

Reddit Conversation Laboratory: Field experiments with conversational AI agents

Jeremy Foote

Brian Lamb School of Communication, Purdue University, USA

Loizos Bitsikokos

Brian Lamb School of Communication, Purdue University, USA

Hitesh Goel

Department of Computer Science and Engineering, International Institute of Information Technology Hyderabad, India

Deepak Kumar

Department of Computer Science, University of California San Diego, USA

Abstract

LLM-based generative AI agents are the first autonomous technologies that can act as true conversational partners. Communication researchers and others have already begun to explore the influence of AI conversations on their interlocutors. We present a methodological framework and software tool for conducting field experiments with AI agents on Reddit. The Reddit Conversational Laboratory is Python-based software that identifies potential participants, messages and consents them, and conducts conversational experiments using researcher-designed AI chatbots. In addition to storing all conversations, the software can also record participant behavior before and after conversations. In this paper, we outline design principles, best practices for using the tool, and possibilities for future extensions.

Keywords: Generative AI, chatbots, Reddit, research software, field experiments

Introduction

Generative AI (GAI) chatbots have shown incredible promise as conversational partners and are being used across multiple domains, from general knowledge queries to mental health interventions to companionship (e.g., Brandtzaeg et al., 2022; Lawrence et al., 2024). As chatbots continue to be integrated into more and more aspects of our lives, communication scholars should be at the forefront of research into understanding their influence.

One important question is the impact that conversations with chatbots have on their conversational partners. For example, recent research suggests that chatbots can influence political attitudes (Argyle et al., 2025; Hackenburg et al., 2025; Lin et al., 2025), attitudes about social issues (Salvi et al., 2024), and even reduce belief in conspiracy theories (Costello et al., 2024). However, it is often difficult and expensive to study the long-term impact of any intervention, including conversations with chatbots. We argue that Reddit is a promising context for doing large-scale experimental work with generative AI bots. The Reddit API provides rich, longitudinal, cross-community data about participants (Foote, 2022), allowing researchers to causally identify changes in behavior due to interactions with bots.

In this paper, we introduce RCL—the Reddit Conversation Laboratory. RCL is open-source Python software designed to run automated conversational AI experiments with Reddit users. Using the Reddit messaging system, the software recruits participants, obtains their consent, and randomly assigns them to a control group or one of a set of configurable chatbots. It then manages and stores the conversations that the chatbots have with participants. Finally, an additional set of scripts tracks participants’ activity before and after their conversations.

We begin with an overview of approaches to studying human-AI interaction and introduce the design of RCL. We then explain how to install and run the system, with tips and best practices. We end with a discussion of ethical and practical considerations for doing research with our system, limitations of the software and this approach, and where future work might focus.

Studying human-AI interactions

As LLM-based chatbots have grown in capability to become conversational partners, they have attracted rapidly increasing interest from communication scholars. The data for this kind of research can be gathered in a variety of ways, each with their own benefits and drawbacks.

First, scholars have engaged in post-hoc methods, like interviews and surveys, to ask people about their experiences with and perceptions of chatbots. For example, Brandtzaeg et al. (2022) interviewed 19 people who had friendships with social chatbots, comparing their perceptions of human-AI and human-human friendships. Other researchers have used surveys to understand the motivations and psychological traits of those who use conversational AI tools (e.g., Hu et al., 2023; Skjuve et al., 2024). Related observational research has looked at how people talk about GAI bots in online communities or social media (e.g., Li & Zhang, 2024). These approaches

can allow for rich, thick understandings of how people perceive their communication with generative AI chatbots. However, they also have multiple drawbacks. First, they don't typically have access to the actual conversations that people have with chatbots. They also don't allow for a longitudinal understanding of the impact of GAI conversations over time, nor do they allow researchers to provide *causal* estimates of how these conversations influence participants.

Second, social scientists have done lab experiments to directly explore how people interact with GAI. Many of these focus on the use of GAI as a tool, studying the ability of GAI to do things like craft persuasive messages (Goldstein et al., 2024; Lim & Schmälzle, 2023). Another strand of experiments looks at the role of generative AI as a conversational partner. Here again, the focus has often been on persuasion, with researchers showing that AI agents can successfully change the attitudes or beliefs of their human conversational partners (e.g., Argyle, 2025; Costello et al., 2024; Hackenburg et al., 2025; Salvi et al., 2024). These approaches allow researchers to answer causal questions, but suffer from the same limitations as most lab experiments: for example, they are often undertaken with a sample that may not be representative of the population of interest, and the conditions may not be similar to those in the “real world” (Gerber & Green, 2012). In addition, they are typically cross-sectional: participant beliefs and behaviors are only measured at the same time that the intervention is given, as further follow-up is expensive and difficult.

Online field experiments

In field experiments, researchers randomize some aspect of how treatments are assigned to subjects in a naturalistic setting (Baldassarri & Abascal, 2017). Some behavior in the “real world” is then observed. In theory, field experiments have many of the benefits of lab experiments; in particular, using randomization eliminates the risk of unobservable variables to plausibly identify causal effects. In addition, they have greater external validity, as people are behaving in natural settings, often without even knowing they are being observed (Gerber & Green, 2012). Of course, there are also trade-offs. Compared to lab experiments, field experiments often suffer from problems of differing rates of compliance, greater self-selection to different conditions, and interference between units (Gerber & Green, 2012).

Online field experiments—that is, field experiments which use online platforms as their context—have become increasingly common. For example, Munger (2017) identified Twitter users who engaged in racist harassment,

and randomly assigned them to be sanctioned by either an in-group or out-group bot, with varying numbers of followers. He then tracked the users' behavior for two months after the intervention, finding that sanctions from high-follower in-group members reduced the use of racial slurs on the platform. Salganik (2017) argues that—when carefully designed—digital field experiments have a number of advantages. First, they can benefit from low or non-existent marginal costs: it is often no more expensive or difficult to run an experiment with 1,000 or 10,000 participants than an experiment with 100 participants. Second, online experiments often benefit from “always-on” data collection, allowing researchers to gather pre-intervention and post-intervention behavior, as well as to use pre-intervention covariates to increase statistical precision or identify heterogeneous treatment effects. Finally, online experiments can reduce some of the risks of typical field experiments; for example, treatment assignment can be enforced via software, reducing opportunities for subjects to change conditions. On the other hand, some risks to validity or generalizability can be greater for online experiments; for example, online communication networks make it easy for subjects to talk about the intervention with others, possibly influencing each other's responses (Muisse & Pan, 2019).

Reddit Conversation Laboratory

Reddit Conversation Laboratory (RCL) is designed to bring the benefits of online field experiments to the study of human-AI interactions, enabling researchers to gather granular, longitudinal, real-world behavioral data to answer rich causal questions about the impacts of generative AI conversations. The field site for RCL is Reddit: the largest platform for online communities, with tens of millions of active users participating in hundreds of thousands of individual “subreddits” about many diverse topics. Reddit is also one of the most open web platforms, with relatively permissive APIs which allow researchers to gather rich retrospective data about participants' communicative actions (comments and posts) before experiments as well as to track their actions afterward. This type of data enables multiple methodological approaches, such as qualitative or computational text analyses, network analyses, or time-series analyses of communication change over time.

To use RCL, researchers use previous activity on Reddit to identify potential participants. They then design a set of GAI chatbots which can be randomly assigned as conversational partners for these users, who chat with the bots using the Reddit platform. Subsequent behavior can then be

compared to a control group which does not receive any intervention.

There are many questions that can be answered under this framework, primarily around the efficacy of different persuasive approaches. For example, health communication researchers might test different informational interventions for people expressing certain health concerns online, while political communication researchers might target people who belong to extremist political subreddits, and organizational communication scholars might explore the role of AI agents in explaining and enforcing rules and norms. They might then look for causal evidence of changes in the terms that people use or which communities they participate in after their conversations.

Our approach has a number of benefits over alternative ways to study human-AI interactions. Unlike surveys or interviews, researchers have access to the conversations that participants have with chatbots. Unlike lab experiments, researchers have access to rich, longitudinal behavioral data both before and after the interventions. In addition, experiments can be performed in situ, with greater promise of identifying generalizable results. Finally, many experiments may represent actual interventions that could be rolled out more broadly using RCL if they are successful.

Intended purpose and design principles

The primary components of RCL are designed to 1) identify potential experimental participants, invite them to participate, and obtain consent; 2) interface between GAI providers (e.g., OpenAI, Anthropic, or Gemini) and the Reddit messaging system to conduct and store participant-chatbot conversations; and 3) track the pre-intervention and post-intervention behavior of participants.

RCL is designed to be *modular*: that is, the components operate as independently as possible, to make it easier to adjust or replace one component or another. For example, the output of the participant identification component is a simple CSV file of usernames for the system to contact. While we provide a few example functions which use the Reddit API to identify participants, researchers can use whatever process they would like to create that CSV file. This includes modifying our code, writing bespoke code, or even manually curating a list.

The software is also designed to be *resilient*. Because it is connecting multiple external services (e.g., Reddit and the GAI provider), it is not uncommon for some component to fail temporarily. RCL includes simple failure detection and checkpointing logic, and uses the cron task manager to re-

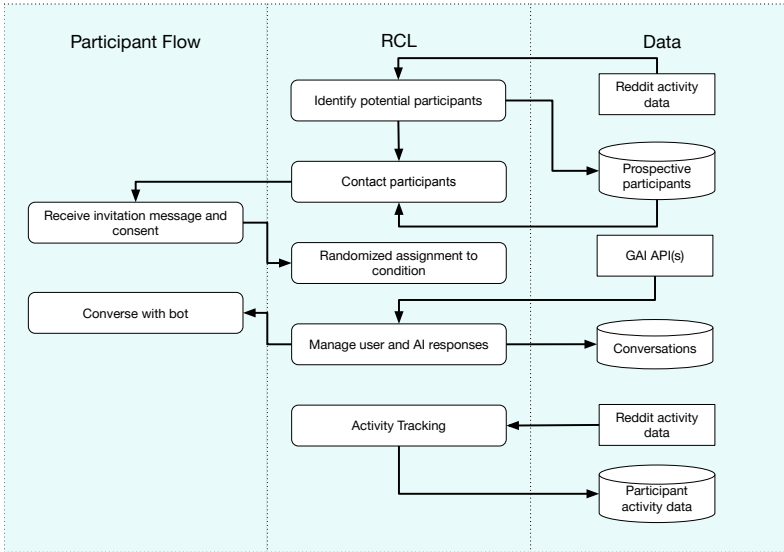


Figure 1: Flow diagram showing the primary components and actions of the system. Squares represent external data accessed through APIs and cylinders represent data stored on the server.

peatedly run a set of Python scripts. This architecture allows the software to run autonomously and automatically resume after disruptions.

Finally, RCL is designed to be *extensible* and minimal. The current version does not have a graphical user interface or an API. It is designed for researchers who have experience in Python but who may not have experience working with Reddit and/or working with GAI chatbots. It exemplifies best practices and gives a framework that can be easily built upon and expanded.

Using Reddit Conversation Laboratory

Install and setup

RCL should be run from a UNIX-based server. Detailed installation instructions and hardware requirements are outlined on the GitHub project page;¹ both compute and storage requirements are quite modest. As discussed in more detail below, in an initial study using this software (Foote et al., 2026), chatbots reached out to as many as a few hundred participants per day and on some days engaged in ongoing conversations with more than 20 users.

¹<https://github.com/jdfoote/Reddit-Conversation-Laboratory>

Based on this experience, we estimate that four cores and 16 GB of memory should be more than sufficient for even fairly large experiments.

Setup begins with cloning the git repository, which includes a set of Python scripts and directories for storing data. Before running the scripts, researchers will need to create a new account on Reddit for the bot which will be used to contact users, and to create an app for the account on the Reddit developer site.² They will also need to create developer accounts with whichever GAI provider(s) the researchers would like to use. Authentication information and keys are stored in a YAML configuration file. As explained in more detail below, we suggest working with moderators and sending invitation messages through modmail. To test that modmail is set up correctly, we recommend that researchers use a second Reddit account to create a new subreddit, and add the bot account as a moderator for that subreddit.

Participant identification

The participant identification step involves creating a CSV file of usernames that will be invited to talk with the chatbot. Researchers establish a set of criteria to select potential participants, and write a function to collect candidate usernames using the Reddit API. These criteria can be fairly complex: for example, a political communication project might identify all of the users who have participated recently in a set of extreme communities. Then, the researchers might filter to only those who post at least five comments per week or those who have commented about a certain topic. They could also write code to enact stratified random sampling, to ensure, for example, that participants from different subreddits are equally represented in the sample. RCL provides two example functions: the first identifies all users who have recently used any of a set of keywords in a list of subreddits. The second is more complicated, and requires that the bot be given moderator access: it retrieves removed comments from moderator logs, retrieves toxicity scores from Jigsaw's Perspective API³ (Lees et al., 2022), and creates a list of users who have recently posted toxic content across a list of subreddits. Researchers can build on these examples, write their own code to generate prospective participants, or even generate this list another way (e.g., survey respondents from MTurk or Prolific).

Because this step only produces a CSV and doesn't actually contact users,

²<https://www.reddit.com/prefs/apps>

³In order to use the Perspective API, researchers will need to obtain an API Key and store it in the config file.

researchers can run and re-run the participant identification script and do black box testing on the output, ensuring that the correct accounts are identified and added to the `to_contact.csv` file.

Chatbot design and testing

RCL will use this list to invite users to have a conversation with a randomly selected GAI chatbot, so the next step is to design and test the different bots that will chat with participants. These bots are instructed via a researcher-designed prompt which is embedded in all conversations. Prompt engineering is a nascent but rapidly formalizing field, increasingly recognized as a structured methodology for optimizing LLM behavior through the careful design, evaluation, and iterative refinement of prompts (Velásquez-Henao et al., 2023). Best practices in prompt engineering emphasize clarifying the model's intent, assigning specific roles, and incorporating exemplars (Schulhoff et al., 2025). Researchers in this area emphasize the need for iterative cycles of formulating, testing, and revising prompts to achieve desired outcomes (Desmond & Brachman, 2024). This testing is most easily done outside of Reddit; we include a very simple command-line utility that allows for conversation testing. Researchers should engage in multiple conversations with bots, tweaking the prompts until the bots reliably produce the expected behavior. Of course, some aspects of design are impossible to test beforehand; participants engage with bots in unexpected and unstructured ways, which shapes conversational outcomes and bot behaviors.

Once some candidate prompts have been identified, researchers can edit the `bot_conditions` variable in the config file. This variable is a dictionary, where the key is the name of the condition that will be stored in the participant table, and the value is the prompt that the GAI provider will use to instruct the chatbot. Researchers should also edit the `GAI_provider` variable, which stores the provider and model name that chatbots will use. For both `bot_conditions` and `GAI_provider`, if the dictionary contains multiple items, then RCL will randomly choose a prompt and/or model to use with each participant, and store which choice was made. Listing 1 shows the structure of the configuration file.

Conversations on Reddit

The general conversational flow for RCL is that potential users will first receive the `initial_message` to recruit them. This initial message should include basic information about the study, and possibly a link to additional

```

max_interactions : 50
GAI_models:
  openai_models :
    - 'gpt-3.5-turbo'
    - 'gpt-4-1106-preview'

conversations_file : '../data/conversations.csv'
to_contact_file : '../data/to_contact.csv'
participants_file : '../data/participants.csv'
subreddits_file : '../data/subreddit_rules.csv'

goodbye_message : "Thanks for chatting with me..."
initial_message :
  default: "Hello, I am a bot designed by..."

clarifying_message : "Please reply with 'yes' if you would like to chat..."
handoff_message : "Thank you for your interest..."

first_consented_message :
  first_message_1: "Thank you for agreeing to chat..."
  first_message_2: "Thank you for agreeing to chat..."

bot_conditions :
  chatbot_1: "Context: You are a chatbot..."
  chatbot_2: "Context: You are a chatbot..."

```

Listing 1: Portion of a sample YAML configuration file for RCL. Multiple variables can be randomized as part of experiments, including bot conditions, recruitment materials, and GAI models.

information. Participants should be instructed to reply “Yes” in order to consent. If the system receives a message that begins with “No” then it ceases contact with that user. If it receives a message that says “Yes”, then it sends the `handoff_message`. Otherwise, it sends a new message asking the user to respond with “Yes” if they would like to participate. An example of a consent message is shown in Figure 2.

For users who consent, the `handoff_message` confirms consent and instructs them that if they are chosen to participate, they will receive a direct message from the bot. For participants in control, the `handoff_message` is the last message they will receive; for those in a treatment group, they will be sent the `first_consented_message` to initiate the conversation, and then all subsequent messages as generated by the GAI model. `initial_message`, `handoff_message`, and `first_consented_message` are all variables in the `config.yml` file, and should be edited to match the

researchers' study. For example, the `first_consented_message` should typically ask participants a question to initiate their conversation with the chatbot. Like the `bot_conditions` variable, both the `initial_message` and `first_consented_message` variables are dictionaries. If researchers include multiple entries, then RCL will randomly select a version of the message to send and will track which version was sent based on the key.

RCL also includes different options for *how* messages are sent, configured in the `message_flow` variable, which has three options: `handoff`, `modmail-only`, and `dm-only`. We recommend the `handoff` option; with this option, users are first contacted by the bot through “modmail”—a prioritized messaging system only available to subreddit moderators. Users who consent are randomized into either control or to talk with one of the bots in the `bot_conditions` dictionary, and are sent the `handoff_message` via modmail. After this, subsequent messages are all sent as direct messages from the bot, including the `first_consented_message` and all GAI-generated responses. In the `modmail-only` condition, RCL sends all messages through modmail. We don't recommend this approach, as the messages can clog up the modmail queue for other moderators. Finally, in the `dm-only` condition, the software sends all messages by direct message, including the initial invitation message. We don't recommend this approach either, because modmail messages are prioritized and much more likely to be seen by users, and DMs are much more likely to be perceived (and reported) as spam. However, `dm-only` may be the only option if the researchers are not partnering with subreddit moderators.

Once the configuration is complete, the next step is to test the chatbot. As explained in more detail below, the bot is intended to be repeatedly run as a cron job, but initially we recommend simply executing the `chatbot.py` script manually in order to facilitate debugging. While testing, we strongly suggest temporarily replacing the potential participants file with a CSV containing only accounts owned by the research team. The first time the script is run, it should send invitation messages to the users in this list; Figure 2 shows how these initial messages look when received by participants. Through Reddit, the researchers can send messages to the bot—testing the consent process, chatting with the bot, etc. After each message is sent, the `chatbot.py` file should be run again in order for the return message to be generated and sent by the bot.

If it appears that messages are sending correctly, confirm that they are being saved to the `conversations.csv` file, and that the information about the users is being saved to the `participants.csv` file. The system will

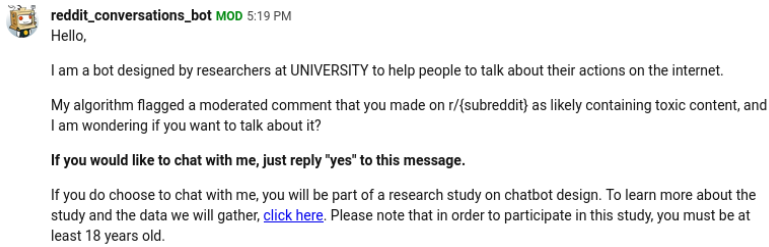


Figure 2: Example of how an invitation modmail message appears to users. Users can reply, and will be alerted every time the bot replies to them.

check the participants file before sending a new invite message, to ensure that users aren't invited to participate more than once. If there are errors, or if the researchers want to do further testing of the message flow, they can delete the `participants.csv` file and the bot will re-send the invitation message to each account in the potential participants file.

Collecting participant data

RCL also includes scripts designed to collect data about the actions of participants. The data collected can include moderation events (if the bot is a moderator with access to modlogs), as well as posts and comments written by participants. Researchers can then compare the communicative behavior of participants with those in other experimental conditions or with those in the control condition. This is rich text data and can be analyzed in many ways to answer different questions. For example, a researcher studying polarization might look for fewer inflammatory statements or lower activity in extreme subreddits. Researchers can test these data collection scripts to ensure that they are collecting data for all of the participants.

Deployment

When the researchers are confident that the GAI prompts are working as expected, and that the `chatbot.py` file is retrieving and sending messages and storing the outputs correctly, they can deploy the bot. In practice, this means moving the code to a server and setting up cron jobs to run the scripts.

`cron` is a utility designed to run programs at set intervals. RCL includes an example crontab file, which runs the chatbot every minute, the participant identification script twice per day, and the participant data collection script and modlog collection script once per day. For the chatbot script in

particular, it is necessary to ensure that only one instance of the script is running at a time, using `flock` or similar. The supplied `crontab` file shows one example of how to do this.

Feasibility

As mentioned above, RCL was used in a project directed at users posting toxic content on Reddit (Foote et al., 2026), which we briefly discuss here to show the feasibility of this framework. An initial concern is that people might not be willing to talk with chatbots on Reddit, or might not be willing to engage in good faith. However, even for this population—which we might expect to be less amenable to participating in research—approximately 12% of invited users consented to participate, and the bot engaged over 500 participants in conversations. As shown in Figure 3, many of these interactions were quite short, but others were long, engaged conversations. Figure 4 shows one of many examples of people having involved, thoughtful conversations with the bots.

We also might worry that users would find it unnatural to chat with an AI agent through Reddit’s messaging interface, or that the asynchronous nature of the system would feel frustrating compared to the immediate responses of platforms like ChatGPT. However, a qualitative review of the responses did not reveal criticisms of the system’s performance or complaints about the interface itself. In short, our experience suggests that people find the bot responsive enough and the context natural enough to hold meaningful conversations, even when mediated through a social media platform rather than a dedicated chat interface.

Additional concerns revolve around Reddit API limitations. The initial project tracked comments, moderation events, and suspensions for nearly 1,000 participants. The Reddit API limits were easily generous enough to allow daily updates to these measures for all participants. Experiments with much larger populations may require different approaches, such as less frequent updates.

Extensions

RCL is designed to make it easy to run the kind of research it was originally designed for—conversational experiments with Reddit users. Common extensions, like adding new bots, are designed to be trivial, simply by modifying configuration files. Other common extensions, such as changing the participant identification process, require writing new Python code. Where

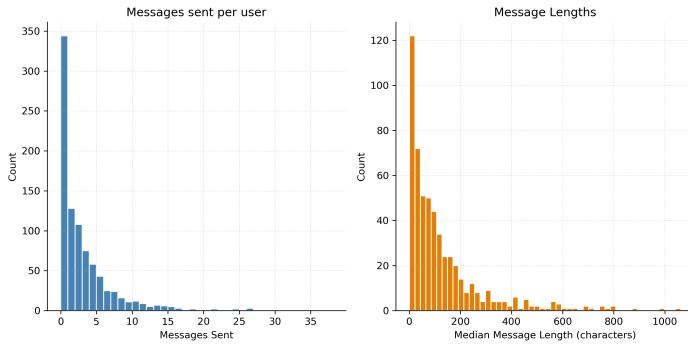


Figure 3: Number of messages sent and median comment length (number of characters) for each human-AI conversation, based on data from Foote et al. (2026).

possible, we have tried to make the code modular so that portions can also be repurposed for other, less related projects. For example, the participant identification and consent portions of the code could be used to recruit participants for survey or interview studies instead of chatbot experiments.

Considerations

Reddit terms of service

Some approaches to research with RCL may be in a gray area regarding some of Reddit’s rules. In particular, there is a rule against spam, defined as, “Sending large amounts of private messages to users who are not expecting them.”⁴ A related risk is community backlash, if users find an approach annoying or distasteful. Both of these risks can be greatly mitigated by recruiting community moderators as partners when possible. Messages sent through modmail to community members are much less likely to be considered spam. If this isn’t possible, researchers should limit the number of messages they send, but there will remain a risk of the bot being banned by Reddit. In the toxicity project (Foote et al., 2026), the bot account was given moderator access and sent recruitment messages as a moderator. This approach did not result in any warnings, bans, or suspensions due to spam concerns, despite messaging thousands of users.

⁴<https://support.reddithelp.com/hc/en-us/articles/360043504051-What-constitutes-spam-Am-I-a-spammer>

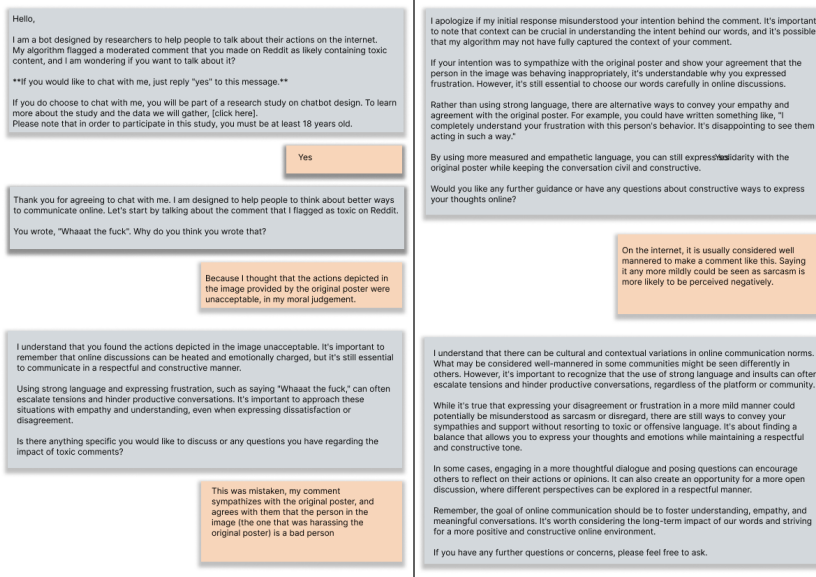


Figure 4: Example conversation between the chatbot (gray) and the user (orange). The conversation continued with user debating the toxicity and the content of the posted comment. The conversation eventually ended with the bot and the user expressing gratitude for the discussion.

Ethical considerations

There are a number of other ethical considerations that researchers should consider when doing this kind of research, including disclosure, participant tracking, and data sharing.

In the example configuration file, the consent and recruitment scripts make it clear to participants that they will be participating with a bot. This is reiterated in the sample prompts for the bots. However, some research questions may require deception, where bots pretend to be humans or processes do not include consent. In general, researchers should be very wary of this approach. A recent unpublished study that used GAI chatbots to persuade Reddit users without identifying itself as AI has been roundly criticized (O’Grady, 2025). Researchers should seek consent and transparency when possible, and work with institutional ethics boards to identify strategies to reduce risks and harms to participants when this is not possible.

A second consideration is tracking the behavior of participants and non-participants. For participants where consent to track activity is given, this is ethically unambiguous. However, it may be useful to understand

the behavior of non-participants (e.g., comparing participants to those who didn't respond to the invitation message), and RCL provides example code to do this. Researchers should be more cautious when gathering data from non-consented users, and consider carefully whether it is necessary to collect public behavior (e.g., comments and posts), non-public behavior (e.g., moderation events), or both for this set of users.

Even for consented participants, there are privacy concerns. The data gathered by RCL can include a mixture of public data, non-public data, and more private data (e.g., conversations with chatbots). Researchers should be careful to safeguard the privacy of their participants. RCL helps by using the UUID Python package to create 32 digit unique ids to replace usernames, and only stores usernames in a mapping file. We suggest keeping this mapping file on the server running the chatbot, and only exporting the anonymized files for analysis.

However, because posts and comments are public, replacing usernames is not enough to ensure anonymity: if comments are released alongside chatbot conversations, for example, then malicious actors could identify which users had which conversations simply by searching for comments on Reddit and identifying who made them. We recommend either releasing only aggregate information about participants' comments and posts, not releasing chatbot conversation data publicly, or using more advanced anonymization or sharing approaches. One option for data sharing is what Salganik (2017) calls a "walled garden" approach, where data is anonymized and shared only with approved researchers via repositories like the Harvard Dataverse.

Limitations and future work

While we believe that conducting AI field experiments on Reddit is an exciting, promising framework for researchers, there are some risks and limitations. First, GAI is a rapidly advancing technology. At a technical level, this means that models are constantly improving. When using commercial AI platforms, they may change invisibly during the course of a study (Bail, 2024). But more broadly, findings on the behavior or capabilities of current bots may not apply to future generations. At the same time, broader attitudes toward and cultural understandings of AI are also rapidly evolving.

RCL also has some risks as a software tool. For example, it relies on a number of external services and APIs, and these external dependencies could change or disappear. Some components are less fragile in this regard than others. For example, there are multiple GAI agent APIs which are

quite similar to each other, and converting the project to a new GAI agent API or locally running instance would be fairly simple. On the other hand, RCL uses the Reddit API via PRAW in order to identify, recruit, and track participant behavior, and the Reddit messaging interface is the user interface for participants to interact with chatbots. Reddit sometimes makes changes to their API access rules and messaging interface, and while the key features used by RCL have remained stable, it is possible that future changes could make some uses of our software difficult or impossible.

RCL could be extended in a number of different directions. First, survey software like Qualtrics could be integrated. Researchers may want to survey participants before and/or after conversations with chatbots, and extensions could be developed to do things like generate Qualtrics links with participant IDs embedded. Second, the software could be extended to be used on additional platforms with amenable APIs, such as Mastodon or Bluesky, or even a standalone web interface. Finally, RCL often prioritizes simplicity over efficiency. Our team plans to improve performance through changes like storing data in a database and moving away from the cron-based scheduling system, to enable a more parallelized approach. There are many other potential improvements, and we look forward to collaborating with computational communication scholars to further extend this open source project in the future. To support governance and sustainability, we plan to maintain RCL as an open-source repository on GitHub, with version control, documentation, and community contributions helping ensure its long-term usability and adaptability.

Conclusion

The Reddit Conversation Laboratory aims to provide communication scholars with an accessible tool for conducting large-scale, in situ experiments with generative AI chatbots. This methodological approach enables both collection of chatbot conversations in authentic online contexts as well as unobtrusive observation of post-experiment behaviors. We believe that this approach holds great promise for advancing our understanding of human-AI interaction and online communities, and we hope that by making this kind of work easier, our tool will stimulate and enable a new wave of research in these areas.

Acknowledgments

This project was supported by a Google Academic Research Award, and Loizos Bitsikokos was supported by an Onassis Foundation scholarship.

References

- Argyle, L. P. (2025). Political persuasion by artificial intelligence. *Science*, 390(6777), 983–984. <https://doi.org/10.1126/science.aec9293>
- Argyle, L. P., Busby, E. C., Gubler, J. R., Lyman, A., Olcott, J., Pond, J., & Wingate, D. (2025). Testing theories of political persuasion using AI. *Proceedings of the National Academy of Sciences*, 122(18), e2412815122. <https://doi.org/10.1073/pnas.2412815122>
- Bail, C. A. (2024). Can Generative AI improve social science? *Proceedings of the National Academy of Sciences*, 121(21), e2314021121. <https://doi.org/10.1073/pnas.2314021121>
- Baldassarri, D., & Abascal, M. (2017). Field Experiments Across the Social Sciences. *Annual Review of Sociology*, 43(Volume 43, 2017), 41–73. <https://doi.org/10.1146/annurev-soc-073014-112445>
- Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI Friend: How Users of a Social Chatbot Understand Their Human–AI Friendship. *Human Communication Research*, 48(3), 404–429. <https://doi.org/10.1093/hcr/hqac008>
- Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, 385(6714), eadq1814. <https://doi.org/10.1126/science.adq1814>
- Desmond, M., & Brachman, M. (2024). Exploring prompt engineering practices in the enterprise. *ArXiv*, *abs/2403.08950*.
- Foote, J. (2022). A Systems Approach to Studying Online Communities. *Media and Communication*, 10(2), 29–40. <https://doi.org/10.17645/mac.v10i2.5042>
- Foote, J., Kumar, D., Jha, B., Funkhouser, R., Bitsikokos, L., Goel, H., & Chiu, H.-C. (2026). Taming Toxic Talk: Using chatbots to intervene with users posting toxic comments. <https://doi.org/10.48550/arXiv.2601.20100>
- Gerber, A. S., & Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. WW Norton.
- Goldstein, J. A., Chao, J., Grossman, S., Stamos, A., & Tomz, M. (2024). How persuasive is AI-generated propaganda? *PNAS Nexus*, 3(2), pgae034. <https://doi.org/10.1093/pnasnexus/pgae034>
- Hackenburg, K., Tappin, B. M., Hewitt, L., Saunders, E., Black, S., Lin, H., Fist, C., Margetts, H., Rand, D. G., & Summerfield, C. (2025). The Levers of Political Persuasion with Conversational AI. <https://doi.org/10.48550/arXiv.2507.13919>
- Hu, B., Mao, Y., & Kim, K. J. (2023). How social anxiety leads to problematic use of conversational AI: The roles of loneliness, rumination, and mind perception. *Computers in Human Behavior*, 145, 107760. <https://doi.org/10.1016/j.chb.2023.107760>
- Lawrence, H. R., Schneider, R. A., Rubin, S. B., Matarić, M. J., McDuff, D. J., & Bell, M. J. (2024). The Opportunities and Risks of Large Language Models in Mental Health. *JMIR Mental Health*, 11(1), e59479. <https://doi.org/10.2196/59479>

- Lees, A., Tran, V. Q., Tay, Y., Sorensen, J., Gupta, J., Metzler, D., & Vasserman, L. (2022). A New Generation of Perspective API: Efficient Multilingual Character-level Transformers. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3197–3207. <https://doi.org/10.1145/3534678.3539147>
- Li, H., & Zhang, R. (2024). Finding love in algorithms: Deciphering the emotional contexts of close encounters with AI chatbots (M. Yao, Ed.). *Journal of Computer-Mediated Communication*, 29(5), zmae015. <https://doi.org/10.1093/jcmc/zmae015>
- Lim, S., & Schmälzle, R. (2023). Artificial intelligence for health message generation: An empirical study using a large language model (LLM) and prompt engineering. *Frontiers in Communication*, 8. <https://doi.org/10.3389/fcomm.2023.1129082>
- Lin, H., Czarnek, G., Lewis, B., White, J. P., Berinsky, A. J., Costello, T., Pennycook, G., & Rand, D. G. (2025). Persuading voters using human–artificial intelligence dialogues. *Nature*, 1–8. <https://doi.org/10.1038/s41586-025-09771-9>
- Muise, D., & Pan, J. (2019). Online field experiments. *Asian Journal of Communication*, 29(3), 217–234. <https://doi.org/10.1080/01292986.2018.1453850>
- Munger, K. (2017). Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment. *Political Behavior*, 39(3), 629–649. <https://doi.org/10.1007/s11109-016-9373-5>
- O’Grady, C. (2025). ‘Unethical’ AI research on Reddit under fire. *Science*, 388(6747), 570–571. <https://doi.org/10.1126/science.ady8074>
- Salganik, M. J. (2017). *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.
- Salvi, F., Ribeiro, M. H., Gallotti, R., & West, R. (2024). On the Conversational Persuasiveness of Large Language Models: A Randomized Controlled Trial. <https://doi.org/10.21203/rs.3.rs-4429707/v1>
- Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., Li, Y., Gupta, A., Han, H., Schulhoff, S., Dulepet, P. S., Vidyadhara, S., Ki, D., Agrawal, S., Pham, C., Kroiz, G., Li, F., Tao, H., Srivastava, A., . . . Resnik, P. (2025). The prompt report: A systematic survey of prompt engineering techniques.
- Skjuve, M., Brandtzaeg, P. B., & Følstad, A. (2024). Why do people use ChatGPT? Exploring user motivations for generative conversational AI. *First Monday*. <https://doi.org/10.5210/fm.v29i1.13541>
- Velásquez-Henao, J., Franco, C., & Cadavid, L. (2023). Prompt Engineering: A methodology for optimizing interactions with AI-Language Models in the field of engineering. *DYNA*, 90, 9–17. <https://doi.org/10.15446/dyna.v90n230.111700>