

*Paper 1:
Leveraging AI for
democratic discourse:
Chat interventions can
improve online
political conversations
at scale (Argly et al.,
2023)*

Prosocial Behavior


How language technologies
can encourage prosocial
human behaviors?

*Paper 2:
Human-AI
collaboration enables
more empathic
conversations in
text-based
peer-to-peer mental
health support
(Sharma et al., 2023)*

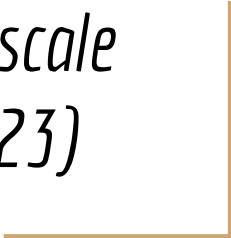
Agenda (for each paper)

1. Instructor: High-level summary
2. Instructor: Important and interesting figures/visualizations
3. Peer reviewer: analyzing strengths and weaknesses
4. Open discussion with reflective questions
5. Academic researcher: proposing new research directions
6. Industry practitioners: innovative products and applications

Conclude with common high-level themes



*Paper 1:
Leveraging AI for
democratic discourse: Chat
interventions can improve
online political
conversations at scale
(Argly et al., 2023)*



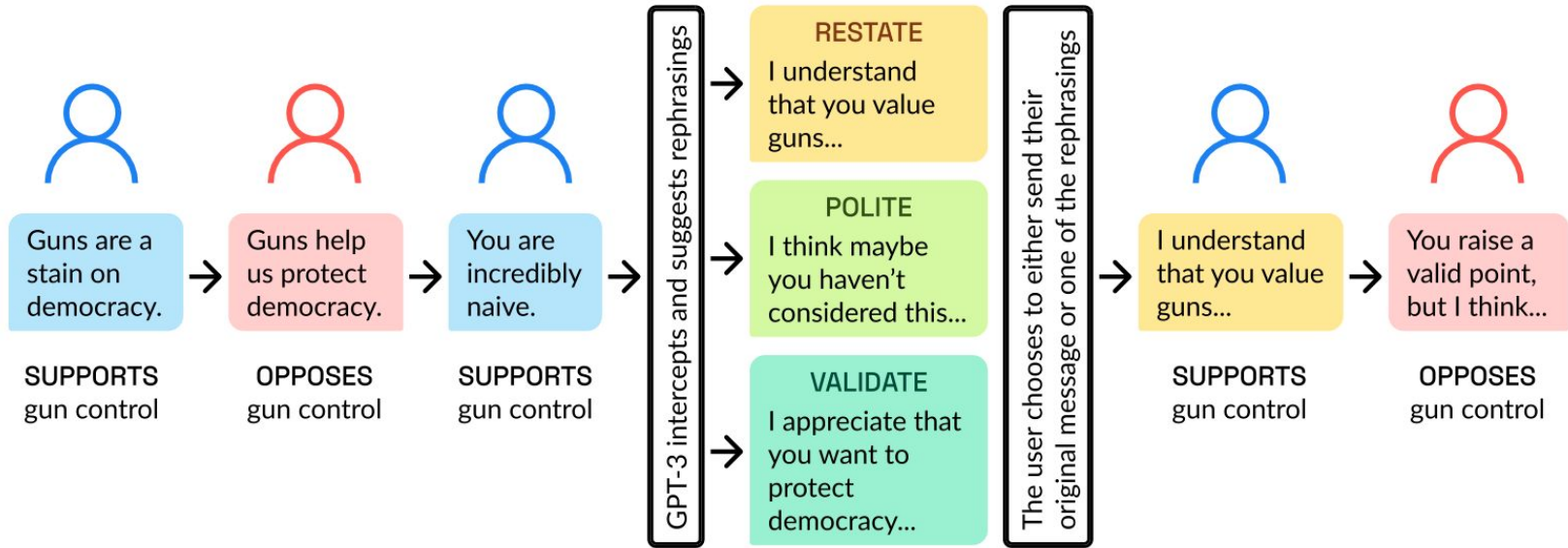
Summary: Leveraging AI for democratic discourse (Argly et al., 2023)

Objective: The study examines how AI tools can improve online conversations about divisive topics, enhancing democratic discourse.

Methodology: A large-scale experiment involving proponents and opponents of gun regulation in the U.S. Participants were randomly assigned to use a GPT-3 powered chat assistant that provided real-time, evidence-based suggestions to improve conversation quality and democratic reciprocity.

Key Interventions:

- Restatement: Repeating back a person's main point to show understanding.
- Validation: Affirming the legitimacy of different opinions without requiring agreement.
- Politeness: Modifying statements to use more polite language.



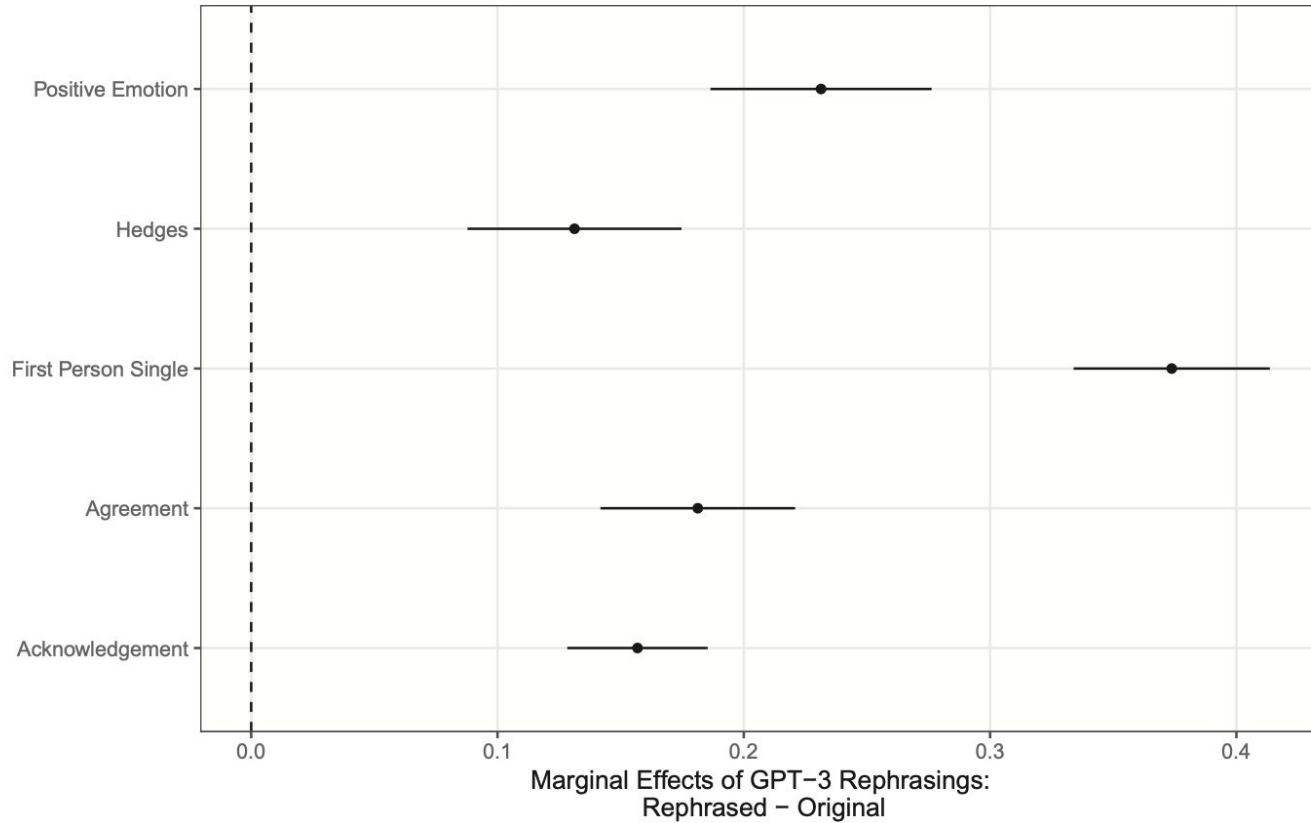
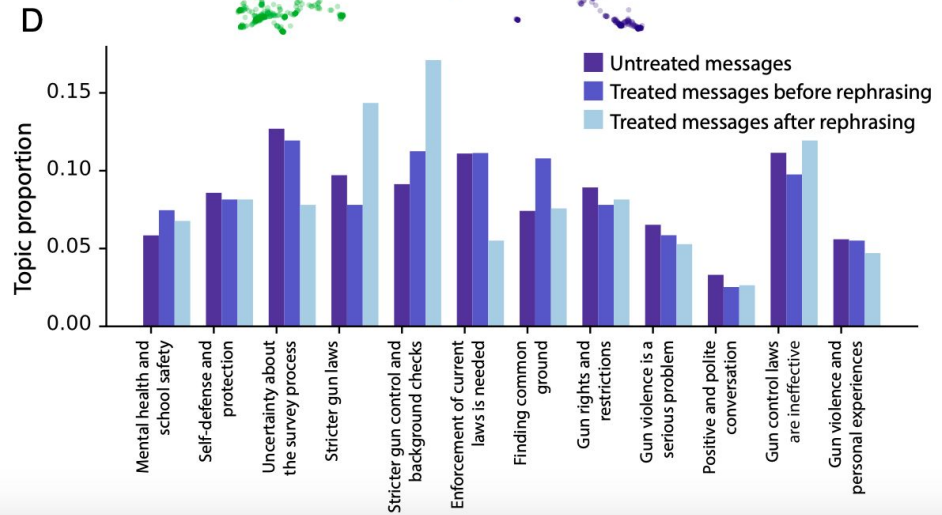
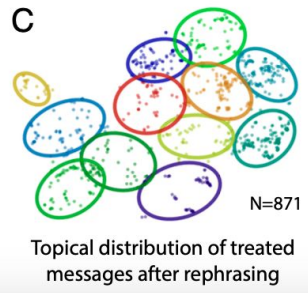
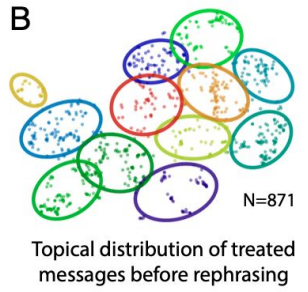
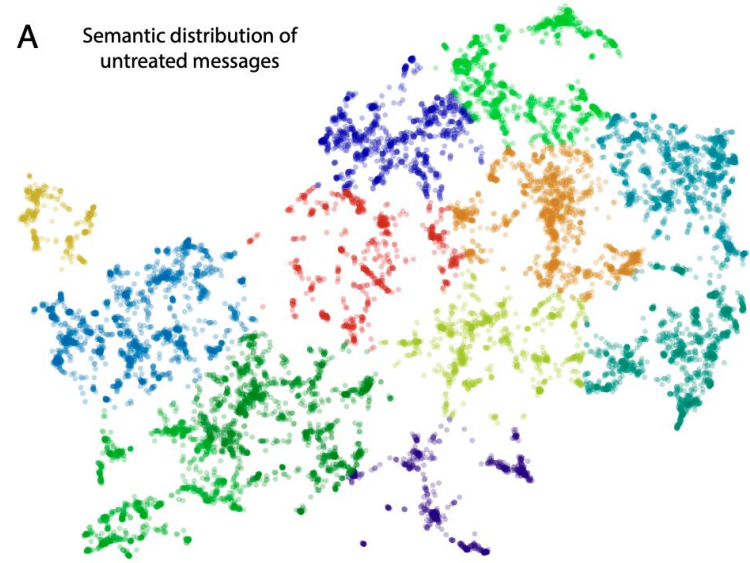


Fig. 2. Text analysis of rephrased messages tone: Marginal difference, with 95% CIs, between rephrased message scores on five politeness package features, and baseline scores from participants' original messages they would have sent had they not chosen the rephrasing.

- Mental health and school safety
- Self-defense and protection
- Uncertainty about the survey process
- Stricter gun laws
- Stricter gun control and background checks
- Enforcement of current laws is needed
- Finding common ground
- Gun rights and restrictions
- Gun violence is a serious problem
- Positive and polite conversation
- Gun control laws are ineffective
- Gun violence and personal experiences



Summary: Leveraging AI for democratic discourse (Argly et al., 2023)

Results:

- Improved Conversation Quality: Participants with the AI assistant reported higher conversation quality and a greater sense of being understood.
- Increased Democratic Reciprocity: The intervention promoted a willingness to grant political opponents space to express their views.
- No Change in Policy Attitudes: The intervention did not affect participants' policy positions.

Important Takeaways:

- AI tools can enhance the quality of political conversations and promote democratic reciprocity *without manipulating participants' views*, demonstrating the potential of AI to foster constructive dialogue in online settings.

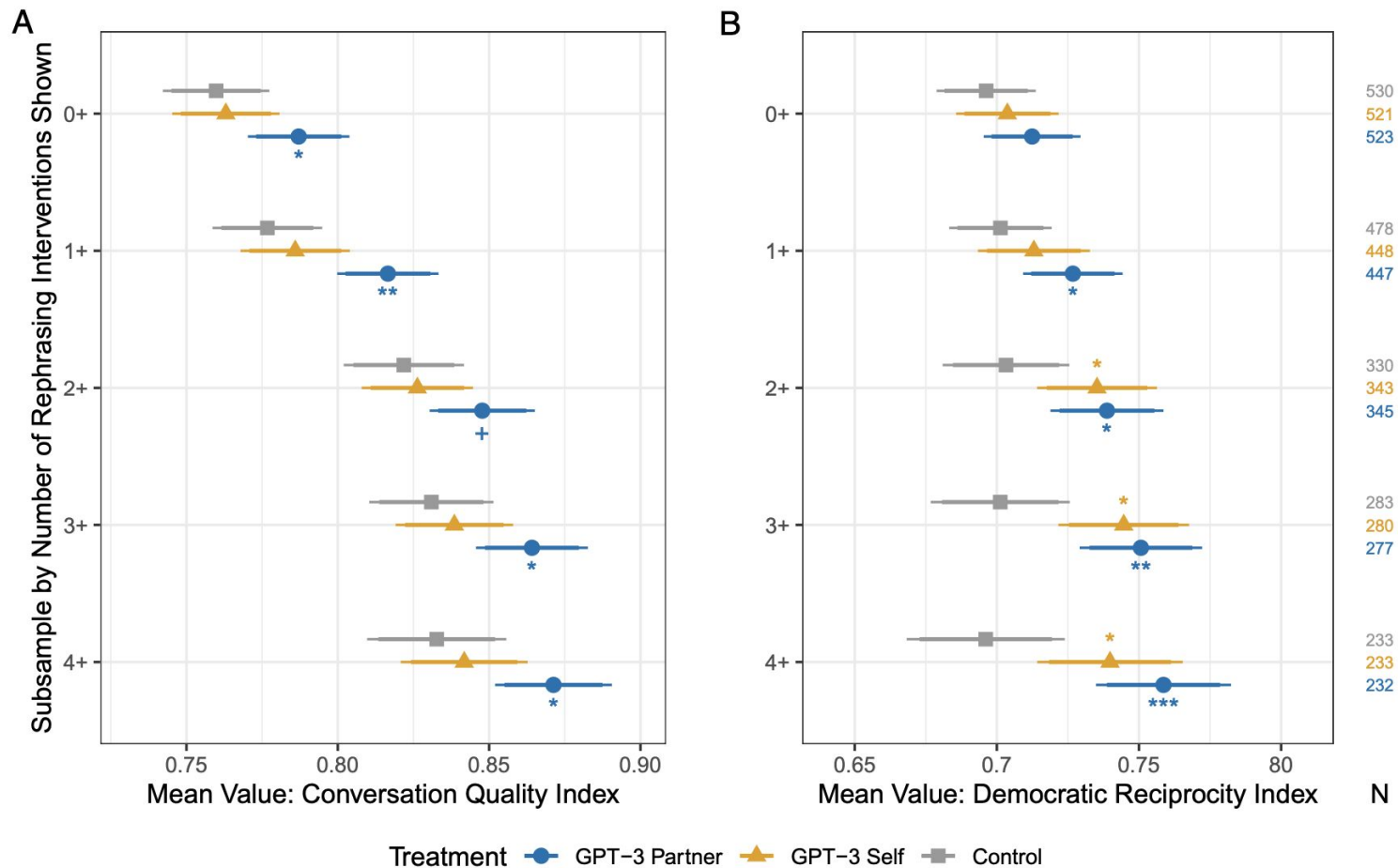
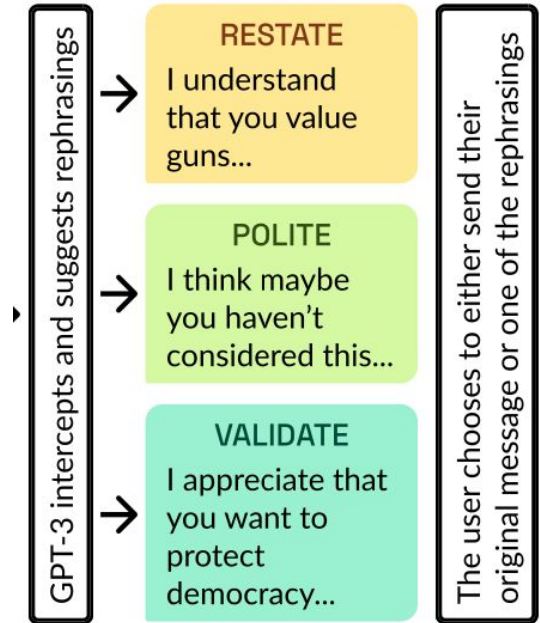


Fig. 4. Conversational quality Panel (A) and democratic reciprocity Panel (B) subgroup means and confidence intervals. Higher values signify greater quality/support. The number of rephrasing interventions are overlapping sets, such that 0+ includes all observations. The 90% and 95% CIs are based on unadjusted standard errors. Significance indicators stem from *t* test comparisons to the control condition in each subgroup: +*P* < 0.1, **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

Peer Reviewer Comments

Pros:

1. A hot, important research topic by its nature
2. Offered a three-dimensional strategy to enhance the feeling of being understood (increase the robustness of the technique)
 - Restatement
 - Validation
 - Politeness
3. A multi-method study (topic modeling, textual measures from other papers (e.g., receptiveness from Yeomans et al., 2020))



Peer Reviewer Comments

Pros:

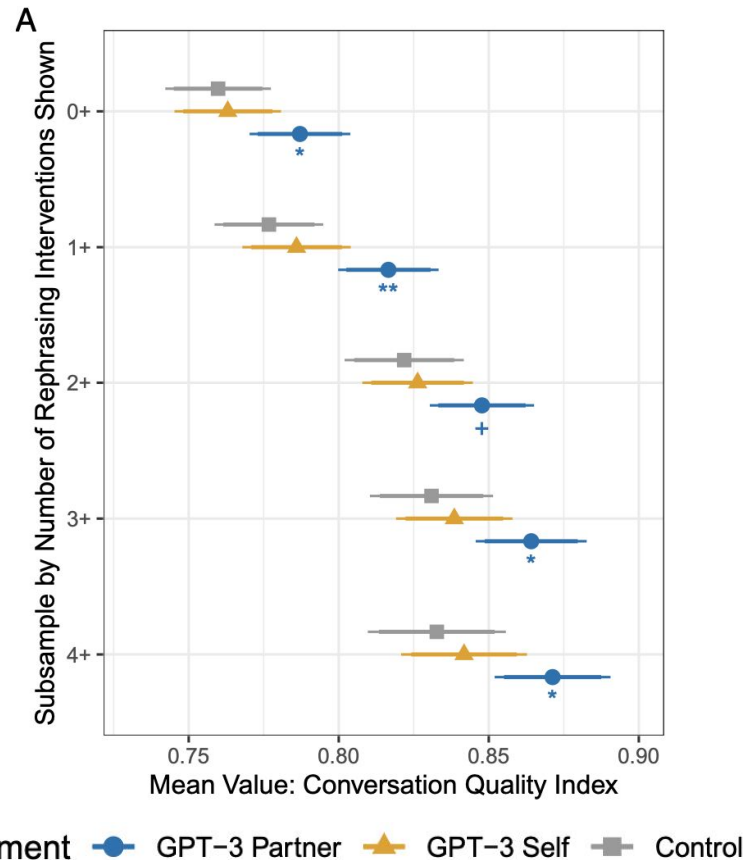
3. Validity of the random assignment and message improvement strategy

- AI-suggested rephrasings were accepted by chat participants two-thirds (1,798) of the time
- Accepted rephrasings were roughly evenly split between the restate (30%), validate (30%), and politeness (40%) interventions.

4. A mutual evaluation of outcomes

- Participants who got AI assistance
- Participants who are recipients of the message

5. Dataset (a large-scale field study)



Peer Reviewer Comments

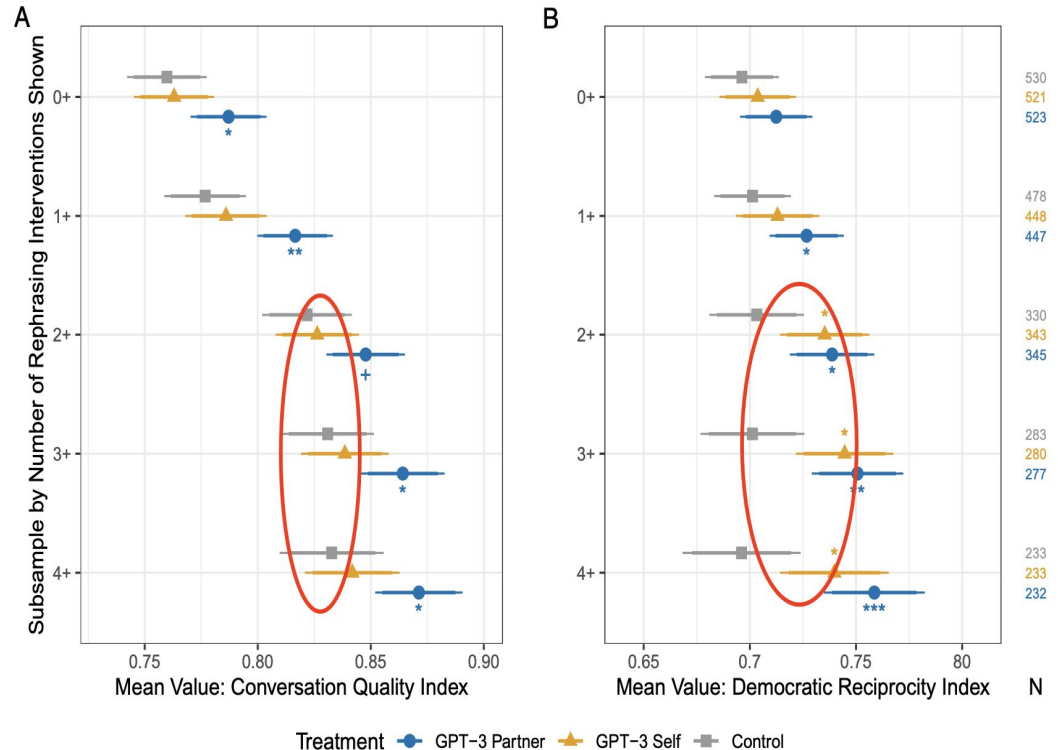
Cons:

1. We know little about any outcomes beyond the conversation...
 - Main dependent variables are only:
 - Quality of the conversation
 - Acknowledge the perspectives of others (i.e., so-called democratic reciprocity)
 - What about other more influential outcomes?
 - **Receptiveness:** The degree of openness to access opposing perspectives or consider opinions from disagreeing others (how receptive to receiving more info..)
 - **Information seeking intentions:** How willing are you to seek more information from your conversation partner's perspective?
 - **Word of mouth:** how willing are you to share this conversation with your family members/friends/colleagues?

Peer Reviewer Comments

Cons:

3. There is no difference between GPT-3 self and control for perceived conversation quality (why differ from reciprocity?)
4. One-shot experiment, what about the long-term effect?



Discussion Starters

Role of AI in Mitigating Polarization: Can AI effectively reduce political polarization, or might it inadvertently entrench existing biases? How can these risks be mitigated?

Ethical AI Interventions: What are the ethical implications of using AI to influence online political discourse, even if the aim is to improve conversation quality and democratic reciprocity?

Long-Term Effects: What might be the long-term impacts of AI-assisted conversation tools on public discourse and democratic engagement? Are there potential unintended consequences?

Academic Researcher

Title: **Culturally-Aware Discourse Intervention:** Can AI Help in Improving Online Discourse at scale?

What: Investigating the efficacy of AI powered chat interventions to improve online interactions across various cultures, countries, and languages.



Why Study Culturally-Aware Interventions?

We are aware of how LLMs can help in promoting polite, constructive online discourse.

However, current research predominantly focuses on **English and US-centric topics**.

Gun violence in the US

Why?

Cultural
variances in
discourse

Different ways of providing validations and responses across cultures.

Original Work: Dominated by white participants

Changes in
translation and
language processing

All languages have distinct features and loss of linguistic nuances.

Original Work: Uses only English

Temporal
nature of
topics

Topics and context changes over time.

Original Work: Single Topic

How we will conduct the study



Diverse Topics: Analyze topics from each continent and in different languages.



Cultural Representation: Include participants from varied cultural backgrounds.



Temporal Relevance: Use current and old news topics to ensure relevance.



Multilingual Discourse: Experiment with AI-assisted translation in discussions between speakers of different languages.

Expected outcomes

Outcome Goals:

- Develop **guidelines for culturally-aware discourse interventions** using LLMs to enhance global communication effectiveness.
- A systematic study that lists **the good, the bad and the ugly consequences of using AI** in culturally diverse settings.
- Direction to create an AI powered chat moderator that can **actually be used at scale**.

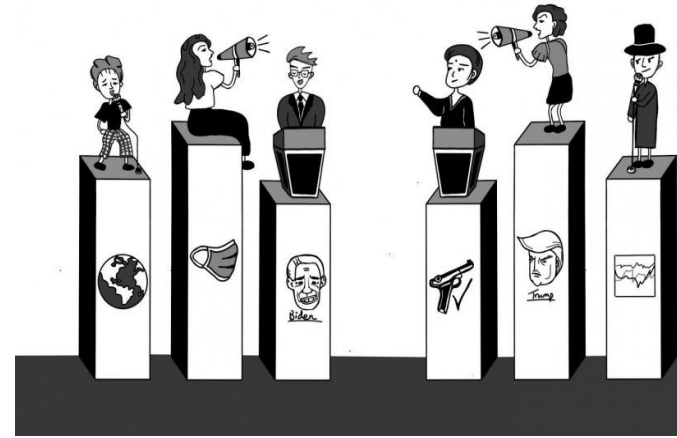
Industry Practitioner

We Introduce: Harmony AI: AI chat assistant for online political conversations

- Improves conversation quality, fosters democratic reciprocity
- Uses LLM (similar to GPT-3) for real-time, evidence-based suggestions. Embedded in social media, forums, and messaging apps

Why?

- Online discourse can be divisive, unconstructive conversations harm social cooperation and traditional moderation isn't scalable for large volumes of online interactions.



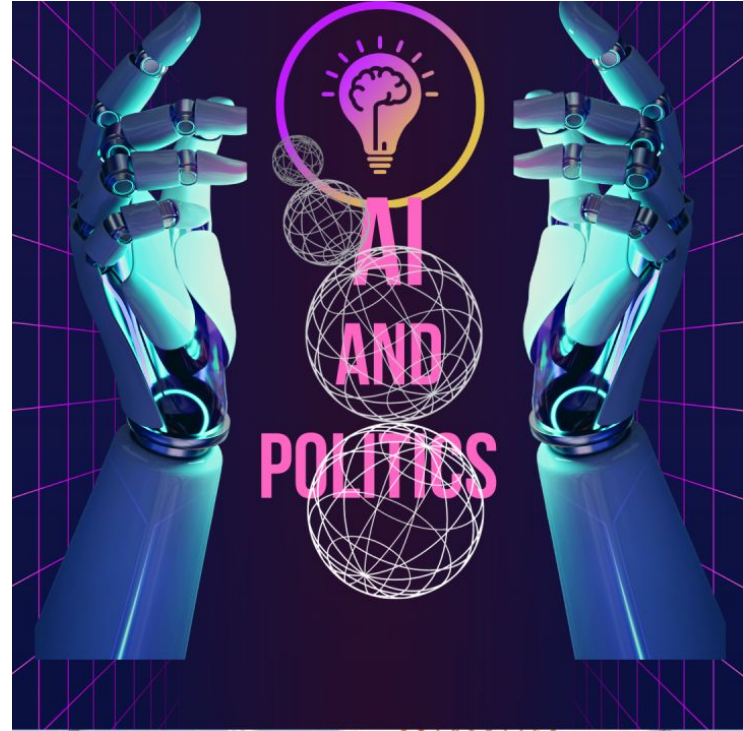
Positive Reasons


- Enhanced Discourse: More respectful, productive conversations
- Scalability: Reach millions of users simultaneously
- User Empowerment: Promotes mutual understanding, reduces polarization




Potential Drawbacks

- Algorithmic Bias: Risk of reinforcing existing biases; requires monitoring
- Privacy Concerns: Ensure data privacy and transparency
- Risk of exploitation: Hackers controlling the bot to direct their own narratives (Big brother)





*Paper 2:
Human-AI collaboration
enables more empathic
conversations in text-based
peer-to-peer mental health
support (Sharma et al.,
2023)*



Summary: Human-AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support (Sharma et al., 2023)

Objective: The study investigates how AI can assist human supporters in providing more empathic responses in online mental health support conversations.

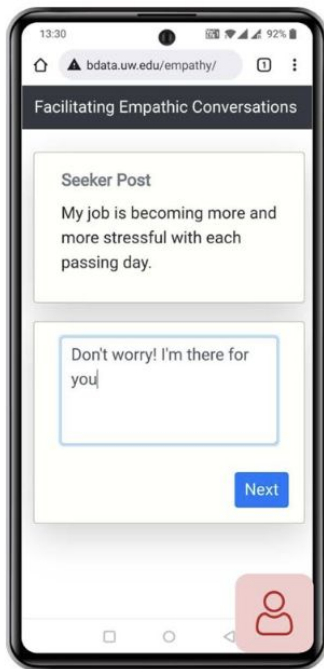
HAILEY SYSTEM: An AI-in-the-loop system named HAILEY was developed to provide real-time feedback to peer supporters, enhancing the empathy in their responses. This feedback includes suggestions to insert or replace text to increase empathy.

Methodology: A randomized controlled trial was conducted with 300 participants from the TalkLife platform, divided into two groups: one receiving AI feedback (human+AI group) and the other not (human-only group). Both groups received basic empathy training.

Human-only
(control) group



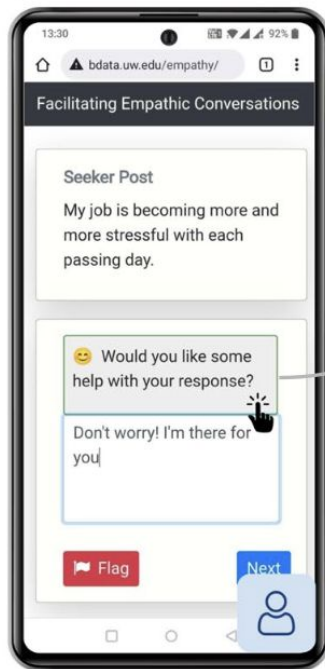
a



Human + AI
(treatment) group



b

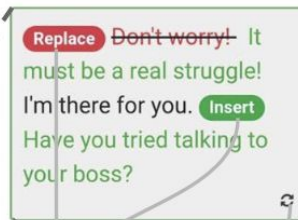


c



Feedback prompts

Feedback



Actions to edit
response

Reload feedback
if required

Summary: Human-AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support (Sharma et al., 2023)

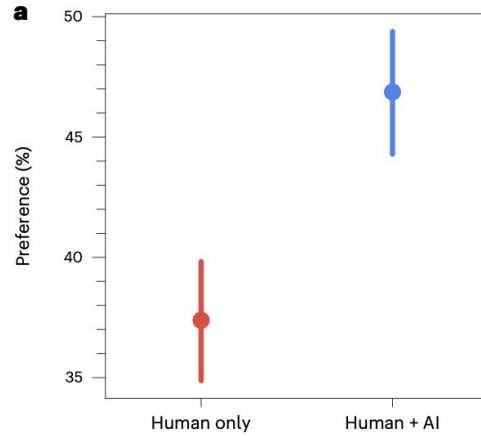
Results:

- **Increased Empathy:** The human+AI group showed a 19.6% increase in conversational empathy compared to the human-only group.
- **Higher Gains for Struggling Participants:** Those who reported difficulties in providing support saw a 38.9% increase in empathy with AI assistance.
- **Use Patterns:** Participants used AI feedback both directly and indirectly, incorporating suggestions or using them to generate their own empathic responses.

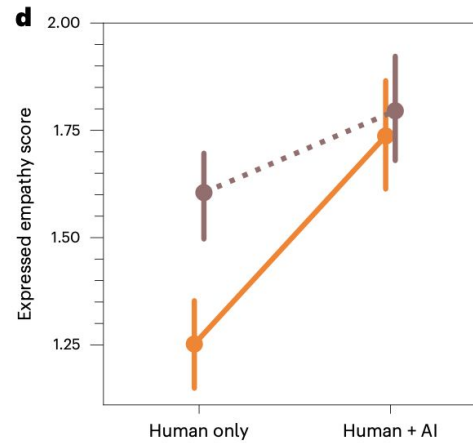
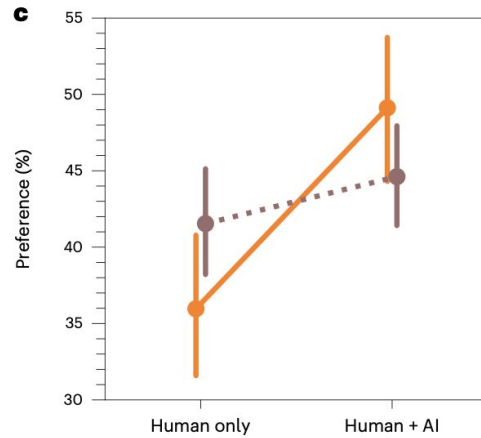
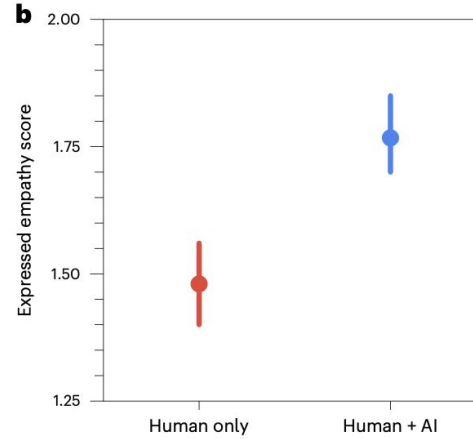
Important takeaways:

- AI-assisted systems can enhance human performance in empathetic communication, particularly in settings requiring high empathy levels.
- Challenges remain in developing human-centered AI models and user-friendly interfaces.

Human evaluation:
which response is more empathic?



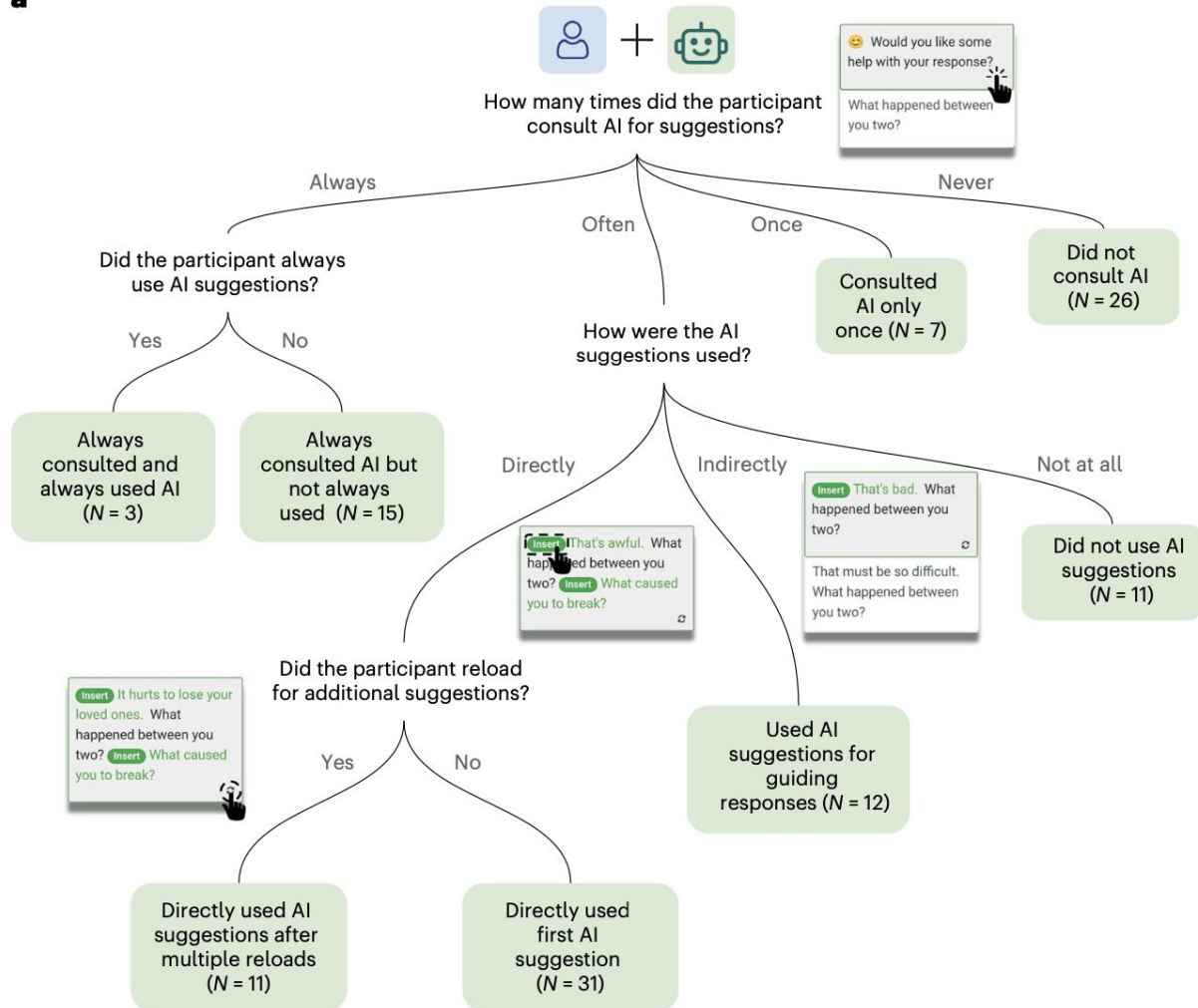
Automatic/AI-based evaluation:
expressed empathy score



— Writing responses was challenging (N = 36)
- - - Writing responses was not challenging (N = 54)

— Writing responses was challenging (N = 91)
- - - Writing responses was not challenging (N = 142)

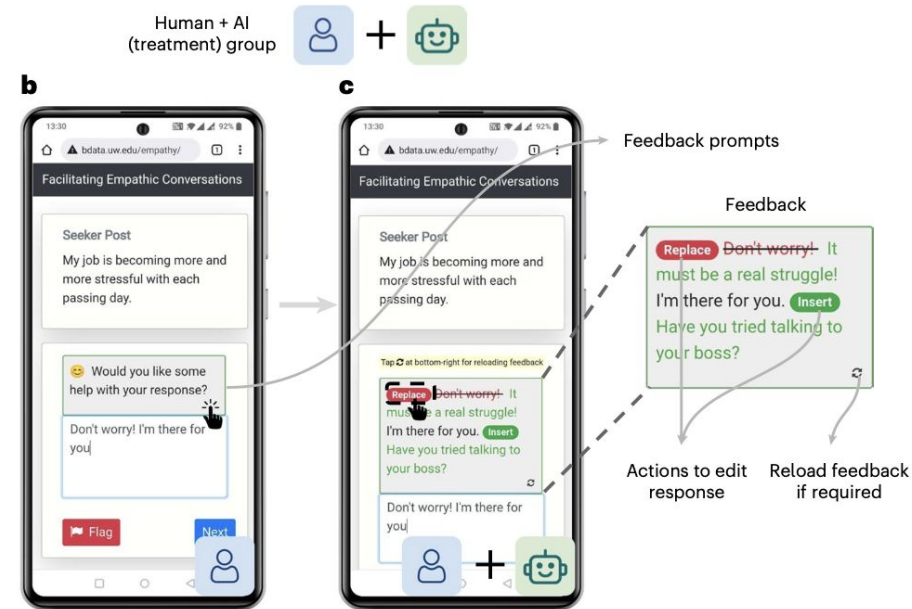
a



Peer Reviewer Comments

Pros:

1. Importance of the research question
 - The supply-demand mismatch between patients and specialists
 - The need for automated and streamlined training for peer supporters
2. Study design (Human only vs. AI-in-the-loop)
 - A more collaborative, dynamic, and less intrusive interaction between AI and human vs. human-in-the-loop vs. Human first and AI second



Peer Reviewer Comments

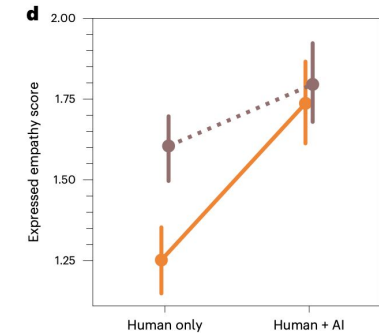
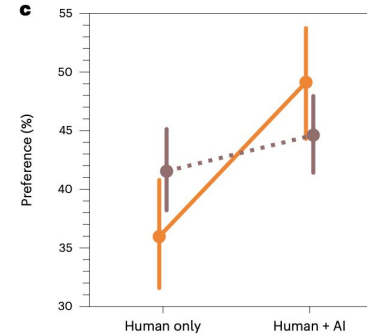
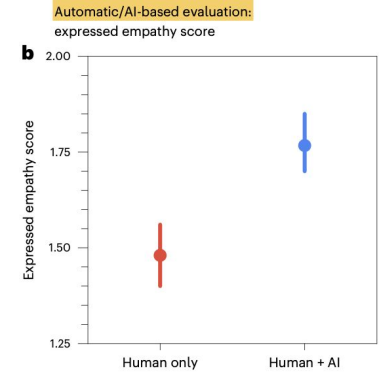
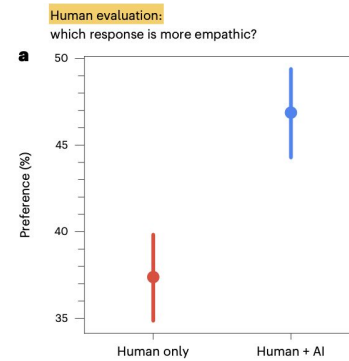
Pros:

3. Robustness check of the empathy effect

- Independent human coders
- AI coders

4. Pinpoint which subpopulation the effect worked on (challenging vs. not challenging)

5. Dataset (a large-scale field study)



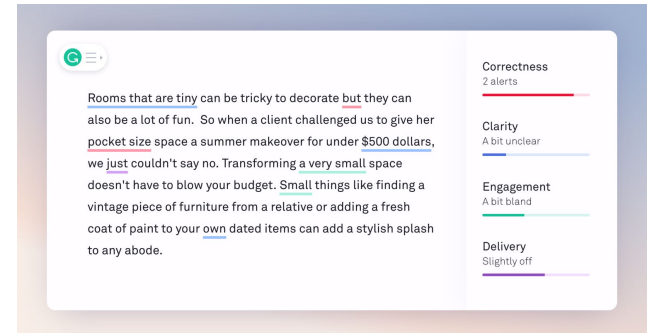
— Writing responses was challenging (N = 36)
- - - Writing responses was not challenging (N = 54)

— Writing responses was challenging (N = 91)
- - - Writing responses was not challenging (N = 142)

Peer Reviewer Comments

Cons:

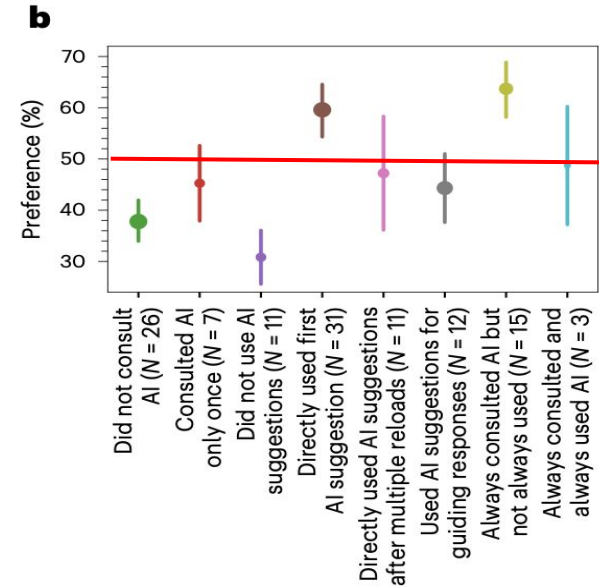
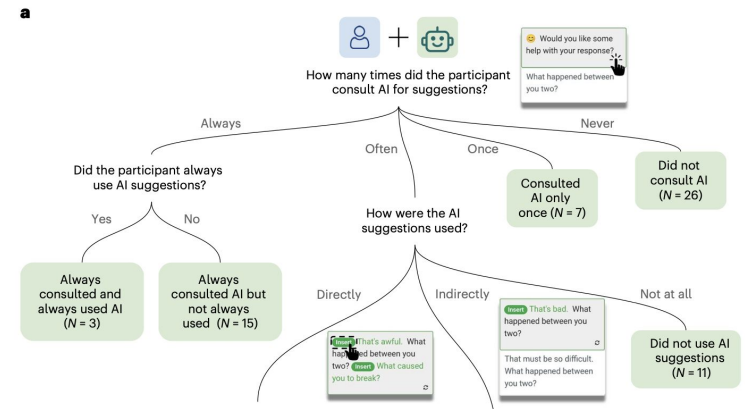
1. How important is empathy in mental health treatment?
 - The study did not test the actual downstream consequences.
 - Are there some other more important attributing factors for the treatment?
2. Practical contribution: How is this different from the other AI tools?
 - HAILEY can increase conversational empathy, but what about other AI tools (e.g., ChatGPT, Grammarly)? How much does HAILEY add to the AI market?
 - HAILEY vs. ChatGPT (prompt)
 - HAILEY vs. Gemini (prompt)
 - HAILEY vs. Grammarly (AI-in-the-loop)



Peer Reviewer Comments

Cons:

3. A quasi-randomized experiment, not a pure randomized study - **selection effect**
4. The preference effect is not very robust among participants
5. Who's evaluation of perceived empathy really matters? - support seekers



Discussion Starters

User Autonomy vs. AI Assistance: How can we strike a balance between providing helpful AI feedback and maintaining the autonomy and authenticity of the human supporter's responses?

Ethical Considerations: What ethical challenges arise when integrating AI into sensitive areas like mental health support? How can we ensure these systems are used responsibly?

Scalability and Accessibility: What are the potential benefits and drawbacks of scaling AI-assisted mental health support to underserved communities or regions with limited access to mental health services?

Academic Researcher

Title: **Simulating Support Seekers Using LLMs for realistic multi-turn feedback**

What: Sharma et. al provided existing posts and asked users to write empathic responses. We propose to extend the study to a **conversational setting with support seekers being simulated using LLMs.**



AI Support
Seeker



Human Support
Provider

Why simulate support seeker?



Risks: Doing a study with real humans who need support has several risks.



Simulations: LLMs can simulate human well to a certain extent.



Feedback: Can help simulate the real setting where support provider is also getting feedback on their response from the support seeker



Diverse supports: Different support seekers have different needs, some want empathy, some seek solutions to their problems.

Expected Outcomes

Skill Training: advanced training modules that help support providers refine their skills in empathy-driven conversations.

Insights Dashboard: Additionally, it will provide valuable insights into enhancing human-AI collaboration in sensitive communication tasks.

Broader Implications: Other domains requiring sophisticated emotional interactions, such as customer service or therapeutic environments.

Future Works: Move to speech to understand tone and other nuances.

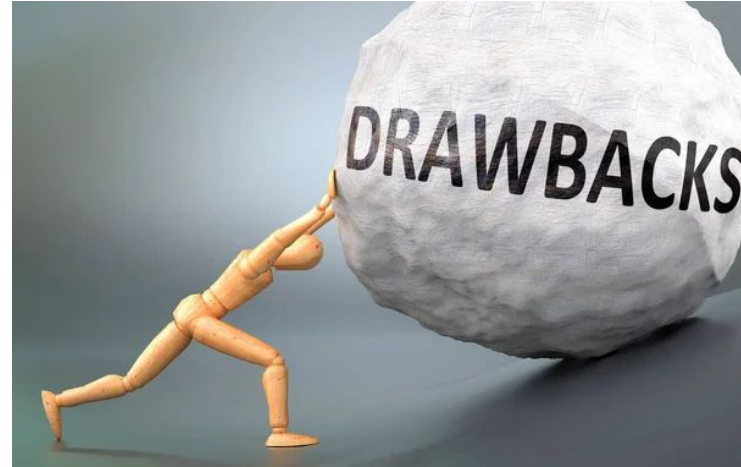
Positive Reasons

- Increased engagement, 70% drop out of mental health care now due to poor engagement; HAILEY can improve patient retention (National Alliance on Mental Illness, 2024)
- Standardizing empathetic communication reduces treatment disparities
- Can Increase productivity and burnout
- Treat more people and provide greater access to mental healthcare



Potential Drawbacks

- Risk of Peer supporters being over reliant on AI
- Unsure if this will generalize to realistic contexts and to diverse contexts (e.g., elderly)
- Need to ensure HAILEY is used as a supportive tool, not a replacement for human judgment
- Implement robust data protection measures to maintain user trust and compliance



High-level themes

1. **Enhancing Human Interaction:** Both papers focus on how AI can improve human interactions, whether in mental health support or political discourse.
2. **AI as a Support Tool:** AI is seen as a tool to *assist rather than replace* human effort, providing enhancements to natural human capabilities.
3. **Real-Time Assistance:** Both studies highlight the importance of real-time AI feedback in improving the quality of interactions.
4. **Ethical AI Use:** Emphasis on the ethical application of AI, ensuring it supports positive outcomes *without undue influence or manipulation*.
5. **User Empowerment:** Both studies show that AI can empower users by making complex tasks more manageable, whether it's providing empathetic support or engaging in constructive political conversations.