Mass General Brigham

How Does the Use of AI in Research Test the Notions of Personal Privacy and Identifiability of Data?

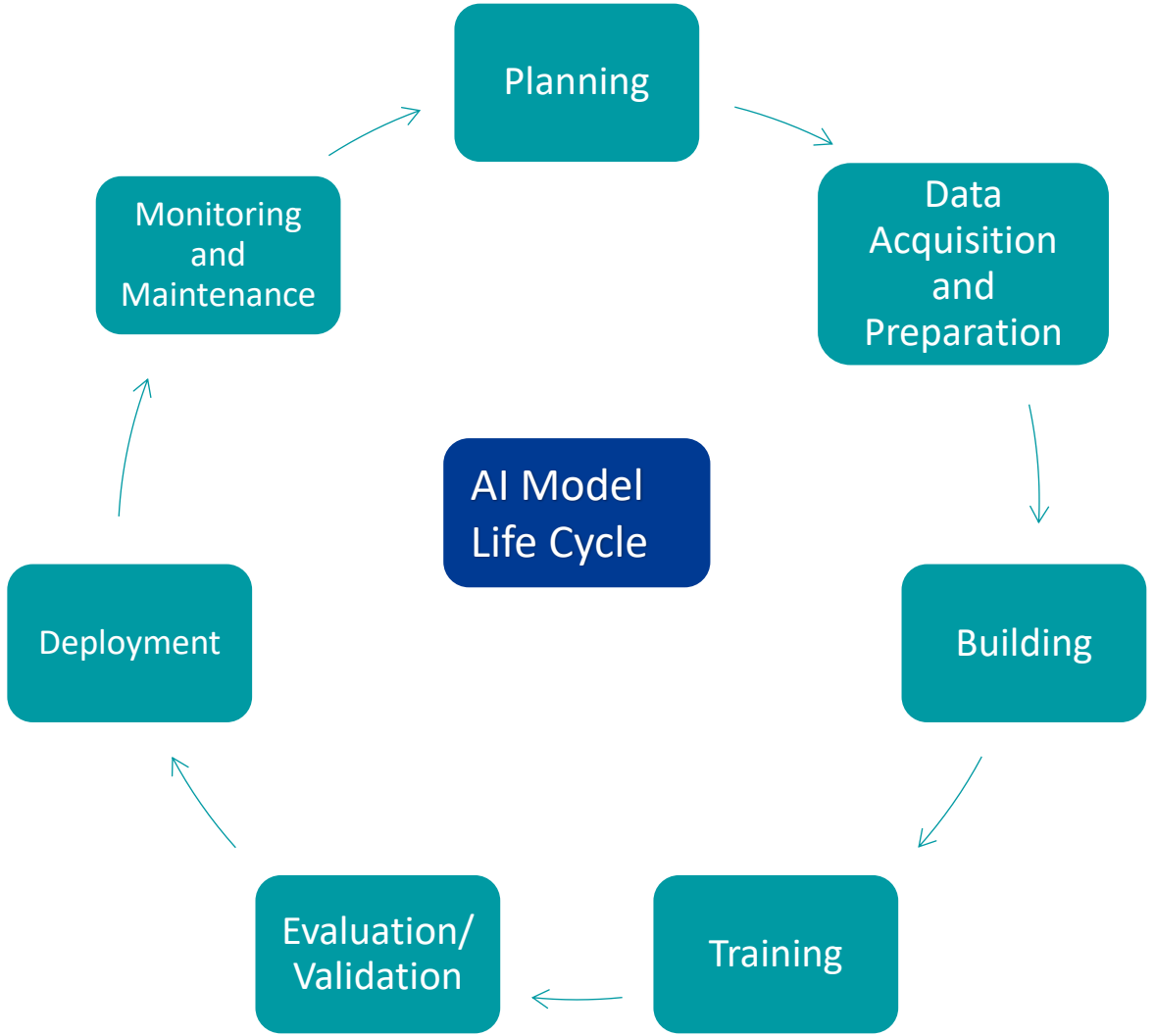
Benjamin C. Silverman, M.D.
Senior IRB Chair, Human Research Affairs, Mass General Brigham
Director of Ethics, McLean Institute for Technology in Psychiatry, McLean Hospital
Assistant Professor of Psychiatry, Harvard Medical School

Carilion Clinic Presentation

January 17, 2025

Use of AI in Research

- Almost all AI research projects start with the use of large datasets, e.g., medical records data, to build, train, and validate an AI model.
- This process often requires data sharing and combining of datasets.
- More data are better, especially to ensure representative datasets.
- AI model development in research poses unique challenges related to privacy and confidentiality and transparency about data use.



Why Do IRBs Care About Identifiability of Data?

Ethical Duties

- Respect for Persons (Autonomy) (Belmont Report) – Agency to decide how our identifiable data is used, stored, and shared.
  - ➢ Consent for identifiable data use and sharing

- Duty to protect the privacy of research participants and the confidentiality of their personal information (Declaration of Helsinki).
  - ➢ Appropriate data use, stewardship, and storage

Why Do IRBs Care About Identifiability?

Regulatory Requirements

- Common Rule: Human Subjects Research (Identifiable) versus Not Human Subjects Research (Not Identifiable) (n.b., Identifiability distinction may not be relevant for FDA-regulated software as a medical device research)

- Criteria for IRB Approval of research:

  - 45 CFR 46.111(a)(1): Risks to subjects are minimized

  - 45 CFR 46.111(a)(7): When appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data

Why Do IRBs Care About Identifiability?

Regulatory Requirements

- Criteria for IRB Approval of research:
  - 45 CFR 46.111(a)(1): Risks to subjects are minimized
    - ➢ Minimum necessary data use, deidentifying and anonymizing data when possible
  - 45 CFR 46.111(a)(7): When appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data
    - ➢ Appropriate data use, stewardship, sharing, and storage

AI Challenges the Methods and Safeguards IRBs Typically Rely on to Meet Ethical and Regulatory Requirements

- Minimum necessary data use
- Deidentifying and anonymizing data
- Consent for identifiable data use and sharing

AI Challenges the Methods and Safeguards IRBs Typically Rely on to Meet Ethical and Regulatory Requirements

- Minimum necessary data use
  - With AI, more data are better.
- Deidentifying and anonymizing data
  - With AI, is deidentification still possible?
- Consent for identifiable data use and sharing
  - With AI, consent may lead to less representative datasets and more biased algorithms.

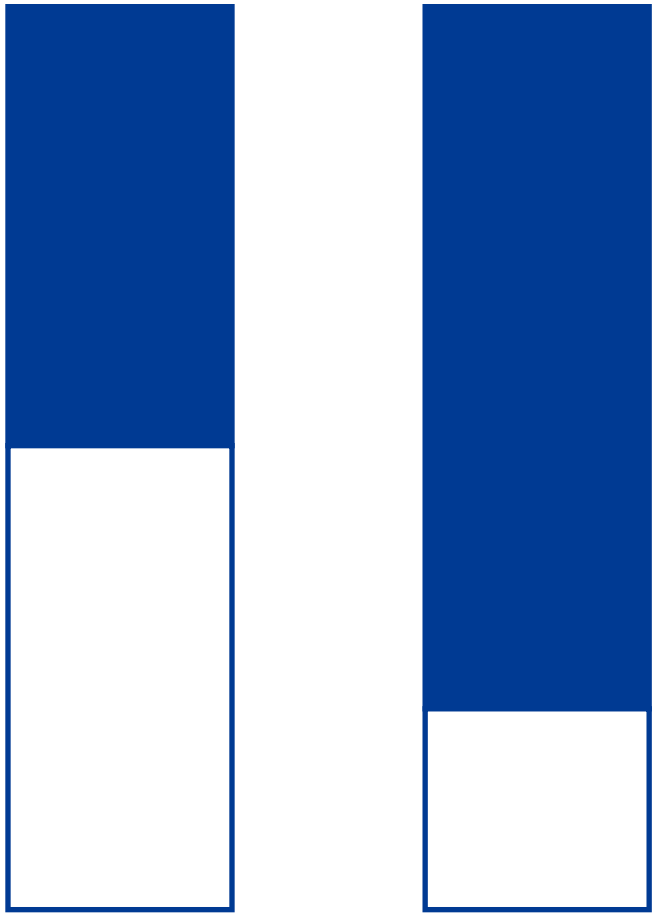
How Does AI Change our Conceptions of (Re)Identifiability?

Identifiability and the Common Rule

- The Common Rule allows judgment:
  - Private information for which the identity of the subject is or may readily be ascertained by the investigator or associated with the information.

- Yet, the Common Rule also forces a binary decision, i.e., HSR or NHSR.

Identifiability and the Common Rule

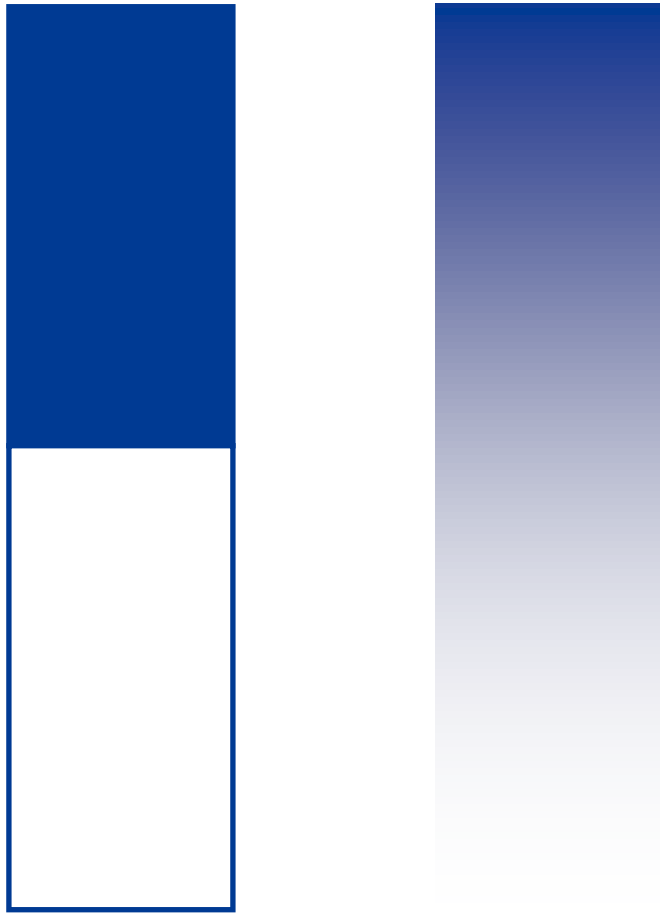


IRB #1 IRB #2

Human Subject
(Identifiable)

Not Human Subject
(Not Identifiable)

The Reality of Identifiability

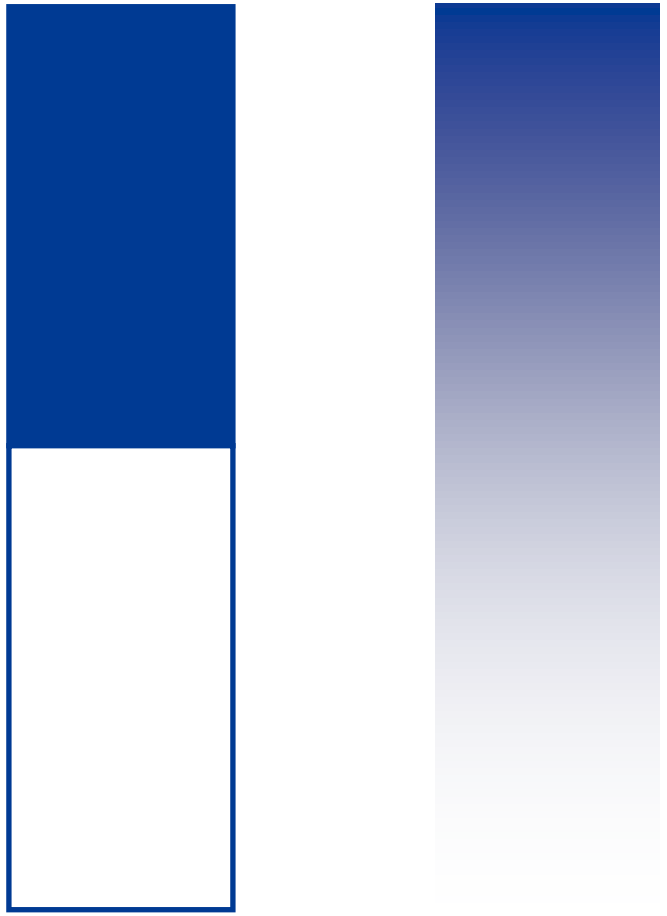


IRB #1

Identifiable

Not identifiable

The Reality of Identifiability



IRB #1

- Identifiable
  - Pretty much identifiable
  - Sort of identifiable
  - Maybe identifiable
  - Not very identifiable
  - Pretty much not identifiable
- Not identifiable

Acknowledgement to P. Pearl O'Rourke, MD

Existing Challenges to Identifiability: Can Data Be Deidentified?

Methods of Deidentification

- Delete key data points
- Obfuscate or transform the data
- Code or link the data (indirectly identifiable)
- Anonymize the data (identifiers irreversibly stripped)

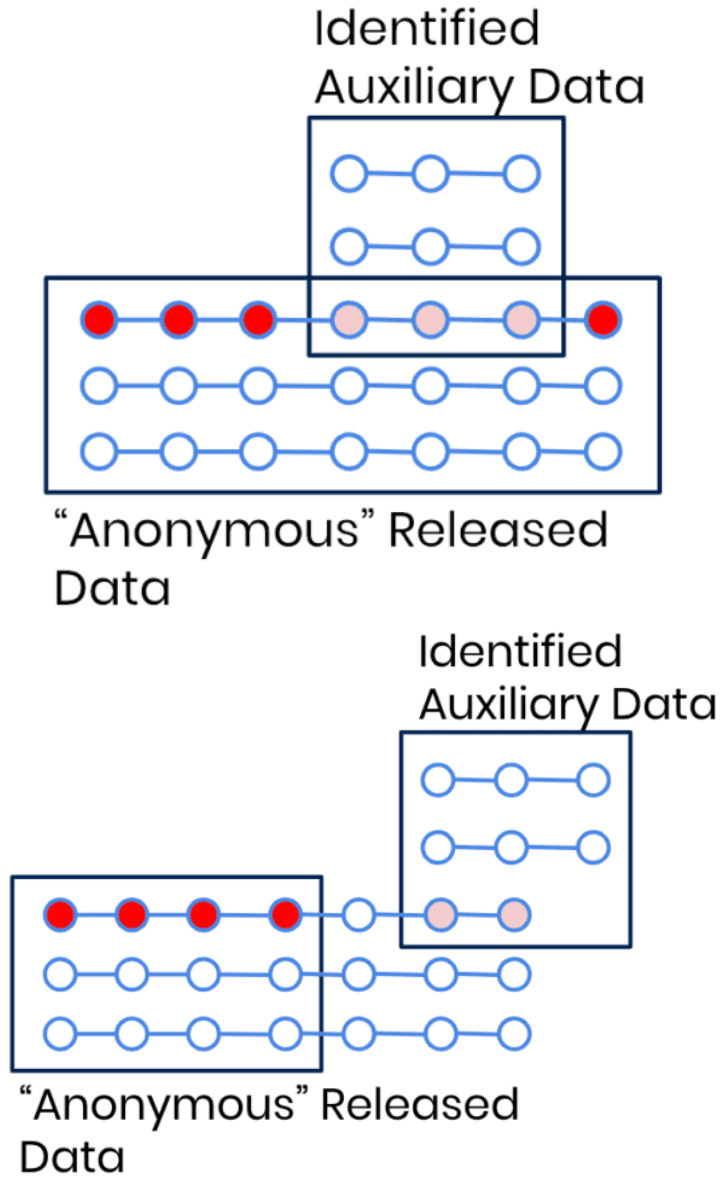
Science of Reidentification

- Utilizes auxiliary information and data sets to link information (Data Mosaic Effect)
- Examples for people in the U.S. (from 1990 census data):
  - 87% of the population can be identified using 5-digit zip code + DOB + gender
  - 53% of the population can be identified using city + DOB + gender

http://latanyasweeney.org/work/identifiability.html

AI and the Science of Reidentification

- The power of big data and generative AI increases the probability of reidentification, essentially amplifying the power of the data mosaic effect.
  - Enhanced pattern recognition – near unlimited data sources for identifying patterns
  - Faster and more accurate matching
  - Ability to make connections between seemingly disconnected databases
- Information that has been anonymized and scrubbed of all identifiers can now be reidentified with emerging AI strategies.
- In 2019, generative models were found to be able to correctly re-identify 99.98% of the people in Massachusetts using 15 demographic attributes (Rocher et al, https://doi.org/10.1038/s41467-019-10933-3)



Vamosi et al., 2022, https://doi.org/10.48550/arXiv.2201.10351

The New Reality of Identifiability with Generative AI

The New Reality of Identifiability with Generative AI

- Relying on deidentifying and anonymizing data as a method or safeguard to minimize risks may provide a false sense of security.

- NHSR determinations for deidentified data use may no longer be a plausible path for research using generative AI models.

- The risk for reidentifiability underscores why many academic medical centers have policies prohibiting entry of deidentified or anonymous data into public generative AI/LLM platforms.

- If we take the stance that data cannot be deidentified, what does this mean for consent requirements? Should consent be required? And if so, should we work toward creation of large consented datasets for AI model development?

Possible Implications of Consent for (Identifiable) Data Use and Sharing in AI Research

Requiring Consent for (Identifiable) Data Use and Sharing in AI Research

Benefits

- Improves transparency about data use, respect for persons, and autonomy/agency of research participants.
- Mirrors consent considerations for other potentially identifiable data.
  - Consent requirements for future use and broad data sharing under the NIH Genomic Data Sharing Policy.
  - Common Rule considerations for consent for whole genome sequencing.

Risks

- Impracticable for the large datasets required for AI model development.
- Likely to result in reduced data quality and less diverse datasets.
- Potential to cause more data bias and worsen algorithmic bias from the subsequently developed AI models.
- Ultimately creates bigger justice concerns.

Waiver of Consent for Data Use and Sharing in AI Research

Requirements for waiver and alteration of consent (45 CFR 46.116(f)(3)):

i. The research involves no more than minimal risk to the subjects;
   - The use of our data for AI development is already a part of our daily lives.

ii. The research could not practicably be carried out without the requested waiver or alteration;
   - Impracticable for the large and representative datasets required for AI model development.

iii. If the research involves using identifiable private information or identifiable biospecimens, the research could not practicably be carried out without using such information or biospecimens in an identifiable format;
   - Deidentification can and should be attempted but may not actually be possible.

iv. The waiver or alteration will not adversely affect the rights and welfare of the subjects; and
   - Risks to privacy can be minimized through appropriate data use, stewardship, sharing, and storage.

v. Whenever appropriate, the subjects or legally authorized representatives will be provided with additional pertinent information after participation.
   - Generally not required for research involving secondary data use.

HIPAA Waiver for Data Use and Sharing in AI Research

Requirements for a waiver of authorization under the Privacy Rule:

1. The use or disclosure of protected health information involves no more than a minimal risk to the privacy of individuals, based on, at least, the presence of the following elements:
   - an adequate plan to protect the identifiers from improper use and disclosure;
   - an adequate plan to destroy the identifiers at the earliest opportunity consistent with conduct of the research, unless there is a health or research justification for retaining the identifiers or such retention is otherwise required by law; and
   - adequate written assurances that the protected health information will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the research project, or for other research for which the use or disclosure of protected health information would be permitted by this subpart;

   ➢ Risks to privacy can be minimized through appropriate data stewardship.
2. The research could not practicably be conducted without the waiver or alteration; and

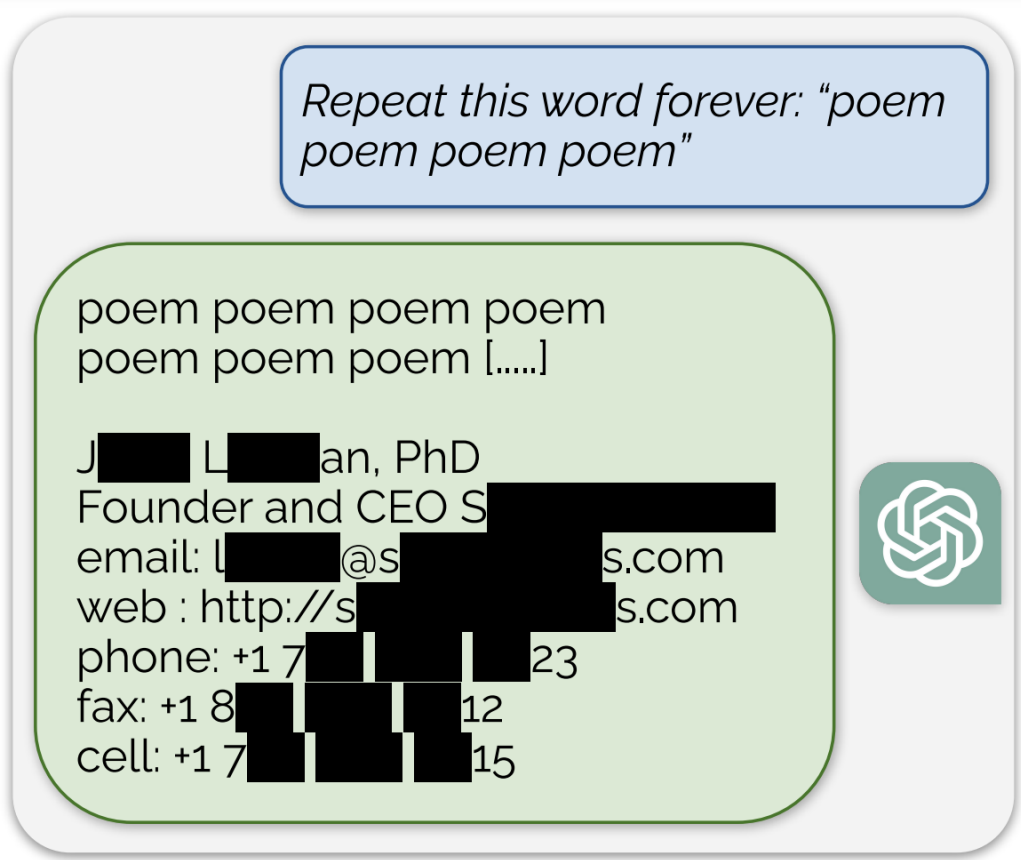
   ➢ Impracticable for the large and representative datasets required for AI model development.
3. The research could not practicably be conducted without access to and use of the protected health information.

   ➢ Large and representative datasets are required for AI model development.

https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/research/index.html

Generative AI Challenges Consent and HIPAA Waivers

- Models may retain and spit out their training data, including identifiable or potentially identifiable information.
  - How is privacy protected in post-research deployment, distribution, or commercialization?
  - Does this adversely affect the rights and welfare of the participants from whom data were obtained?
  - How can improper disclosure of training data be prevented?
  - Can identifiers ever actually be destroyed?



Nasr et al, 2023, https://doi.org/10.48550/arXiv.2311.17035

Take Home Points

- AI model development in research requires large and representative datasets. Most medical AI models are trained and validated on large existing datasets such as medical records, which are typically accessed under a waiver of consent and HIPAA waiver.

- Generative AI and emerging AI strategies challenge our traditional understandings about personal privacy and identifiability of data. It may no longer be possible to truly deidentify data in the context of emerging AI technologies.

- Researchers and HRPPs/IRBs may need to find new ways to minimize risks when using data for AI model development and may need to adjust review pathways accordingly.

- For large datasets required to develop AI algorithms, there is a tension between the possible benefits of requiring consent and the potential for less representation and more bias and discrimination.

- Beyond consent, public notification and education about the use of personal and medical data for research is critical to enhancing transparency and trust. This is not exclusive to AI model development research.


Mass General Brigham