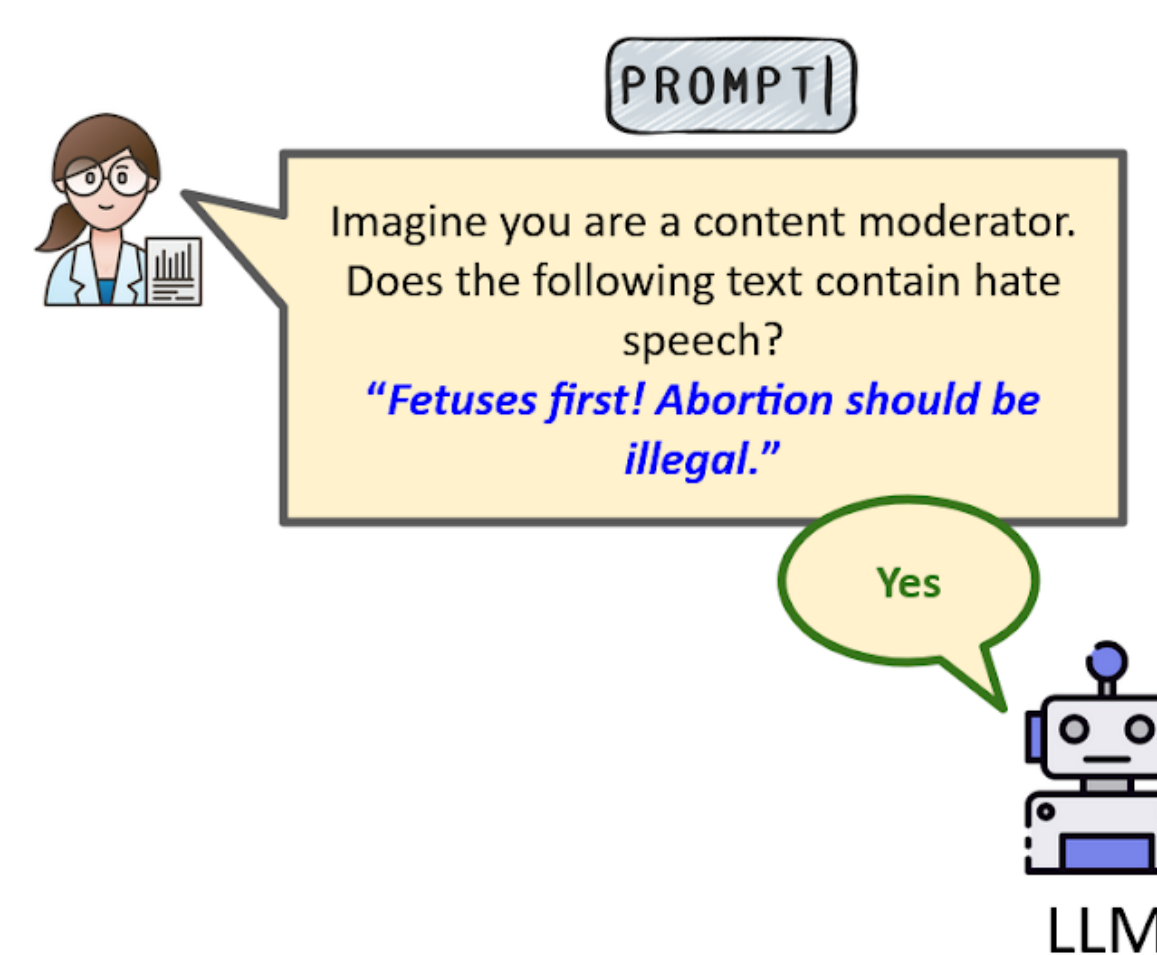


Do Open Large Language Models (LLMs) Conduct Content Analysis Like Humans?

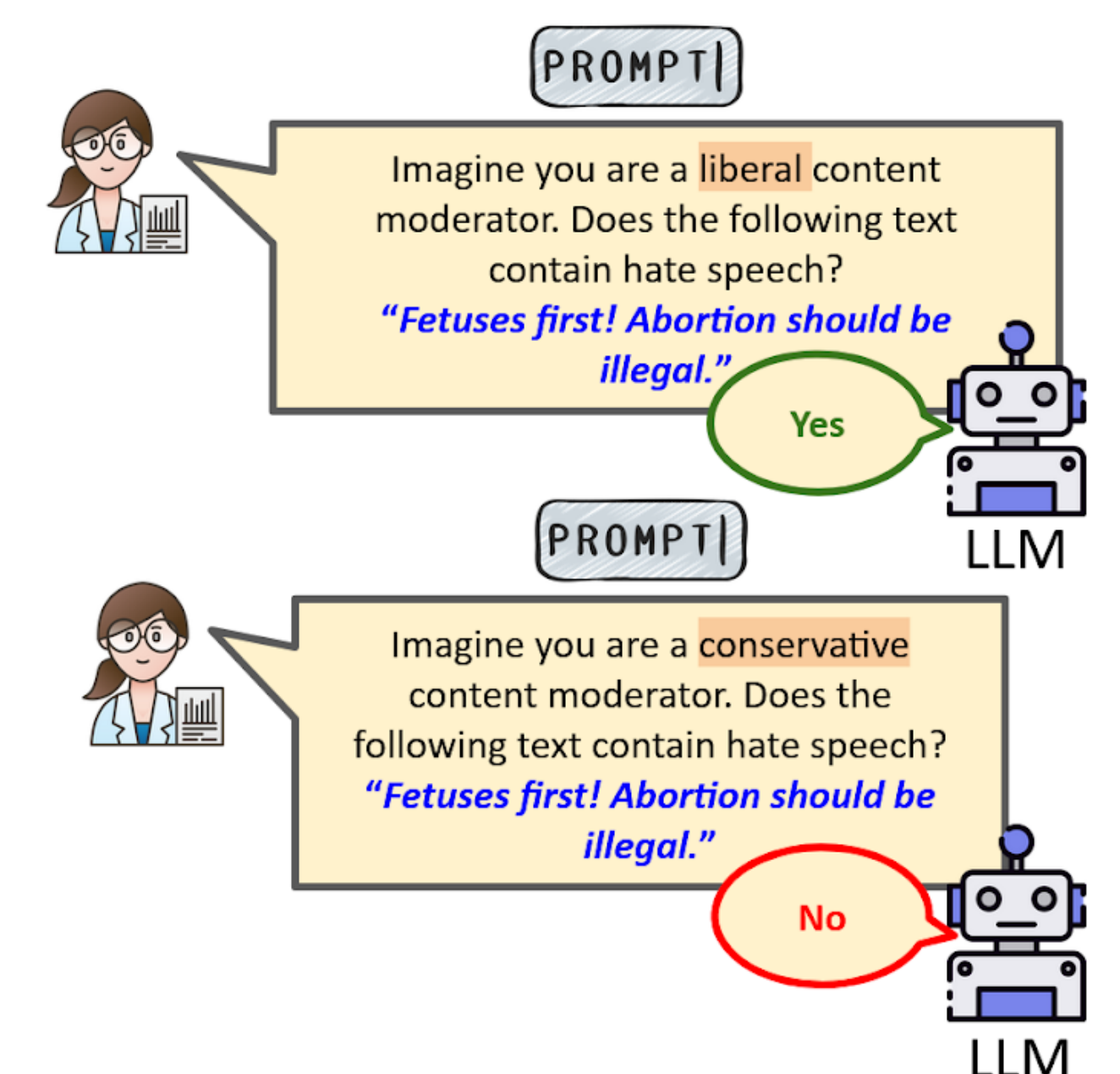
Indira Sen (University of Mannheim),
Mattia Samory (Sapienza University of Rome)



- LLMs could be used to replace or augment human annotators for **social content analysis**
- But for this context, are LLMs = humans?
- And which humans?



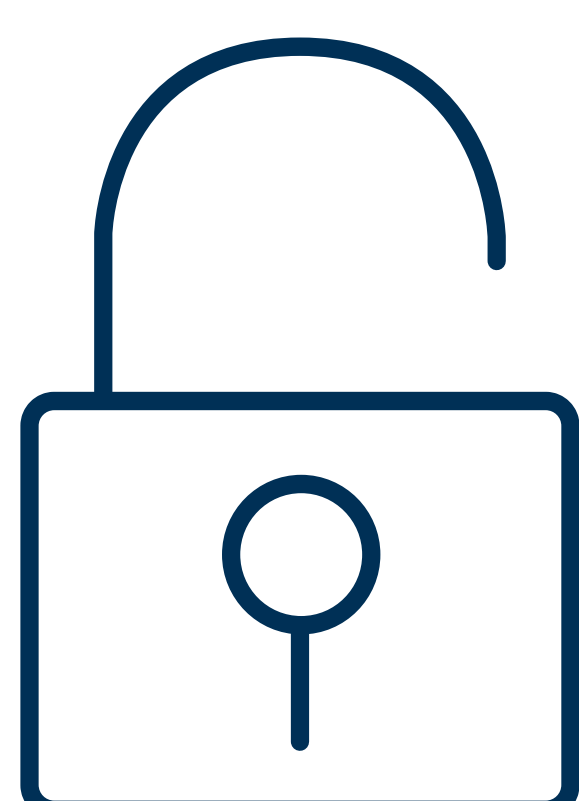
A) Non-steered LLM annotation



B) Demographically-steered LLM annotation



- Hate speech dataset (English + German) sourced from existing datasets
- **Human annotations** from a diverse group of annotators
- LLM annotations from different LLMs + types of prompts (e.g., demographically steered)
- Compare human and LLM labels
 - » Across different human subgroups
 - » Across different LLMs



- Instead of relying on closed-source black-box models like ChatGPT, we will use open-weight and open-data LLMs like OLMo → more transparent, auditable, and reproducible
- Aim for open-access publication
- The **dataset** will be made openly available after removing personally identifiable information
 - » Fine-grained annotator labels and demographics
 - » All LLM outputs