AI/Bots and Online Surveys

The Issue: The use of AIs and other technological advancements are important tools in a researcher's toolbox. Unfortunately, they are also effective tools for fraud, typically used in online attacks.

Highest Risk: AI/Bot attacks primarily target online surveys with an incentive component and where recruitment is conducted through social media.

IRB Responsibility: The IRB understands that some studies need to use social media platforms and/or online survey platforms in research to promote scientific validity. However, the use of such techniques raises special concerns for the IRB in the perspective of research ethics and security. Thus, the IRB reviews the protocol using social media platforms/online surveys more carefully.

1. What is AI and what are Bots?

- **Artificial Intelligence (AI)**: is a type of technology allowing computers to perform tasks that typically require the use of human intelligence. AI can learn from experiences and work to complete tasks, including but not limited to performance improvement, data analysis, translate both spoken and written communication, offer recommendations, and perform tasks in unpredictable conditions without the use of human interaction or oversight.
- **Bot(s)**: are an autonomous program on the internet and/or other networks that can interact with systems or users. Bots can follow specific instructions to imitate human behavior, but bots are faster and more accurate. A bot can also run independently without human intervention present.

2. IRB review and approval of AI/Bots and Online Surveys

- **Technological Safeguards**
  - If recruiting via social media, send respondents to a dedicated survey portal where you can control access and/or restrict the number of times a respondent can take a survey when possible. Examples include, but not limited to:
    - Limiting the number of times a person from a certain IP address can take the survey.
    - Use a portal like Qualtrics where additional safeguards are used (see below).
    - Use a Quick Response (QR) Code, (if possible).
    - Whenever possible, use encrypted platforms.
  - If funding is available, use a professional survey service or survey recruiting service.
  - When possible, use embedded attention checkpoints within researcher surveys (see example in Qualtrics section).
  - When possible, randomize survey question order as a means of baffling "quick click" bots. This works for situations where the bot is trying to complete several surveys at the same time.
  - Use reCAPTCHA for user verification. "reCAPTCHA is a free service from Google that helps protect websites from spam and abuse. A "CAPTCHA" is a Turing test to tell human and bots apart. It is easy for humans to solve, but hard for "bots" and other malicious software to figure out. By adding reCAPTCHA to a site, you can block automated software while helping your welcome users to enter with ease. Try it out at https://www.google.com/recaptcha/api2/demo."*
  - When possible, do not use open/public access links. PIs should use unique or personalized survey links that can only be accessible through a controlled study site/survey site/survey service. This helps limit responses to one per participant.
  - If possible, use "Honeypots" in your survey. A Honeypot is a question(s) that a person cannot see as they are in a hidden form field (invisible to people) but can be seen by a bot. The bot generally will not know this "question" needs to be skipped.
  - Qualtrics Specific: Below are additional guidelines, resources, and tools to consider adding into a Qualtrics-based survey.
    - Adding a cap on surveys (ex: 300 participants)
    - New security settings in Qualtrics
    - Addition of "Commitment Checks": https://www.qualtrics.com/blog/attention-checks-and-data-quality/

- Fraud detection: https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/?utm_medium=product&utm_source=survey-builder#RelevantID
- Fraud detection cont.: https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection/
- Geo Location: https://www.qualtrics.com/support/survey-platform/survey-module/using-logic/#GeoIPLocationLogic
- Caps on surveys: https://www.qualtrics.com/support/survey-platform/survey-module/survey-tools/quotas/#SpecifyingQuotaActions

- **IRB protocol Safeguards**
  - When preparing an IRB protocol, please use the "pick two" method, and do not rely solely on one method of protection from Bot/AI attacks.
  - IRB reviewers will ask where the information/survey links will be posted, and if they are posted on public/private feeds.
  - In the survey, repeat random questions, but phrased in a different way or requires a different way of responding (ex. "what is your age" w/ a drop drown menu and a "how old are you" with a fill in text box).
  - Verbiage that can be added in the Informed Consent (IC):
    - "Participants must take the survey in good faith"
    - "Incentives will not be awarded to participant(s) if it is determined fraud is used"
    - "Participants must be living individuals"
    - "Only one incentive awarded per person/address/etc." (depending on study design)
    - "AI-generated/Bot responses will automatically be discarded"
  - PI may want to consider stating country/geographic region applicable to the survey (Ex: MUST be in the US to patriciate)
  - Please note that social media sites are NOT secure. When feasible, the IRB encourages PIs to use professional/ dedicated sites for recruiting before using public social media (ex. ResearchMatch).

- **Other item(s)**
  - PI must read the terms of service/terms of use agreement(s) for ANY social media platform/site they intend on using for their project. Depending on the social media platform/site, the PI/researcher(s) needs to look for the following key points:
    - *Who owns the data from the social media platform/site?*
    - *Who else may be able to use the data collected (ex. third party vendors)?*
    - *Are there any research-related restrictions when using this social media platform/site?*
    - *Are there any research-specific policies to keep in mind when using this social media platform/site?*
  - The BSU IRB will not require PIs to pay out for fraudulent AI/Bot attacks, ***if the PI took reasonable precautions and safeguards***.
  - The PI will need to separate legitimate responses from fraudulent ones as the IRB expects that people who took a survey in good faith would still receive the incentive.
    - BSU's IT Security department may be of help, as are the support staff who administrate Qualtrics (if Qualtrics was used).
  - When determining if a respondent/response is real:
    - Check to see where the IP address is located.
    - Make sure the inclusion criteria is actually met.
    - Check how long it took the respondent to take the survey and if it is close to the estimated time listed in the IRB protocol and consent form (ex. a 20 minute survey was completed in 1 min.)
    - Check email addresses from respondents to see if there is a pattern (ex. johndoe1@XYZ, johndoe2@XYZ, etc.)
    - Check when responses came in and how many in a particular timeframe. For example, 100 responses came in within 2 minutes.
    - Look to see if open ended responses are too generic or are inconsistent with other responses (ex. "I work in the industry"; male is checked in the demographic section, but questions about female health are answered.)
  - Restrict the use of public social media to minimal risk studies. **NEVER** use for identifiable confidential/ sensitive information or data collection.
  - After the survey is closed, if "participants" start contacting you about the incentive and you are reasonably sure the survey has been compromised, it is OK to ask for some type of validation/verification (see items below).

3. **What to do if…:**
   a. **You are contacted initially by "participants" interested in taking part of your study. Here are some key factors to look for:**
      i. Consider why they are contacting you and not going to the survey themselves.
      ii. While these contact emails may include your title, name, etc. how did you list yourself on the contact/recruitment materials and how did they use your info on the email (ex. Dr. Christopher (Chris) Zerg, but the email greeting is the exact same instead of "Hi Chris or Dear Dr. Zerg".
      iii. What are the time/date stamps indicated on the received emails? Typically, AI-generated/Bot emails come in multiples at once. For example, you may receive five emails that all came in at 7:30pm.
      iv. What does the email formatting look like? Are all the emails formatted similarly? Are the emails void of details or very generic? Are there several groups of emails that look similar but with different names? Please see three examples below sent by three interested participants – all coming in at the same time:
         1. Example one: *Tuesday, 4/29/2025 12:30pm*
            a. *Hello, I trust this email finds you well. I write to express my keen interest in your study program on the above subject matter. Please I would like to share my view. Thanks for your consideration. And best regards!*
         2. Example two: *Tuesday, 4/29/2025 12:30pm*
            a. *Hello, please I would love to take part in your study on the above subject. It will be a great pleasure to work with you. Thank you very much.*
         3. Example three: *Tuesday, 4/29/2025 12:30pm*
            a. *Hello, I trust this email finds you well. Please I write to express my outmost interest to Participate in your study on the above subject. Thanks in advance for your consideration. Looking forward to hearing from you. Best regards!*
      v. What does their email handle look like? Are they a BSU email you recognize? Do they have similar handles that follow a similar formatting? Examples to look for can be the following:
            a. A person's first and last name followed by numbers. Example: You receive an email from a Jane Smith; SmithJane012345@gmail.com

If you wish to respond to these emails, please use the "verification template" and "additional verification guidance" below in the Other Resources section. In the circumstance above, consider the following:

1. Ask the potential participants that reached out directly to you to submit their full name and the name of the school/business/university/organizations/etc. with which they are associated.
2. Let them know you will reach out to the above's administrators to confirm their association with the above without collecting further identifiable information.
3. Remind them that their final interview data/survey results will not be connected with their identity if it is confirmed they are eligible.
4. If the potential participant provides contact information, do not use that, but verify it independently. For ex. PI is provided with 555-555-5555 and johndoe@XYZ.org, look up XYZ and see what their contact information is.

4. **Other resources:**

   a. Verification template:
      i. *Unfortunately, it was determined that a vast majority of responses to the survey were either fraudulent or the result of an AI/Bot attack. Based on my meeting with IT and our IRB office, we agreed compensation is only appropriate if it can be verified that you met study criteria listed in the informed consent __(insert criteria here)__. If this is the case, can you please provide ___(verification here)___. If I can verify you meet study eligibility, then the incentive can be given. Otherwise, I cannot provide compensation.*
   b. Additional verification guidance:
      i. If a person who used a Bot/AI tries to take someone else's name at a company/university/etc. when you are trying to verify, consider speaking directly with the person at the company and directly verify if they took part in the survey. Ex. Jane Doe works at ABC healthcare, the fraudster uses Jane's name to "verify" their identity, it is ok to contact ABC Healthcare and verify Jane works there and ask to speak with her about taking part in the study.

ii. If the potential participant provides contact information, do not use that, but verify it independently. For ex. PI is provided with 555-555-5555 and johndoe@XYZ.org, look up XYZ and see what their contact information is.

5. **Appendix:**

   a. *Google reCAPTCHA definition: https://support.google.com/recaptcha/?hl=en