Botender: Supporting Communities in Collaboratively Designing Community AI Agents through Case-Grounded Prototyping

REU: Sophia Liu[1], Mentors: Tzu-Sheng Kuo[2], Kenneth Holstein[2], Haiyi Zhu[2]
[1] UC Berkeley, [2] Carnegie Mellon University

Human-Computer Interaction Institute

This material is based upon work supported by the National Science Foundation under Award No. 2349558.

MOTIVATION

AI Agents for Online Communities: **AI agents**, often called ***bots***, play important roles in online communities—from **server management** to **content moderation.** Yet **most bots are built by outsiders**, with limited ability to **collaborate with communities** and understand their norms. As a result, these agents often **fail to reflect community values, goals, and nuances.**

LLMs in Agent Design: LLMs make AI agent design more accessible—but **key challenges** must be overcome to enable meaningful community involvement:

✍ Shallow Prompting Practices

Stakeholders **without prior LLM prompting experience** tend to focus on **single-turn interactions** when designing bots, overlooking broader consequences across diverse scenarios.

Zamfirescu-Pereira et al. CHI'23

🧠 Conflicting Mental Models

Community members often hold **divergent views** on what constitutes appropriate agent behavior—and how to realize it—complicating **collective norm-setting** and enforcement.

Sorensen et al. ICML'24

🛠 Fragmented Design Workflows

Existing tools are often **disconnected from the platforms** where bots are deployed, making it difficult for communities to **iteratively design, test, and revise** agent behavior **in context.**

Kuo et al. CHI'24; Subramonyam et al. CHI'25

RESEARCH QUESTION

*How might we **empower online communities** to **collaboratively design** their **community AI agents**?*

slack Discord reddit

Botender facilitates collaboration by giving communities concrete tools to collaboratively shape how their bots behave.

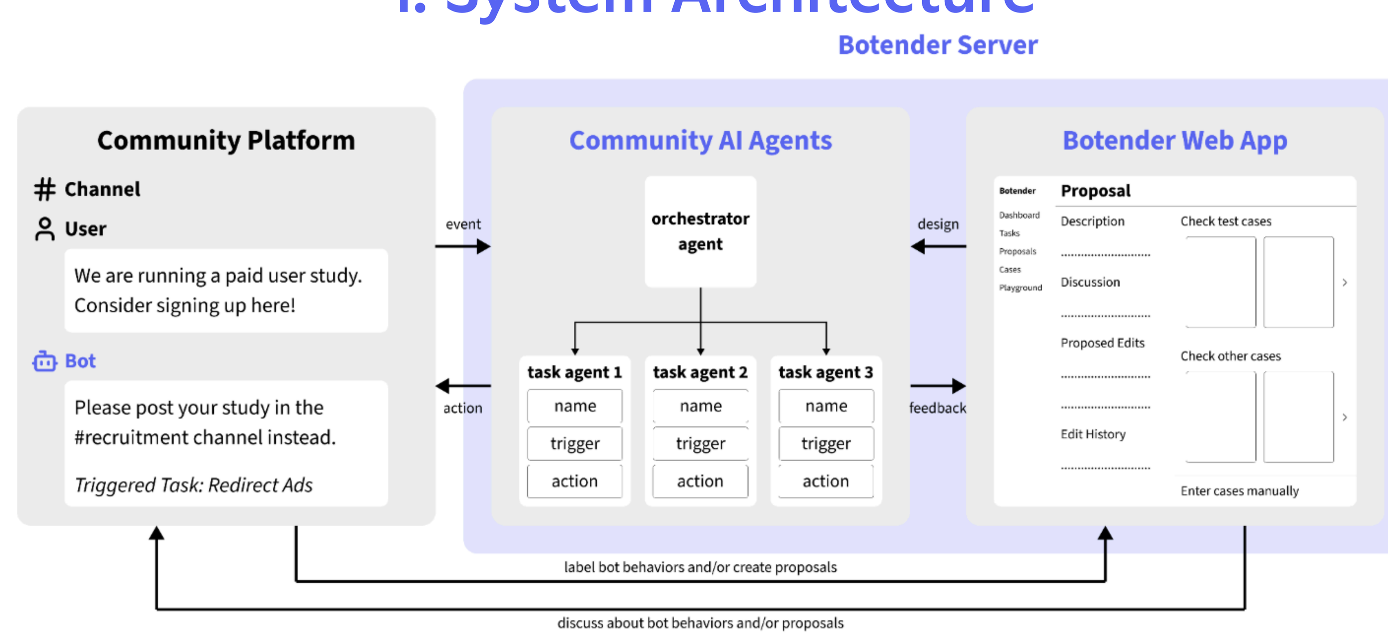
METHODS

Botender consists of:

i. The **System**, a Discord plug-in and a web-based interface for editing, testing, and discussing bot behavior.

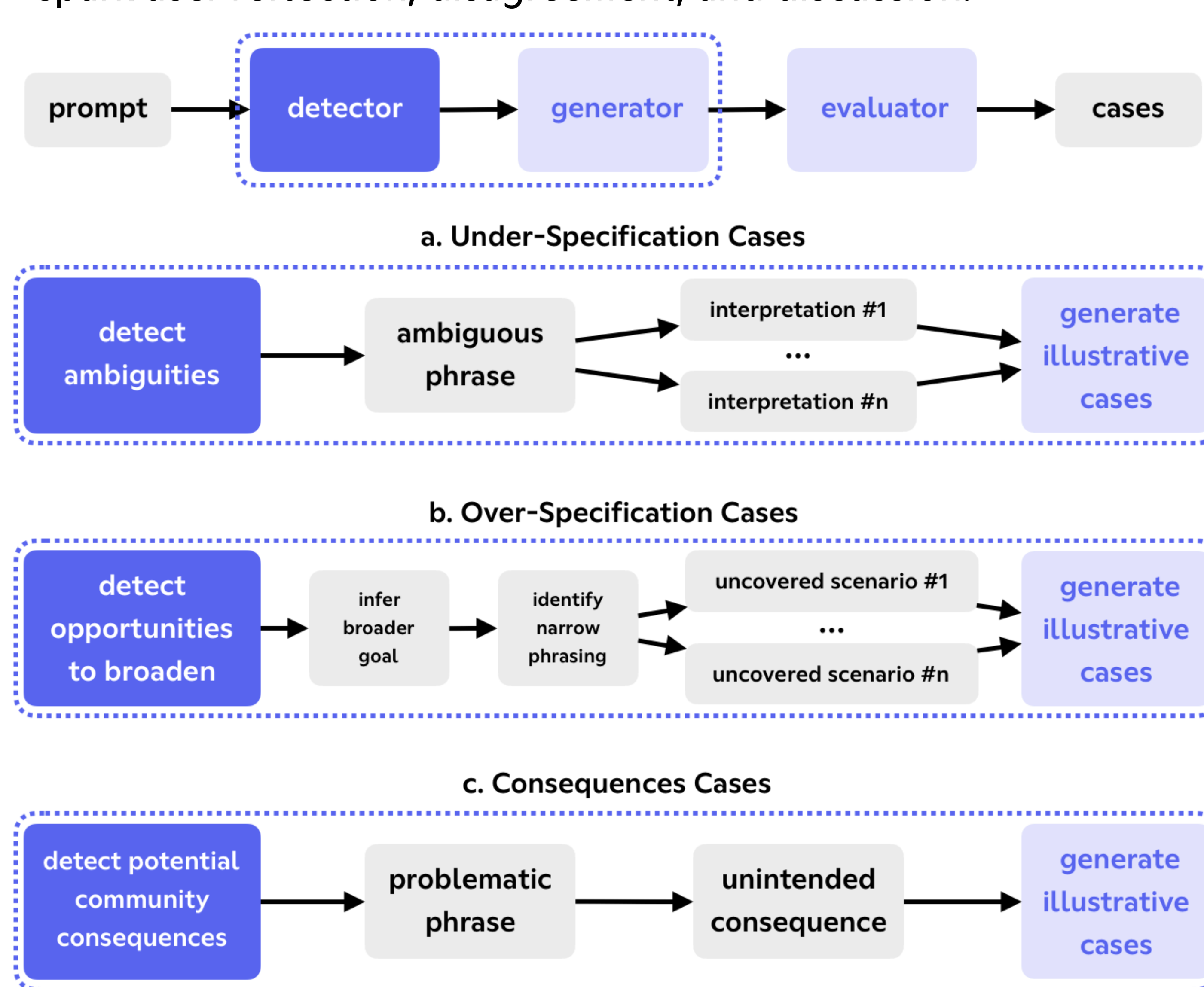
ii. A **Case Generation Algorithm** that surfaces design opportunities by encouraging reflection and facilitating community-level discussion on prompt behavior.

i. System Architecture



ii. Case Generation Algorithm

Three pipelines detect different types of **prompt flaws**—under-specification, over-specification, or unintended community consequences—and **generate illustrative cases** designed to spark user reflection, disagreement, and discussion.



METHODS (cont.)

Pipeline-Specific Evaluators: Assess whether each case **clearly demonstrates the specific flaw** it was meant to reveal—under-specification, over-specification, or an unintended community consequence.

General Evaluator: Rates from 1-5 how well a case **surfaces a meaningful issue** in the prompt that is likely to **provoke individual reflection** or **community discussion.**

ONGOING WORK

Study 1: Algorithm Evaluation

Goal: Understand how well the case-based provocations...

- support users in **identifying opportunities** to iterate on a bot's prompt
- help reveal **differences in perspective** across multiple users

Procedure: Online survey

Study 2: System Evaluation

Goal: Understand the **overall user experience** and assess whether Botender effectively facilitates **collaborative community agent design**

Procedure: Multi-day field study

TAKEAWAYS & FUTURE WORK

💡 Early Insight

Preliminary use suggests cases help users spot **prompt flaws** and **community value tensions.**

🤝 Design Principle

Botender supports **meaningful participation** from **non-experts** in agent design.

🔎 In Progress

Ongoing user studies will assess how Botender impacts **prompt iteration** and **value alignment.**

ACKNOWLEDGEMENTS

We are grateful to our collaborators Quan Ze Chen (AI & Democracy Foundation), Amy X. Zhang (University of Washington), and Joseph Seering (KAIST).

This project is generously supported by CMU's Block Center for Technology and Society, the K&L Gates Initiative in Ethics and Computational Technologies, and the NIST–CMU AI Measurement Science & Engineering (AIMSEC) Center.