## Phase 1

**RQ1:** Do perceptions of constructiveness vary based on argumentation styles?

Online thread with homophobic or Islamophobic comments

2 cultures {India, US} X 2 issues {homophobia, Islamophobia} X 2 threads X 2 stances {against, support} X 2 argumentation styles {logical, dialectical} X 2 comments

= 64 LLM-generated **constructive** comments

Pairs of LLM-generated Logical vs
Dialectical comments

Reviewed by
Indian and American
participants

- LLMs view **dialectical arguments as more constructive** than logical ones, even more so than humans do.
- Potential misalignment between how humans and LLMs characterize constructive comments.

## Phase 2

**RQ2:** Can LLMs help people write constructive comments on divisive social issues?

Online thread with homophobic or Islamophobic comments

Control: Humanwritten constructive comments

Quantitative and qualitative evaluation of LLM-generated vs Human-Written vs Human-AI co-written comments

- LLM-generated comments are significantly **more constructive** than human-written comments.
- In the Test group, LLM's suggestions made people's comments significantly more constructive, more positive, and less toxic.
- Although LLM captured people's points well without homogenizing writing, it often misrepresented people's stances on divisive issues.