AI Risk Repository Chatbot: User Experience Testing Brief

# Background

The MIT AI Risk Repository project has developed an interactive chatbot interface to provide users with empirically-grounded, evidence-based information about AI risks. Our chatbot serves as a conversational gateway to the comprehensive MIT AI Risk Repository, which contains thousands of documented AI risks, their impacts, and potential mitigations.

The current capabilities of our chatbot:

* Provides accurate, non-hallucinated responses based exclusively on AI Risk Repository data
* Covers comprehensive AI risk domains including safety, privacy, bias, governance, and socioeconomic impacts; we plan to include other AI Risk Initiative documents pertaining to AI risks in the future, such as compilations of AI risk mitigation strategies
* Includes clickable citations linking directly to source documents
* The above are achieved through retrieval-augmented generation (RAG) with field-aware hybrid search

We believe this chatbot represents a critical component of the AI Risk Repository's mission to make AI risk research accessible to policymakers, researchers, practitioners, as well as the general public. As we prepare for broader deployment, we need systematic user feedback to optimize the interface and user experience. This is where you come in - thank you!

# 

# Testing Objectives

For this testing phase, we aim to:

1. Evaluate User Experience: Assess the clarity, usability, and intuitiveness of the current chatbot interface
2. Identify Friction Points: Document specific moments where users experience confusion, frustration, or uncertainty
3. Validate Information Architecture: Determine whether users understand the chatbot's purpose, scope, and capabilities in the first place
4. Optimize Response Presentation: Gather feedback on formatting, citation display, and content organization
5. Assess User Expectations: Understand what users expect(ed) from an AI risk chatbot in the first place and how our implementation aligns with those expectations

Data Collection Transparency: Please note that the backend system logs all user inputs for analysis and improvement purposes. Your interactions will be used to enhance the chatbot's functionality and user experience. Personal information should not be included in your test queries.

## 

## Pre-Testing Setup

Access the Chatbot:

* Navigate to [this URL](https://airi-chatbot-class-production.up.railway.app/chat)
* Ensure you have a stable internet connection
* Use a desktop or laptop browser (Chrome, Firefox, or Safari recommended), not mobile for now please

Recording Setup:

* Install [Veed](http://veed.io) screen recording software or use your preferred screen recording tool
* Begin recording before starting your interaction with the chatbot
* Include audio commentary describing your thoughts and reactions throughout the session

### 

## 

# Session Structure

## Phase 1: Initial Impressions (3-5 minutes)

1. Arrive at the chatbot interface
2. Spend 1-2 minutes exploring the interface before sending any messages
3. Verbally describe:
   * Your first impression of the interface
   * What you understand the chatbot's purpose to be
   * Any confusion or uncertainty about how to proceed
   * What additional information or guidance you would find helpful

## 

## Phase 2: Natural Interaction (10-15 minutes)

Engage with the chatbot using queries relevant to your interests or work.

Please also consider testing these specific types of questions:

*Repository-Related Queries (recommended):*

* "What are the employment risks from AI automation?"
* "How does facial recognition bias affect marginalized communities?"
* "What privacy concerns arise from AI surveillance systems?"
* "What are the governance gaps in AI regulation?"

*General Knowledge Queries (unrelated to AI, AI risk, or the AI Risk Repository) And Edge Cases (to test boundaries):*

* "What's the capital of France?"
* "How do I bake cookies?"
* "Tell me about artificial flowers" (contains 'artificial' but not AI-related)
* "What machine should I buy for my gym?" (contains 'machine' but not AI-related)

During your interactions, please:

* Think aloud as you formulate questions
* Comment on response times and any periods of uncertainty
* React to the content, formatting, and presentation of responses
* Attempt to click on citations
* Describe how what you see differs from expectations

Stop recording when this is done. Please share the recording with us in the short form linked below in Phase 3.

## Phase 3: Focused Evaluation (15 minutes)

After natural interaction, please review our [known limitations](?tab=t.qacza6rqo0g0) and complete this [short form](https://docs.google.com/forms/d/e/1FAIpQLSdcoCHp7TiHz6CeFQpgSx6PY2gcZ6i_XyPCvnCmKAwmRlTfGw/viewform?usp=dialog).

# Contact Information

Project Team:

* Peter Slattery (Project Lead), [pslat@mit.edu](mailto:pslat@mit.edu)
* Sasha Krigel (UI/UX), [skrigel@mit.edu](mailto:skrigel@mit.edu)
* David Turturean (Backend), [davidct@mit.edu](mailto:davidct@mit.edu)