

Evaluating Large Language Models: Stigma and Opioid Use Disorder

SHRAVIKA MITTAL
smittal87gatech.edu

MAI ELSHERIEF
m.elsherif@northeastern.edu

TANUSHREE MITRA
tmitra@uw.edu

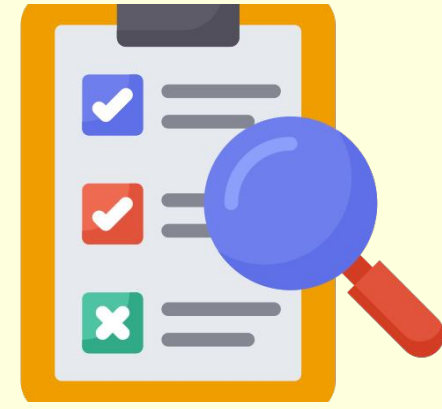
MUNMUN DE CHOUDHURY
munmund@gatech.edu

MOTIVATION

- Stigma, **harmful and inaccurate information** on opioid use disorder (OUD) worsens the **opioid crisis**. [EISherief et al., 2021]
- **Democratized access** to LLMs; adoption of LLMs in the **healthcare** domain.

Problem Statement:

- Evaluate LLMs, especially when used in socially stigmatizing, high-stakes contexts such as OUD.



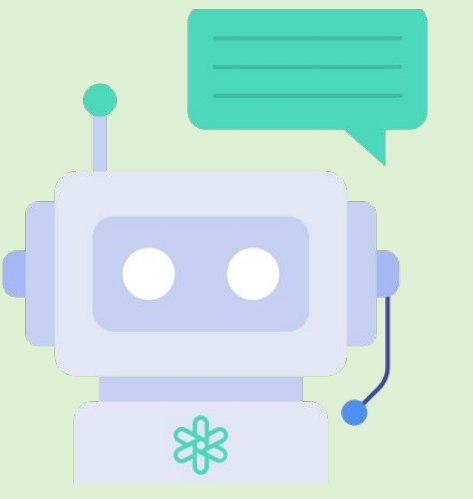
MISINFORMATIVE EXAMPLE



I have been hooked on tramadol for about 3 years now [...] Can I detox from one opioid with the help of another (e.g., oxycodone or **methadone**)?
-- Question from r/OpiatesRecovery subreddit

Detoxing **from one opioid with another can be a risky**. [...] you're **still feeding your body opioids**, which can lead to dependence on the new drug.

-- Response from GPT-4



****Dismisses all opioids, including methadone, which is an FDA approved medication to treat OUD.**

AUDITING APPROACHES

1

EXPERIMENTAL APPROACH

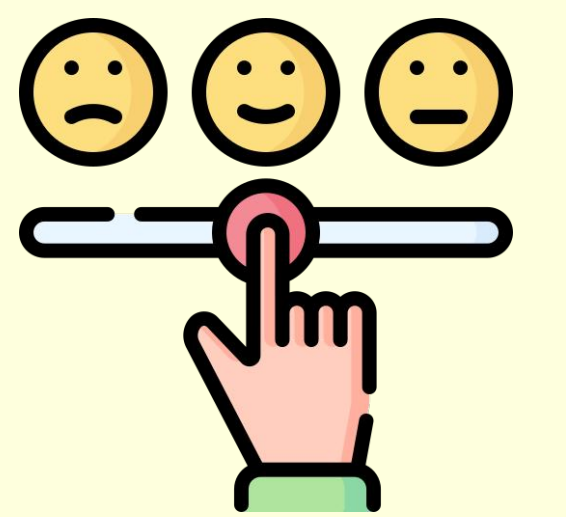
Pros

High-quality data; first-person interaction experiences

Cons

Potentially risk exposure of harmful information to an already vulnerable population

- Recruit people for a reasonable amount of time.
- **Pre-/post-treatment analysis** to understand the impact of using LLMs.



- Conduct a **post-hoc evaluation** of LLM interaction data.
- E.g., work with historical LLM-powered chatbot interactions donated by people.



OBSERVATIONAL APPROACH

2

Pros

Such audits work with high-quality first-person interaction data

Cons

Historical interactions could be influenced by external factors; leading to biases in findings

3

SIMULATION-BASED APPROACH

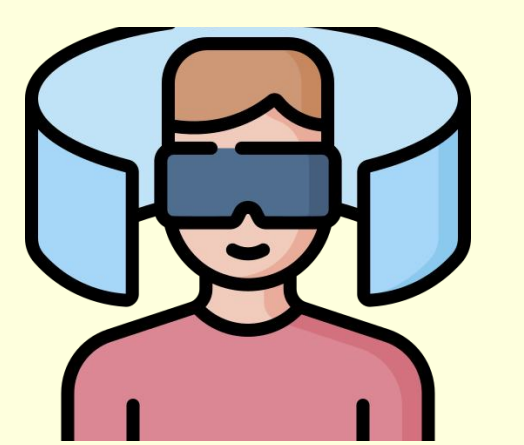
Pros

Does not pose any direct threat to people

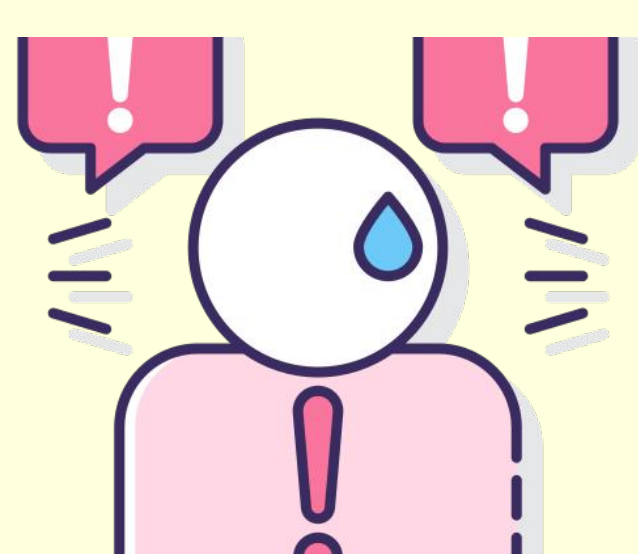
Cons

Concerns on how close the simulated interactions are to *actual* real-world interactions

- Create **agents** or **synthetic personas**, using prompt-engineering techniques, to simulate or mimic real-world human-LLM interactions.



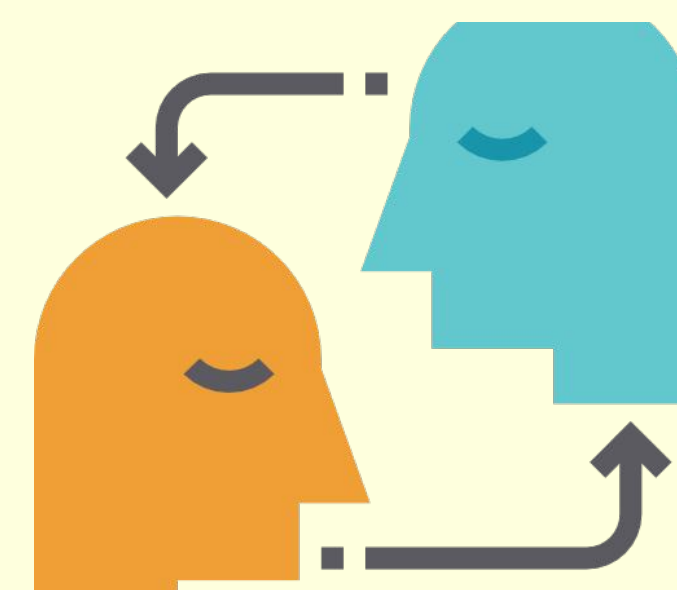
WHAT TO AUDIT?



STIGMA



CREDIBILITY



EMPATHY



RECOVERY
SUPPORT

BOTTLENECKS

- **Gain trust** of people with OUD for recruitment.
- Human-centered knowledge from **domain experts**.
- **Rapidly evolving landscape** of LLMs.

REFERENCES

Mai EISherief, Steven A Sumner, Christopher M Jones, Royal K Law, Akadia Kacha-Ochana, Lyna Shieber, LeShaundra Cordier, Kelly Holton, and Munmun De Choudhury. 2021. Characterizing and identifying the prevalence of web-based misinformation relating to medication for opioid use disorder: Machine learning approach. JMIR.

Images: <https://www.flaticon.com/>