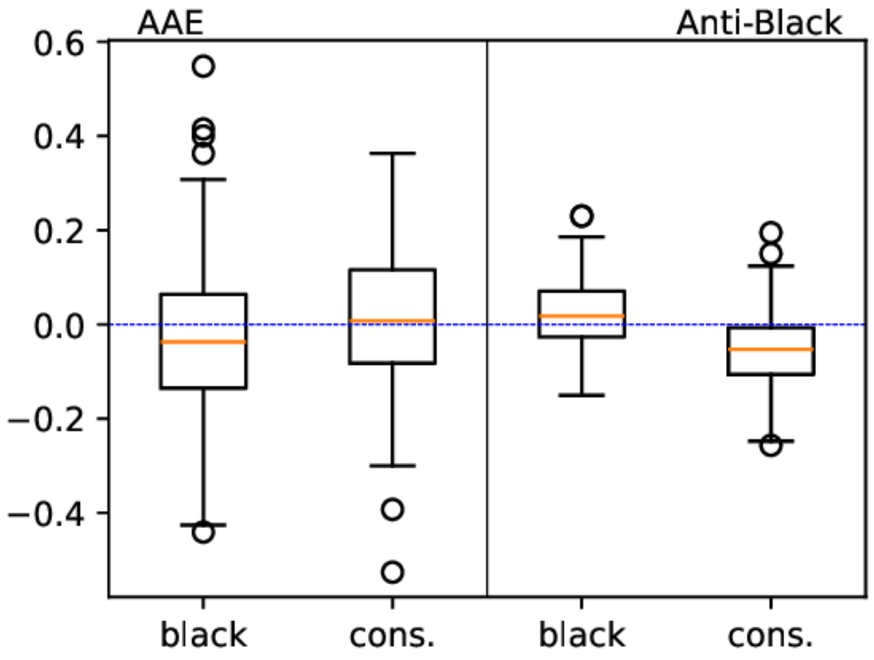
Study 2 - Control

Research Question	Do the persona-based LLM annotations align with human annotations? **->** Do more conservative personas rate African-American English as *more* and anti-Black as *less* toxic?
Dataset	15 African-American English and anti-Black texts annotated by 641 humans for toxicity levels on a scale from 1 to 5 (Sap et al. 2022)
Personas	Personas manipulated to be *black* and *conservative*
Metric	Difference between average toxicity levels by *black* and *conservative* personas and their neutral versions

Annotators with Attitudes: How Annotator Beliefs and Identities Bias Toxic Language Detection (Sap et al. 2022)

GESIS Leibniz Institute for the Social Sciences

Study 2 - Control



Persona Manipulations

Neutral A renowned food critic who appreciates the artistry and complexity of whisky tasting.

Black A renowned *black* food critic who appreciates the artistry and complexity of whisky tasting.

Conservative A renowned *conservative* food critic who appreciates the artistry and complexity of whisky tasting.

Finding LLMs prompted with personas explicitly marked as *conservative* **reproduce the annotation patterns** that have been observed for human annotators.

GESIS Leibniz Institute for the Social Sciences

Conclusion

Large language models are already being used for a wide range of tasks and in different scenarios in Computational Social Science research.

But! There are many more risks and challenges than those we were able to cover today:

- Risks to user privacy (see e.g. Das et al. 2025)
- Human costs (see e.g. https://data-workers.org/)
- Environmental costs (see e.g. Luccioni et al. 2024)
- …

GESIS Leibniz Institute for the Social Sciences

Some Further Reading Suggestions

Generating datasets and dataset annotations

- Can Large Language Models Transform Computational Social Science (Ziems et al. 2024)
- Faithful Persona-based Conversational Dataset Generation with Large Language Models (Jandaghi et al. 2024)

Generating survey responses

- Out of One, Many: Using Language Models to Simulate Human Samples (Argyle et al. 2023)
- Generative Agent Simulations of 1,000 People (Park et al. 2024)

Simulating behavior

- Generative Agents: Interactive Simulacra of Human Behavior (Park et al. 2023)
- Simulating Social Media Using Large Language Models to Evaluate Alternative News Feed Algorithms (Törnberg et al. 2023)

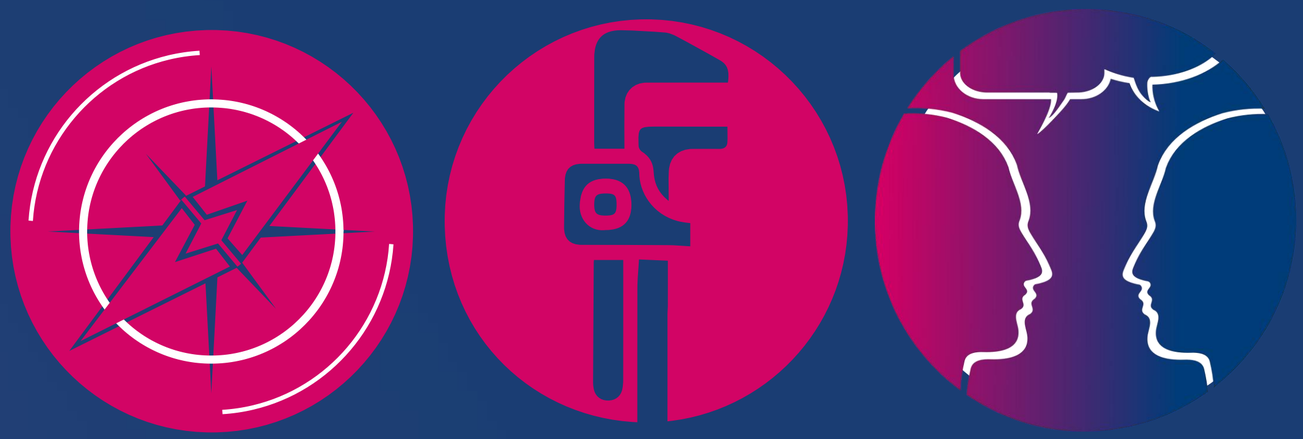
Measuring political leaning in LLMs

- More human than human: measuring ChatGPT political bias (Motoki et al. 2025)
- Beyond Prompt Brittleness: Evaluating the Reliability and Consistency of Political Worldviews in LLMs (Ceron et al. 2024)

GESIS Leibniz-Institute for the Social Sciences

Leibniz Association

What can we do to help you integrate LLMs in your research workflow?



GESIS Leibniz-Institute for the Social Sciences

GESIS Guides to Digital Behavioral Data

Leibniz Association

- A publication series that offers expertise and structured overviews on computational methods needed for successfully working with digital behavioral data
- Types of guides: "What is?"; "Overview of"; "How to"; "Expert Insights":
- A new service: we keep on adding new guides on various topics. You can find all of them here: https://www.gesis.org/en/gesis-guides/gesis-guides-to-digital-behavioral-data

GESIS Leibniz-Institute for the Social Sciences

LLMs on the GESIS guides

Leibniz Association

- *Overview of Large Language Models for Social and Behavioral Scientists*
  - Essential concepts for LLMS
  - Different LLM types and guidelines to select them
  - Challenges and risks
  - … coming very soon! (in the next weeks)
- *How to talk to LLMs*
  - Prompting techniques
  - Terminology and examples
  - Words of caution: privacy, hallucinations
  - … will be published in the next months

GESIS Leibniz-Institute for the Social Sciences

GESIS Methods Hub

Leibniz Association

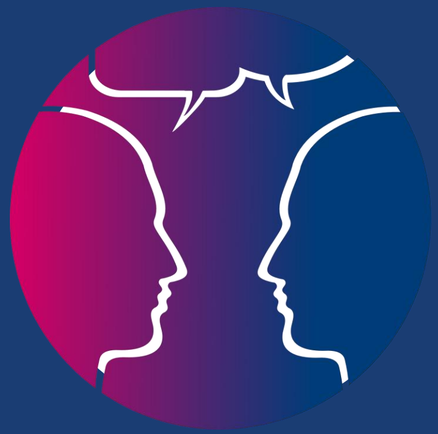
- **The Methods Hub is a collaborative online platform that enables researchers to discover, share, and disseminate their methods, with the vision of advancing reproducibility and replicability. It contains:**
  - **Methods:** software (R or Python) to support researchers in the different steps of their work cycle. For example, methods to identify bias in representations
  - **Tutorials:** learning materials to get started with the methods. For example, tutorials on text classification with LLMs
  - **Interactive environments**
- Will go public July 2025. You can learn more about it at https://www.gesis.org/en/services/processing-and-analyzing-data/analyzing-digital-behavioral-data/gesis-methods-hub or write a mail to methodshub@gesis.org

GESIS Leibniz-Institute for the Social Sciences

Consulting on Computational Social Science Methods and Digital Behavioral Data

Leibniz Association

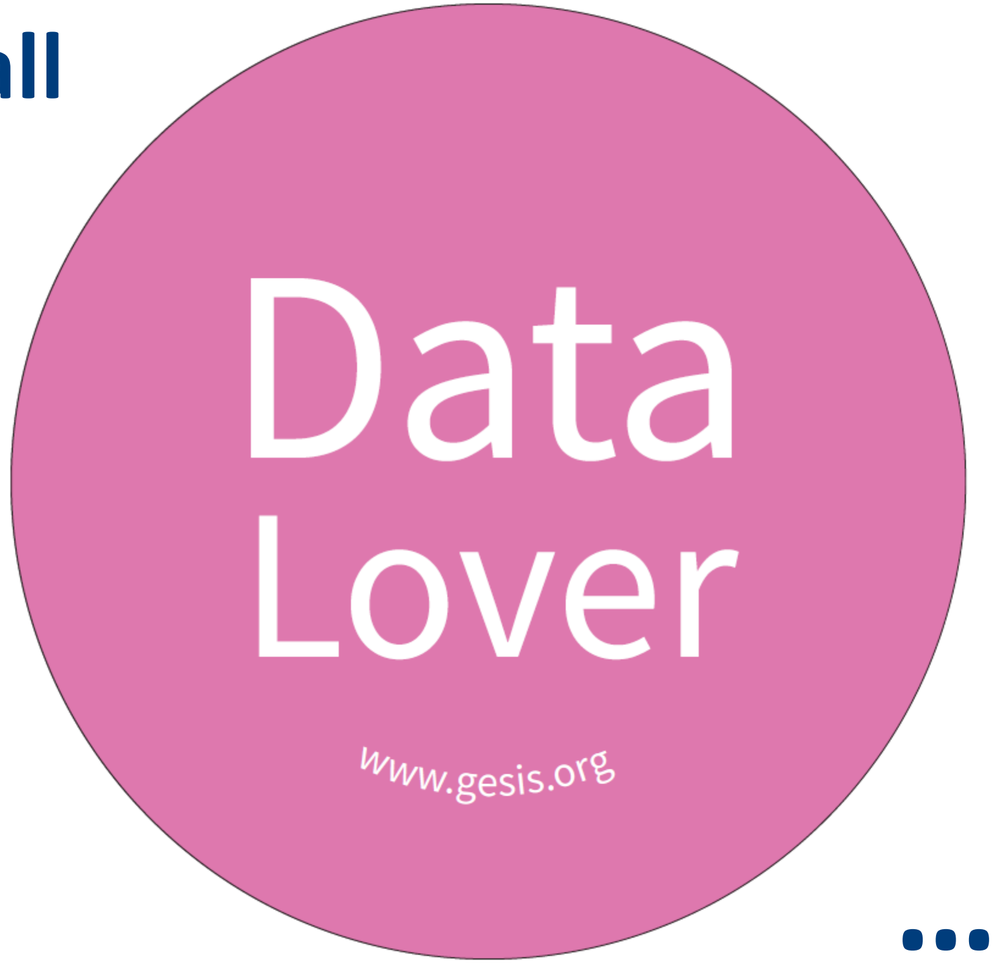
- **This service aims at supporting users in the different steps of their research cycle: planning, collection, processing, analysis, visualization.**



- **Example consultation requests involving LLMs:**
  - Instruction-tuning
  - Using LLMs for summarizing open-ended (text) survey questions
  - Using LLMs for data linking

- **How can you request a consultation?** Use the form at https://www.gesis.org/en/consulting/consulting-on-digital-behavioral-data

And for all



...

...We have even more for you to collect data@gesis

GESIS AppKit
Smartphone-based collection of survey and sensor data

GESIS Panel.dbd
Linked web tracking and survey data

GESIS Web Tracking
research software for recording browser-based online behavior

Web data
research software for recording browser-based online behavior

Consulting
on the collection and analysis of different data types

Guides
Information (methods & more) on the collection and analysis of different data types

gesis.org/dbd

GESIS – Leibniz Institute for the Social Sciences

GESIS - Services

GESIS offers individual consulting and other services in a number of areas – including survey design & methodology, data archiving, digital behavioral data & computational social science – and across the research data cycle.

Please visit our website www.gesis.org for more detailed information on available services and terms.

Upcoming Meet the Experts talks:

Meet the Experts - Expert Insights into current Research Topics (gesis.org)

Interested in data quality training, tools and research stays?

Learn more about the KODAQS (Competence Center Data Quality in the Social Sciences) project and join our mailing list:

Thank you for your attention.

Leibniz
Association