

INTERVIEW GUIDE
v.2022.10.19

WELCOME/INTRODUCTORY REMARKS

- Introduce yourself, thank the interviewee for his/her/their time and willingness to participate.
- Remind the individual that we are recording the Zoom interview but that we will take reasonable steps to redact identifying information (i.e., de-identify) the comments in the transcript. Discourage the interviewee from providing identifiable information.
- Remind the individual that this study is voluntary, that they can decline to answer questions, and that they can end the interview at any time for any reason or no reason.
- Give the interviewee a roadmap of what to expect:
 - First, we will ask some basic questions about the informants so that we are able to describe our interview sample.
 - Next, we will get into substantive questions that we have preplanned.
 - Before we end the interview, we will give you an opportunity to share additional comments or thoughts that you think are important for us to consider as well.
 - After we end the recording, we will gather any information that is necessary to deliver the e-gift card as a thank you for your participation. This information must be kept separate from the study records.
- Do you have any questions before we get started?
- **[BEGIN RECORDING NOW]**
- State details for the record and obtain consent here:
 - Today is [DATE]
 - My name is [INTERVIEWER NAME].
 - I will be interviewing you today as a key informant in the study we are conducting at Penn State titled, “Exploring the Ethics of Synthetic Data and Artificial Intelligence.”
 - Do you give us your consent to participate in this interview study?

BASIC QUESTIONS ABOUT YOU (DEMOGRAPHICS)

These first set of questions are intended to help us describe the informants in this study. We have specific answer options, so we’ll read the question to you and please tell us which answer option you prefer.

1. In which age group do you belong?
 - a. 18 to 25 years old
 - b. 26 to 35 years old
 - c. 36 to 45 years old
 - d. 46 to 55 years old
 - e. 56 to 65 years old
 - f. 66 to 75 years old
 - g. 76 years and older
2. What terms best express how you describe your gender identity? Select all that apply.
 - a. Man
 - b. Woman
 - c. Non-binary
 - d. Transgender
 - e. None of these describe me. I identify as: (free text)
3. Which categories describe you? Select all that apply.
 - a. American Indian or Alaska Native
 - b. Asian
 - c. Black, African American, or African
 - d. Hispanic, Latino, or Spanish
 - e. Middle Eastern or North African
 - f. Native Hawaiian or other Pacific Islander
 - g. White

- h. None of these fully describe me. I identify as: (free text)
- 4. What is the highest grade or year of school you completed?
 - a. Grade 11 or below
 - b. Grade 12 or GED (High school graduate)
 - c. 1 to 3 years after high school (Some college, Associate's degree, or technical school)
 - d. College 4 years or more (College graduate)
 - e. Advanced degree (Master's, Doctorate, etc.)
- 5. Which of the following best describes your current employer?
 - a. Medical school
 - b. Teaching Hospital or Health System
 - c. College or university
 - d. None of these. I would describe my employer as: (free text)
- 6. How would you describe your current professional role or position?
- 7. How much experience do you have in your current professional role?
 - a. 0-5 years
 - b. 5-10 years
 - c. More than 10 years
- 8. In which state or territory in the United States do you currently work?

SUBSTANTIVE QUESTIONS

About Synthetic Data

[Questions will cover matters regarding key ethical issues, social implications, policy implications, and challenges and/or opportunities in addressing these matters]

We would like to talk to you about synthetic data.

- 1. Are you familiar with synthetic data?
- 2. Have you used or encountered synthetic data in your work?
 - a. If used: for what purpose did you use synthetic data?
 - b. If encountered: In what contexts or use cases have you encountered synthetic data being used?
- 3. What is your general opinion of the use of synthetic data in biomedical contexts? (e.g., favorable, unfavorable)
 - a. Can you tell us why you see the use of synthetic data in biomedical contexts as [favorable/unfavorable]? ~~What is the basis for your opinion? For example, is that view due to the utility or potential social benefit?~~
- 4. Do you see opportunities with the use of synthetic data in biomedical contexts? What kind of opportunities? Opportunities to whom?
- 5. Do you see risks associated with the use of synthetic data? What kind of risks? Risks to whom?
- 6. What do you see as the major ethical issues surrounding the creation and use of synthetic data in biomedical contexts?
 - 6.a Do you think synthetic data can address health information privacy concerns? Could you explain your response?
 - 6b. Do you think synthetic data can address biases in biomedical research generally and AI specifically? Could you explain your response?
- 7. What do you see as key social issues related to use of synthetic data in biomedical contexts?
- 8. What do you see as legal implications of synthetic data in biomedical contexts?

9. What do you see as policy implications of synthetic data in biomedical contexts?
 - 9a. Who do you think should create synthetic data or make decisions about the creation of synthetic data?
10. What challenges do you anticipate for addressing the ethical, legal, and social implications of synthetic data?
11. How do you think synthetic data relates to real-world data and clinical trial data?
 - a. For example, do you think one type of data is preferable to another? If so, how or why?
 - b. Advocates for synthetic data have expressed either (1) that synthetic data is a good substitute to actual data or (2) that synthetic data is most useful as a supplement to actual data. Do you have a view on this one way or another?
12. What are your thoughts regarding institutional review boards (IRBs) oversight's role of the creation and use of synthetic data in biomedical research? Is it beneficial? Does it have limitations?
13. Overall, do you think synthetic data is a promising new approach or one we should be worried about, or does its promise depend upon the right ethical and policy safeguards?

[If prompts are needed throughout the interview]: How does synthetic data relate to Bias? Privacy? Consent? Safety? Accountability? Equity?

About ELSI-focused Computational Checklists

[Questions will cover matters regarding benefits, barriers, and other challenges in the development, implementation, and evaluation of ELSI-focused computational checklists and other forms of automated assessments on ethics adherence by AI developers and users] We now want to shift our conversation a bit and look beyond the focus on synthetic data. Standards for documenting and communicating details about shared AI objects (such as datasets and models) are of increasing importance for promoting reproducibility and their responsible re-use.

14. What ethical dimensions, elements, or aspects of AI/ML objects (such as datasets and models) do you think AI/ML developers and users should be thinking about when explaining AI objects?
 - a. For example, are there items that should be disclosed with the development and release of any new algorithm or predictive model?
 - b. Or are there ethical considerations that should be reviewed as part of due diligence when
 - i. deciding to use an algorithm or predictive model?
 - ii. Evaluating performance of an algorithm or predictive model?
15. How familiar are you with computational checklists for AI/ML developers to describe predictive models in standardized ways? Do you have any thoughts about the utility of computational checklists specifically for the ethical issues involved with AI objects?
16. What benefits do you envision an ethics-focused or ELSI-focused (meta)data accompanying AI objects could bring? (For example, could standardized documentation and communication of this type of information be helpful? Could it alleviate/mitigate any problems with AI?)
17. What challenges do you envision an ELSI-focused computational checklists created and used in the biomedical AI community could bring? (For example, would it affect the amount of attention given to ethical issues, would adherence or disregard to such items create liability risks, etc.)
18. [If no challenges, consider SKIPPING this question] What barriers do you envision an ethics-focused or ELSI-focused (meta)data accompanying AI objects could bring? (For example, do you have concerns about documentation burdens on AI/ML researchers? Do you have concerns about how such disclosures would be audited or validated?)
19. If ELSI-focused checklists were developed, who should be involved in that process? (e.g., only ethicists, only AI/ML developers, etc.)
20. Should checklists be enforced by anyone? If so, how? (e.g., should funding agencies enforce this? Should this be self-enforcement from within the AI/ML community? Should institutions where AI/ML objects are created or used enforce these – whether IRBs or other groups?)

About AI Ethics Broadly

[Questions will cover matters regarding key ethical issues, social implications, policy implications, and challenges and/or opportunities in addressing these matters in biomedical AI]

21. What is your perspective on the development of an AI Bill of Rights? Do you have any thoughts about what should be included, the process used to develop such an instrument, or its likelihood of success either when thinking about biomedical AI or synthetic data? Do you have any thoughts about the Blueprint for an AI Bill of Rights recently issued by the White House Office of Science & Technology Policy?

Catch-All Question

22. Is there anything else that we haven't covered during this interview that you think we should know about if we want to learn more about the ethical dimensions of synthetic data and means of promoting ethical and responsible AI practices?

THANK YOU/CLOSING

- Thank you so much for sharing your perspectives with us. I am going to stop the recording now.
- **[STOP RECORDING NOW]**
- We are offering our informants a \$100 e-card. If you would like to receive that gift card, please provide your name and email address so that we can get that payment processed.
- Is there anyone else that you think we should interview as part of this project?
- Thank you again for your participation. Have a wonderful day. If you have any questions later, please don't hesitate to reach out to me. **[END CALL]**