**IRB Tipsheet: Generative Artificial Intelligence (AI) Use in Human Subjects Research**

# **Overview**

This document is intended for researchers who may be using Generative AI in their research. The goal is to help researchers think through relevant IRB-related issues with using AI and how to communicate about AI use to the IRB and research participants. Recognizing that AI is a rapidly changing area, please notify the IRB if you find outdated information in this document.

# **I. Generative AI in Human Subjects Research**

Generative AI can be used in human subjects research in a variety of ways. See some applications below:

1. An AI tool is used to obtain, access, analyze, or transcribe identifiable data about human participants
   * Example: A researcher submits participants’ responses to open-ended survey questions to an AI tool for summarizing.
   * This may also include AI Validation Research if the AI tool is being trained or tested on data from human subjects that contains potentially identifiable private information.
2. An AI tool collects data from humans through interaction or intervention
   * Example: A study examining interactions between participants and an AI chatbot.
3. An AI tool acts as an extension of the researcher
   * Example: An AI tool is used to answer participants’ questions about a study procedure or provide informed consent information to participants

# **II. Primary Concerns**

These are the primary concerns that the Institutional Review Board (IRB) has when generative AI is used in research:

## **Identifiability and Privacy**

* + **Could data be reidentified?**

When an AI tool has access to participant data, it may be possible that deidentified data could become reidentified through the tool’s access to and ability to combine large numbers of data points.

* + - The AI tool’s access to demographic data that could be used to identify an individual should be limited.
    - The AI tool’s access to more sensitive data should be highly restricted [e.g., FERPA protected data, biospecimens, genomic data, data that could put ppts at risk (e.g., audio/video recordings about sensitive or illegal topics)]
      * See the [UW Generative AI Guidance](https://www.wisconsin.edu/uwsa/ots-helpdesk/information-security/generative-ai/)
  + **What is a public vs. private online space** **where AI may be pulling data?**

The line between public and private spaces online is blurred. If an AI tool is scraping data from online sources, researchers will need to document the limitations and parameters placed on the tool to protect privacy (*see Public vs. Private spaces document for on IRB website additional information*).

* + **What will the AI tool do with the data it is given access to?**

By default, content entered into AI tools may be used to train the model, and promising privacy or confidentiality may not be possible if AI has access to participants’ data. This may even be the case for subscription versions. Some AI tools will allow you to turn off this default setting. However, even if this default setting is turned off, there also can be content monitoring that happens by employees, so data aren’t entirely private (again, even for subscription versions). It’s important to understand the privacy and terms of use policies for each specific tool you are using and keep aware of changes to these policies. [Murphy library](https://libguides.uwlax.edu/ai_tools/terms) has a collection of such documents that may be a helpful start.

## **Transparency**

* + Consent forms should notify participants if a generative AI tool will have access to their data in a research study. The consent form should include information in lay terms about:
    - How AI is being used in the research
    - What data the AI tool will have access to
    - What the AI tool will do with the data (if you don’t know, say that)
    - If/when data can be removed from the AI tool
  + **Example language for the “Confidentiality” section of a consent form**: The responses you give to the open-ended survey questions will be entered into ChatGPT for analysis of themes. These responses will first be reviewed to make sure they don’t contain any identifying information. Demographic data you provide (e.g., race, age, gender) will not be accessible by ChatGPT. Once your deidentified survey responses are submitted to the AI tool, they can’t be individually removed and may be used to train or improve the tool. It is also possible the survey responses submitted to the AI tool will be available to and used by others in ways we cannot predict.

# **References**

HHS. (2022, July 21). Considerations for IRB Review of Research Involving Artificial Intelligence. <https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-e-july-25-2022-letter/index.html>

ORIED. (n.d.). AI Tools: Guidance on the use of AI in human subjects research. University of TN-Knoxville. <https://research.utk.edu/research-integrity/artificial-intelligence-ai-tools/>