



# Responsible research in artificial intelligence: lessons from the past

Puneet Sharma<sup>1</sup>

Received: 13 February 2024 / Accepted: 20 March 2024

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

We are still at the beginning of the artificial intelligence (AI) revolution in medicine, but it is already clear this technology will impact the way we deliver care. In my field of neonatology alone, deep learning models have been shown to predict sepsis, bronchopulmonary dysplasia, and necrotizing enterocolitis. These models are largely trained on patient data extracted from the electronic medical record. In other words, they use the data of previous patients to predict the behavior of future patients. This methodology, while effective, raises an ethical question: to whom does these data belong? This is a question being asked in many disciplines currently. A group of famous authors sued OpenAI for infringing on their copyrights while training ChatGPT. Similarly, a central issue in the actor's strike last year was the use of AI to generate the likeness and voice of actors for future roles. Despite these debates taking forefront in the culture, the same concerns have not yet manifested in medicine.

A large portion of medical knowledge is learned from patients, through both observational studies and controlled trials. While much of this research requires informed consent, institutional review boards can exempt investigators from this process as consent may not be feasible. Many AI models fall in this category due to the large amount of retrospective data needed to train them. As such, the patients whose data are used in these models are not always aware and may not have given consent. Of course, researchers take every measure to maintain the security and protect the privacy of these patients, but the point remains. While not explicitly unethical, I do believe if we learn from the past we can do better by patients.

The story of Henrietta Lacks is well known the medical community. In her book *The Immortal Life of Henrietta Lacks*, Rebecca Skloot described how Henrietta had her cervical cells harvested without her knowledge prior to

passing away from cervical cancer in 1951. These so-called HeLa cells were the first immortal human cell line, playing instrumental role in medical research and the generation of over 11,000 patents. This went on for decades without her family being aware until they were solicited by researchers for samples after the accidental contamination of HeLa cells. There have since been attempts to right this wrong. Henrietta has been formally recognized by the World Health Organization, National Institutes of Health, and other organizations. Her family has been given some control over access to the genome of her cells and they have settled with companies for profiting of her cells.

The case of Henrietta Lacks is obviously not identical to modern research using artificial intelligence. No physical samples are taken from patients, and the models are not derived from one patient but rather aggregated, de-identified records. However, it is not hard to see parallels. Like Henrietta, the patients may not know their data are being used and the models will outlive the patients themselves. Furthermore, AI has the potential to be just as impactful to medical care as the HeLa cells, not to mention a similar financial impact. As such, it is important we learn from Henrietta's story and place a priority on consent, recognition, and compensation in AI research.

The principle of informed consent is paramount in medicine, but we must be pragmatic with AI research. Obtaining informed consent from every patient is not reasonable nor can it be expected with the volume of retrospective data needed to train these models. However, it is feasible to disclose to patients admitted to the hospital or seen in clinic that their data may be used in the training and development of AI models. While not explicitly obtaining consent, it is a practice that keeps our patients informed of their role in clinical research and patient care. Recognition, on the other hand, is a little more straightforward. We should commit to idea that any publication has a formal acknowledgment to the patients and/or families that contributed their data to the model. The final lesson to be learned from Henrietta is compensation; this is similarly challenging as it is not just one patient but an aggregate. As such, we may not be able to personally

---

✉ Puneet Sharma  
puneet.sharma@childrens.harvard.edu

<sup>1</sup> Division of Newborn Medicine, Boston Children's Hospital, 300 Longwood Avenue | Enders 961, Boston, MA 02115, USA

compensate every patient whose data were used to develop an AI product or business. However, we can make sure we impact the larger community in which the patients reside. As such, we can commit to the principle that any product trained on patient data is accessible to the community from which it was born and, therefore, improve the health of these patients. With these basic principles, we can do right by our patients and ensure that AI does not become the 21st version of HeLa cells.

**Curmudgeon Corner** Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

**Funding** This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (5T32HD098061).

**Data availability** Not applicable.

## Declarations

**Conflict of interest** The authors have no competing interests to declare that are relevant to the content of this article.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.