Course Concepts: Informed Consent, Diversity and Bias, Fairness, Transparency, Accountability and Governance, Data Privacy, Data Protection, Data Retention and Disposition

Domain: Chatbots in Healthcare

Ethical Checklist for Chatbots in Healthcare

1. Do we have a mechanism for gathering informed consent from users? (Loukides et al., 2018; Stanhaus, 2022a)
2. Have we appropriately adapted our informed consent procedures to our training population, including possible use of surrogate consent? (Shen et al., 2022)
3. Have we ensured that data used in development includes underrepresented groups? (Sundareswaran & Sarkar, 2020; Sandvig, 2020a; Raji, 2020)
4. Have we studied and understood possible sources of bias in our data? (Loukides et al., 2018; Sandvig, 2020a)
5. Do we have a diverse test group (and a diverse development team) representative of the target population group? (Sundareswaran & Sarkar, 2020)
6. Has our project plan addressed potential inequities in access, for instance varying levels of access to mobile technology and to health care services? (Sundareswaran & Sarkar, 2020)
7. Do we have a provision in place to make the model, data source, and collection methodology available for review by the Medical Ethics Commission of this jurisdiction? (Sundareswaran & Sarkar, 2020; Sandvig, 2020b)
8. Do we publish an explanation of the decision or inference, in plain language, infographic or video? (Sundareswaran & Sarkar, 2020; Sandvig, 2020b)
9. Do we publish the limitations of the chatbot (possibility of errors and consequences) its reliability? (Sundareswaran & Sarkar, 2020; Sandvig, 2020b)
10. Do we have a provision in place to inform users when the chatbot fails to understand the user? (Sundareswaran & Sarkar, 2020)
11. Do we have a plan in place to human audit the chatbot at least once every year to maintain its current status and accuracy? (Sundareswaran & Sarkar, 2020; Sandvig, 2020c)
12. Do we have a mechanism for redress if users are harmed by the results? (Loukides et al., 2018)
13. Do we have a written policy on data deidentification and user privacy that is consistent with best practices in medicine? (Shen et al., 2022; Stanhaus, 2022b)
14. Have we ensured compliance with state, federal and international laws governing our chatbot, HIPAA privacy requirements, state data privacy laws, and applicable international privacy laws? (Shen et al., 2022; Office of Civil Rights, U.S. Department of Health and Human Services, retrieved 2022)
15. Have we determined which, if any, third-party vendors will be required to be HIPAA compliant and sign a Business Associate Agreement? (Shen et al., 2022; Office of Civil Rights, U.S. Department of Health and Human Services, retrieved 2022)
16. Have we installed safeguards to identify abnormal behaviour of the chatbot and prevent manipulation? (Sundareswaran & Sarkar, 2020)
17. Do we have a plan in place to perform regular security reviews of the chatbot? (Sundareswaran & Sarkar, 2020)
18. Have we developed a retention policy for conversation data? (Sundareswaran & Sarkar, 2020; Sandvig, 2020d)
19. Do we have a provision of opt-in/opt-out for users to manage their stored data in our system? (Sundareswaran & Sarkar, 2020; Sandvig, 2020d)
20. Have we created accountability for stored conversation data and penalties for leaks/losses? (Sundareswaran & Sarkar, 2020; Sandvig, 2020c)

Design Document

I, along with my data science team, have been tasked with the development of an AI-powered chatbot product for use in healthcare. Essentially, the chatbot would converse with the user, asking for symptoms they experience, and would offer probable diagnoses and recommendations for treatment based on the symptoms.

To that effect, I have created an ethical checklist to accomplish the following broad goals:

1. To direct our workflows through all stages of the chatbot product development – data collection, data storage, analysis, modelling, deployment, and maintenance – such that they adhere to ethical principles by "connecting principle to practice" (Loukides et al., 2018).
2. To ensure that the actions of the chatbot shall not result in avoidable harm to humans or other unintended consequences, including lack of respect for diversity (Sandvig, 2020a).
3. To ensure that the chatbot is fully verified for the efficacy of its purported service, in compliance with accepted international standards given the sensitive nature of the medical information being communicated (Sandvig, 2020c).
4. To enable appropriate safeguards, protection, and disposal of private data collected with informed consent (Sandvig, 2020e; Stanhaus, 2022a).
5. To ensure that a person within the organization is held accountable in case of any intended or unintended consequences arising out of usage of the chatbot (Sandvig, 2020c).

To create this ethical checklist, I have adapted points from multiple sources, such as those from a brochure on the responsible use of conversational AI published by the World Economic Forum (Sundareswaran & Sarkar, 2020), an article from the Journal of Medical Internet Research exploring ethics in digital health research (Shen et al., 2022), and Chapter 2 of the book Ethics and Data Science (Loukides et al., 2018) which discusses some salient points to keep in mind for data science projects.

The points from the latter source have been appropriated as is, mainly in relation to pitfalls in data sampling, biases, and redressal for harm caused, if any. I didn't think these needed to be modified in any way as they seemed to fit my purposes as they stand. The former two sources were my main inspirations for the ethical checklist, as they seemed to perfectly encapsulate the specific ethical problems one might encounter with chatbots in healthcare. I mainly appropriated points relating to informed consent of users who would use the chatbot for medical recommendations (Stanhaus, 2022a), diversity and fairness to ensure that underrepresented groups are included and the team working on this product is diverse (Sandvig, 2020a; Raji, 2020), transparency with respect to the model and the data and external reviews of the product by the Medical Ethics Commission (Sandvig, 2020b; Sandvig, 2020c), accountability in case something goes wrong (Sandvig, 2020c), privacy (Stanhaus, 2022b), data retention, protection and disposition (Sandvig, 2020d; Sandvig, 2020e), and governance (Sandvig, 2020c), with slight modifications in wording so that it fits the "checklist" style of language. Moreover, combining these points ensured that I covered all bases as some sources were missing certain broad issues I wanted to be covered, which the other sources made up for, essentially ensuring that they complement each other.

I have picked these sources (mainly the former two) as inspirations for my ethical checklist for the following reasons:

1. They are deeply rooted in the domain of the specific product I have undertaken and are relevant to fit the specific nuances re chatbots and conversational AI in healthcare.
2. They cover all bases with respect to the ethical usage of the data involved for this product, such as how they should be collected, handled, stored, and then disposed of.

3. They focus on the prevention of some very specific ethical issues typically faced by data science-driven products, such as the gathering of informed consent, respect of data privacy and protection, fairness, transparency, accountability, and compliance with acceptable universal ethical standards (such as the HIPAA framework (Office of Civil Rights, U.S. Department of Health and Human Services, retrieved 2022), which is especially relevant in the medical context).

References

1. Loukides M, Mason H, Patil DJ. (2018). *Ethics and Data Science.* O'Reilly Media, Inc.
2. Office of Civil Rights, U.S. Department of Health and Human Services. (retrieved 2022). HIPAA for Professionals. HHS.gov. https://www.hhs.gov/hipaa/for-professionals/index.html
3. Raji D. (2020). How our data encodes systematic racism. *MIT Technology Review.* https://www.technologyreview.com/2020/12/10/1013617/racism-data-science-artificial-intelligence-ai-opinion/
4. Sandvig C. (2020a). Cumulative disadvantage and protected classes. *Coursera.* https://www.coursera.org/learn/siads503/lecture/5yC8Q/cumulative-disadvantage-and-protected-classes
5. Sandvig C. (2020b). How Transparency Works, and Doesn't Work. *Coursera*. https://www.coursera.org/learn/siads503/lecture/Lz25w/how-transparency-works-and-doesn-t-work
6. Sandvig C. (2020c). Example Technique: External Auditing. *Coursera.* https://www.coursera.org/learn/siads503/lecture/fy5a6/example-technique-external-auditing
7. Sandvig C. (2020d). Case Studies: Data Brokers, AOL Searcher, Clearview AI. *Coursera.* https://www.coursera.org/learn/siads503/lecture/Popnr/case-studies-data-brokers-aol-searcher-clearview-ai
8. Sandvig C. (2020e). Case Study: Mugshots.com. *Coursera*. https://www.coursera.org/learn/siads503/lecture/H3wTW/case-study-mugshots-com
9. Shen F, Silverman B, Monette P, Kimble S, Rauch S, Baker J. (2022). An Ethics Checklist for Digital Health Research in Psychiatry: Viewpoint. *J Med Internet Res*. 24(2):e31146. https://www.jmir.org/2022/2/e31146
10. Stanhaus A. (2022a). Informed Consent. *Coursera.* https://www.coursera.org/learn/siads503/lecture/WF5gk/informed-consent
11. Stanhaus A. (2022b). Privacy Principles. *Coursera*. https://www.coursera.org/learn/siads503/lecture/cFN6X/privacy-principles
12. Sundareswaran V, Sarkar A. (2020). Chatbots RESET: A Framework for Governing Responsible Use of Conversational AI in Healthcare. *World Economic Forum.* https://www3.weforum.org/docs/WEF_Governance_of_Chatbots_in_Healthcare_2020.pdf