SHSU IRB Guidance: Guidance for Online Data Collection and Bot Interference

Introduction:

The increasing prevalence and sophistication of bots (automated software programs) pose a significant risk to the integrity of online research involving human subjects. It is crucial for researchers to be aware of these risks and to proactively implement measures to mitigate potential bot interference in their studies. The IRB plays a critical role in ensuring the ethical conduct of research, which now includes considering the impact of bots on online data collection. Therefore, researchers submitting protocols for online data collection should consider the following:

Understanding Bots and Bot Incidents

- **Definition of a Bot:** A bot is a software program designed to perform tasks on the internet without human involvement. They can automate internet functions such as interacting with websites, completing forms, web scraping, and spamming.
- **Definition of a Bot Incident:** A bot incident is any event in which a bot interferes with or disrupts the integrity of a research study. This can include:
    - Bots entering online surveys and attempting to claim incentives.
    - Unauthorized data collection.
    - Manipulation of responses.
    - Any other actions that compromise the validity of the research and data.
- **Vulnerability of Online Surveys:** Online survey research is particularly vulnerable to bots due to the ease of automation for tasks like clicking boxes and responding to close-ended questions and scales
- **Increasing Sophistication:** With the rise of artificial intelligence, bots are becoming increasingly sophisticated, capable of generating human-like responses to open-ended questions, making them harder to detect.

IRB Considerations and Expectations

- **Proactive Bot Mitigation Plan:** The IRB will expect researchers to include a bot mitigation plan in their online research protocols. This plan should outline the steps that will be taken to prevent and detect bot interference. Relevant citations can be found on the IRB website.
- **Awareness of Suspicious Data:** Researchers are expected to be particularly aware of and attuned to suspicious-looking or outright falsified data, as this can be bot-generated. Examples of suspicious data can include, but are not limited to, …..
- **Ethical Review:** If a bot incident occurs that involves a breach of participant information, problems with consent, improper handling of incentives, or compromised data, the IRB must be involved in an ethical review.
- **Collaboration with Experts:** The IRB may host webinars that include experts as well as encourage researchers to collaborate with IT or cybersecurity departments within the institution to develop effective security protocols.

Proactive Strategies for Researchers

Researchers should consider implementing the following proactive measures in their online study designs:

- **Verification Tools:** Employ verification tools such as CAPTCHA or reCAPTCHA to help distinguish between humans and bots.
- **Attention Checks:** Include attention check questions within the survey to identify responses that are not likely from engaged human participants. For example, asking participants to select a specific answer to a question unrelated to the study topic.
- **Dummy Questions:** Incorporate a few dummy questions within screeners that are relevant to the study but not strict inclusion/exclusion criteria to make it harder for bots to determine eligibility.
- **Redundant Data Collection:** Ask for related information in different parts of the survey and check for consistency (e.g., zip code and county). Be mindful that this can require significant manual checking.
- **PINs and Passwords:** For studies sent via mail or other direct methods, consider using both a unique PIN and a predefined password for survey access.
- **IP Address Collection:** Collect IP addresses (with participant consent and IRB approval) to monitor for suspicious patterns of responses coming from the same address in a short period. Ensure this is disclosed in the consent form as it may change the review category.
- **Response Time Monitoring:** Be aware of unusually fast completion times, which could indicate automated responses. However, consider the target population and the potential for varied response times (e.g., postpartum individuals).
- **Screening Procedures:** Consider requiring potential participants to schedule a screening meeting during which the research can verify the participant meet the criteria, understands study protocols such as both audio and video recording, ….. These steps may be especially relevant for researchers who are asking qualitative research questions and completing interviews to gather data.
- **Incentive Management:** Offering incentives can attract participants to online studies; however, it can also be a target for bots seeking to exploit the compensation offered1 . Researchers should carefully plan their incentive management strategies and clearly communicate these to participants. Consider the following:
    - **Clear Language in the Consent Form:** It is crucial to include clear and specific language in the consent form regarding how and when participants will be compensated. This should outline the conditions for receiving the incentive and any limitations.
    - **Timeframe for Compensation:** Consider including a timeframe for compensation in the consent form (e.g., compensation will be provided within one or two business weeks of survey completion). This allows the research team time to review responses for potential bot activity before distributing incentives.
    - **Delayed Compensation for Larger Studies:** For larger-scale studies with significant incentive budgets, researchers might consider **a delay in compensation** to allow for thorough verification of a larger volume of responses. This should be clearly stated in the consent form.
    - **Fraud Detection Mechanisms:** Implement fraud detection mechanisms to monitor the distribution of incentives for suspicious patterns. This could involve looking for unusual

numbers of requests from the same IP address (though bots can sometimes circumvent this) or other anomalies in incentive claims.

- **Limitations on Participation and Incentives:** Explicitly state in the consent form and recruitment materials if participants are only eligible to complete the survey and receive compensation once. This can help to deter individuals (and bots) from attempting to claim multiple incentives. As one case study highlighted, the absence of such a statement can lead to exploitation.

By carefully planning and clearly communicating their incentive management strategies, researchers can help to mitigate the risk of bot exploitation while still ethically compensating legitimate participants for their time and effort.

- **Survey Design:** Ask similar questions at different points in the survey to check for consistency in responses. Be mindful that real participants may also have slightly different answers, and overly repetitive questions might frustrate participants.
- **Leverage Platform Features:** Utilize security features built into survey platforms like Qualtrics and RedCap. However, be aware that these may not be foolproof.
- **Careful Recruitment Language:** Use caution with language in recruitment materials (especially online postings) that might attract bots. Avoid explicit mentions of large monetary incentives and be aware that bots may scan for terms like "$", "research," or "study". Consider phrases like "compensation may be provided" or "opportunity to enter a raffle".

Reactive Strategies if a Bot Incident Occurs

- **Halt Data Collection:** If suspicious activity is detected, be prepared to immediately halt data collection to prevent further contamination of the data and depletion of incentives.
- **Consult with the IRB:** Immediately inform the IRB of any suspected bot incident, including the extent of the potential impact on data integrity and incentives.
- **Engage IT/Cybersecurity:** Work with your institution's IT or cybersecurity department to investigate the incident and implement measures to prevent future occurrences.
- **Review Data:** Carefully review the collected data to identify and remove suspected bot responses. This may involve looking for patterns in response times, IP addresses, attention check failures, and open-ended responses.
- **Communicate with Participants (if necessary):** The IRB can provide guidance on how to communicate with participants appropriately if compensation needs to be adjusted or if data integrity is questioned. Be cautious about revealing too much information about bot detection methods to avoid educating bot designers.
- **Document the Incident:** Submit an Incident through Cayuse Human Ethics to thoroughly document the bot incident, the steps taken to address it, and lessons learned for future studies. Here is a guide to get you started (link to resource). Researchers also need to consider how to document the incident in any formal dissemination of their research such as during scholarly presentations and in published scholarly articles.

Ethical Considerations

- **Compensation and Fairness:** Strive to ensure that legitimate human participants are compensated for their time. Err on the side of compensating participants when there is uncertainty about whether a response is from a bot or a real person.

- **Potential Misidentification:** Be aware of the possibility of inadvertently flagging real human participants as bots due to overly strict detection methods, especially in vulnerable or unique populations. Consider qualitative data and potential variations in responses from real participants.
- **Transparency in Consent:** Clearly outline in the consent form the possibility of bot interference and the methods the study will employ to detect and mitigate it.

By carefully considering these points and developing a comprehensive plan to address the risks posed by bots, researchers can better protect the integrity of their online data collection and ensure the ethical conduct of their research. The IRB is a resource to help navigate these complex issues, and open communication is essential throughout the research process.