

MAS.S68, Spring 2023

2/15

Generative AI for Constructive Communication

Evaluation and New Research Methods



Agenda

Rishi Bommasani

[Zoom](#) talk

Q&A after his talk

[attendance note]

Second half of class:

Critiques

Break into groups (5 min)

Share critiques within groups (20 minutes)

Share out and discuss with group (25 minutes)

Logistics notes (5 min)



Critiques

Groups

Search

- Bing
- Perplexity
- Siri

Research Tools

- Elicit
- Semantic Scholar
- ResearchRabbit.ai

Writing aids

- Grammarly
- Wordtune

Characters

- Replika
- Inworld

Critiques

Instructions

Break into groups

Rotate giving your 3-5 minute presentation (everyone in group should go)

Prepare to share out:

- What were the identified **benefits** of your app category?
- What were identified as the biggest **risks**?
- Does your group have any **proposals** for research or interventions to mitigate these risks?

Upcoming

Workshops with Doug

- **Workshop 1 (Feb 16 @ 5pm):** Students will build an interactive dashboard that uses the OpenAI GPT-3 API to explore their project idea. We'll go over how to make requests, how to interpret the API results, how to set the parameters, share basic prompting tricks, etc. We'll also go over Streamlit, a frontend framework that makes it easy to build data-sharing apps; and we'll touch on how to do evaluation with sklearn.
- **Workshop 2 (Feb 22 @ 1pm):** We'll go over the content from Workshop 1 more rapidly (10 minutes), and then dive into more advanced stuff including fine-tuning, embeddings, the OpenAI edit API, and semantic search. Depending on the needs and interests of the attendees we may also discuss how to run generative models locally on your computer using the HuggingFace library.

Next class:

- **Michiel Bakker** (DeepMind) in person!
- Second hour: project pitches

Logistics

Logistics announcements:

- Update to grading scheme
 - 25% homeworks (points total will be put online soon)
 - 25% in person participation (contributing to discussions, coming to class, sharing work)
 - 50% final project (including mid-point check-in and final writeup)

Homework for next week:

- Questions for Michiel Bakker
- Research project proposal!
 - We will do lightning project pitches next week
 - API Credits

Other:

- Join our Slack and contribute papers you find interesting or share what you're reading!
- Office hours: 1-2:30 on Fridays in CCC
- Attendance reminder