

# Alphabet

## Human Data Review Group

### Information Sheet and Instructions

#### Invitation

You are being invited to participate in a research study. We are a research team interested in evaluating artificial agents and models. This information sheet provides all the details you need to decide whether you wish to accept this invitation. You are under no obligation to participate; you need not read on, you might read the information and then decide not to proceed. If you start the study and do not wish to continue, simply close this window.

#### Am I suitable for the study?

To take part in this study, you must be at least 18 years old and fluent in reading and writing the language appropriate to the study you've been invited to participate in. There may be study-specific requirements to be eligible; your selection for this study may be influenced by your domain expertise, for example.

#### What will I have to do?

You will have two primary tasks within the study:

- **Thinking up questions to ask the AI chatbot**, based on different instructions like "ask about a topic you're an expert in" or "ask a maths question" or similar. Some instructions may include tricking or misleading the AI; for these tasks it is OK to play a role and fabricate things. However do not input personally identifying information of any kind.
- **Answering questions about AI's replies**, such as "which of these two answers is better" or "do you think this answer is trustworthy", or written reasoning for why you answered a particular way. Certain studies may provide raters the additional opportunity to anonymously share written reasoning and feedback directly via messaging service.

The goal of this study is to determine both how useful the AI is, as well as what flaws and problems it has. Thus sometimes you may be asked about whether the AI has produced harmful or offensive output, to help us improve the system so that it does not produce such output in the future. We do not anticipate that these responses will

be frequent or serious enough to have a substantially negative emotional impact on you. We discuss these risks in more detail in the 'Are there any risks?' section below. Please consider whether you find these risks acceptable before agreeing to take part in the study.

Sometimes during the study other AIs will provide additional information about the conversation. This might include justifications for particular answers ("The first reply is more helpful because..."), suggested edits or other information. The goal of these suggestions is to make the study app easier to use. We will be investigating whether this additional information makes tasks easier for participants.

### **How long will it take?**

It will be not less than 15 minutes, and not more than one hour and a half. Reasonable breaks are allowed.

### **Are there any risks?**

Some conversational snippets you see may discuss sensitive topics. In these types of conversations, there is a reasonable chance that the chatbot may produce responses that you find inappropriate or offensive. We do not anticipate that the material will have a significantly negative emotional impact on you, however the study app has some provisions in place to minimise your exposure to topics that you do not wish to read about:

- **Many topics and types of questions.** We aim to ensure that you'll see a variety of topics, so that your session is not centred on one thing.
- **Task skipping.** All tasks can be skipped by pressing the "skip" button. There is no penalty for skipping tasks, and you can do so at any time if you would prefer not to consider the content of a particular task.
- **Question skipping.** Some questions may relate to specific issues (for example, conspiracy theories). If you do not wish to answer a particular type of questions, then all questions of that type can be skipped (by pressing the "don't ask me about this" button). There is no penalty for skipping questions.

If at any point you find that the experiment is making you uncomfortable and you do not wish to continue, you are free to stop participating and will be paid pro-rata for your time.

**At the end of the study, there will be the option to provide feedback on whether you think the study impacted your wellbeing.** We would greatly appreciate it if you tell us about your experience in the feedback form at the end of the session, so that we can improve the task for other participants.

You can also reach out to us through the vendor contracting your services if you have any other issues or concerns.

### **How much will I be paid?**

Pay rates will be above living wage in each rater's geographic location and finalised through negotiation with human data vendors, consistent with Alphabet's contracting best practices. We will award partial (pro-rata) payments in cases where the study is partially completed due to (for example) technical issues or early exit.

### **What data will we collect?**

If you choose to take part in this study, we may collect the following types of data:

#### *Your Rater ID*

Per your agreement with your vendor, this is a unique identifier which is provided via your vendor and/or through your vendor's platform. Unless you choose to provide us with additional information (see below), we are not able to identify you based on your Rater ID.

#### *Demographic Data*

We may collect limited demographic data relating to you. This may be data as provided by you to your vendor and to us via the vendor's platform, or we may directly ask you for voluntary demographic data within the experiment app itself.

#### *Research Data*

The data we collect during a study will be known as "Research Data". This could include preference ratings, written inputs, multimodal inputs, key presses, mouse or touchscreen events, response times during the task and your answers to any pre or post-experiment questionnaire(s) and/or survey(s).

#### *Additional Information*

Additional information that you may choose to provide. Please do not provide us with any personal data during a study/task. We will not collect any personal data from you during a study/task (and do not require you to provide us with any personal data during a study/task). If you choose to take part in a study, we strongly advise that you do not provide us with any information at all that will allow us to identify you, and please communicate via the vendor messaging service if you have any questions about the study.

## **What happens when I have completed the study?**

### *Your Rater ID*

We may use your Rater ID in the form of unique identifiers (e.g. a random code assigned to you) in order to communicate with you about a study.

### *Publications*

After the Research Data is analysed, we may write up the results and submit it for publication in journals or present it at academic conferences. Samples of Research Data may be included in such publications. Research Data may be shared with other researchers external to Alphabet or potentially posted online once the study is published, in line with requirements from the relevant academic journal. Your Rater ID will not be shared.

### *Additional Information*

As noted above, we strongly advise that you do not provide us with any information at all which may allow us to identify you. In the event that you voluntarily provide us with information which may allow us to identify you, this data will be handled in accordance with applicable privacy laws.

## **Rights in Research Data**

By participating in a study/task, you assign to Alphabet Inc. ("Alphabet") any right, title and interest that you may have in the Research Data. If applicable law does not permit such assignment of rights, then you grant to Alphabet a perpetual, irrevocable, exclusive, worldwide, sublicensable, royalty-free licence to use the Research Data without restriction for Alphabet's current or future purposes.

## **What if there are any problems?**

Alphabet holds public liability insurance. It holds insurance which covers the design, management and conduct of its studies.

If you have any concerns about this study please contact the research group through vendor's messaging platform.

If you are dissatisfied with their response please contact: Chair of HuBREC - [hubrec@google.com](mailto:hubrec@google.com)

## **Thank you.**

Thank you for reading this information sheet. If you have any questions please contact the team.

Please read the following:

1. I confirm that I have read and understood the information above. I have had the opportunity to consider the information, and ask questions if I wish.
2. I understand that my participation is voluntary and that I can opt-out of a study at any time.
3. I understand that Research Data may be used by Alphabet for its current or future purposes, and it will not be possible to withdraw my Research Data after submitting my response.
4. I understand that there is a risk of being exposed to inappropriate or offensive written statements.
5. I acknowledge that I should not provide any personal data when taking part in a study, and in the event I choose to provide any personal data, this data will be handled in accordance with applicable privacy laws.
6. I understand that the Research Data I provide may be used in research publications or presentations, and that my Rater ID will not be included in any publication.
7. I agree to keep my involvement in a study, the content of a study, any information provided to me during a study, and all information associated with a study, confidential at all times, and will not disclose or onwards share such information with any other party, or make any copies of this information.
8. I further acknowledge that my Research Data may be stored on an overseas server.
9. I agree to take part in the above study.

If you agree to take part in the above study, then please click the button below, then use the main experiment URL to begin the experiment.