Published in final edited form as:

Am J Bioeth. 2020 November ; 20(11): W6–W10. doi:10.1080/15265161.2020.1827695.

A framework to evaluate ethical considerations with ML-HCA applications—valuable, even necessary, but never comprehensive

Dr Danton Char,

Stanford, Anesthesia, Stanford, United States

Dr Michael Abramoff,

University of Iowa, Ophthalmology and Visual Sciences, 200 Hawkins Drive, Iowa City, 52242-1002 United States

Dr Chris Feudtner

The Children's Hospital of Philadelphia, Philadelphia, 19104-4399 United States

Machine Learning (ML) is fundamental to multiple visions of healthcare's future, from precision medicine (Petrie Flom 2020, Frohlich 2018) to a model of health delivery and research as a combined 'learning healthcare system' (Davenport 2019). Not surprisingly, as awareness of ethical failings with ML applications outside of healthcare grow (Char 2018), there has been expanding concern about ethical challenges applying ML tools within healthcare (HAI 2019, Rajkomar 2018, Obermeyer 2016, Obermeyer 2019). In fact, ethical concerns with ML use was a significant part of the F.D.A.'s scrutiny prior to granting approval for the first autonomous diagnostic system based on ML (Abramoff 2020). Failure to adequately anticipate and address ethical problems before they emerge has set back other promising technologies for healthcare, such as stem cell therapies (Lo 2009).

Our approach to identifying and considering ethical problems with ML for healthcare applications (ML-HCA) relies on six premises. First, multiple stakeholders are impacted by any ML-HCA. These stakeholders can be identified by examining the contexts into which ML-HCAs are designed and deployed. Second, stakeholder groups are likely to have different values, and significantly different explicit or implicit goals for the ML-HCA, that should and can be ascertained. Third, the process of design and development of an ML-HCA involves making a series of decisions, from initially conceiving of the problem to be addressed by an ML-HCA to actually deploying and maintaining it. Fourth, how a stakeholder makes these decisions, or would want these decisions to be made, reflects their underlying values. In fact, such decisions are often value judgments (Shilton 2018). Fifth, where stakeholder groups disagree or their values are at odds about resolving these decisions—where values collide—are where ethical problems are most likely to emerge (Brey 2012, Shilton 2018). Sixth, while some of these value collisions may mark novel ethical concerns, many are likely able to be resolved by drawing on prior ethical scholarship on similar or related problems.

Given these premises, a mental map of the process from conception to deployment of an ML-HCA, providing a reasonably detailed overview of the major decisions that need to be made at each stage of the process—and highlighting the potential value collisions and the ethical considerations for each of these decisions—is a valuable framework with which to evaluate any ML-HCA for current or future ethical concerns. While we would never argue that our particular mental map is comprehensive and exhaustive, we do argue that ethical evaluations will be improved if they are explicit about the mental map being used, and that the mental map we propose is useful in many regards. We welcome dialogue about examining ethical concerns with ML for healthcare applications (ML-HCA) and wish to enhance understanding for the framework we are proposing.

Risk, Impact on Workflow, and Regulation

Significant safety and judgment failures have occurred around implementation of ML systems or output for work processes in non-healthcare contexts (Nicas 2019, Boudette 2018, Efrati 2018). These failures have raised awareness about the potential for ML approaches to cause “disruptive change” within the practice of medicine (Jameson 2015) and the potential for similar failures, particularly around clinician and patient interactions with ML systems or output, and with inadequate *in situ* assessment of ML impact on workflows (Nuffield 2018). The role of risk posed by ML “disruption” (Ho 2020) and the impact of an ML on inter-stakeholder relationships, including workflows, (Kraft 2020) are important considerations. Such considerations require value judgments by multiple stakeholder groups and should be considered in the conception and deployment stages of the pipeline framework (Kraft 2020, Ho 2020). Importantly, while the framework can identify these areas of ethical challenge, we make no claims about what normative or regulatory guidance is needed to address these potential harms (Ho 2020), only that such oversight would need to be done.

We would also raise the caution that ethical and regulatory groups, once established, then themselves become stakeholders with values. While such values are usually benevolent and intend to protect, recent ethics scholarship has called attention to situations where ethical concerns have led to regulatory approaches that, when implemented, caused unethical outcomes by overly restricting access to scarce research samples (effectively limiting research that could be done) or new technologies (Stolberg 2020, Botkin 2013, All of Us 2018).

Distributive Justice and Bias

Several commentators rightly underscored the importance of addressing distributive injustices and bias against vulnerable groups as ML-HCAs are built and deployed (McLennan 2020, Hardt 2020 and Karnik 2020), and characterize strategies to address these concerns. These strategies include trials with pre-specified endpoints and evaluation of patient outcome preferences (Hardt 2020, Abramoff 2020), addressing societal structural inequalities (McLennan 2020), and increasing underrepresented minorities in the ML design community (Karnik 2020).

In these regards, ML-HCAs can be characterized as having both the “reification of bias” problem and the “zero-sum” problem. Regarding the first of these, ML-HCAs can reify existing biases and distributive injustices (McCradden 2020). Healthcare data has known health, wealth, and race disparities woven into it (Obermeyer 2019). Predictions formulated from this data may be accurate but may perpetuate these disparities (Hardt 2020). For example, a mortality prediction algorithm to guide an oncologist’s therapeutic choices may correctly predict that an individual with a health, wealth, or race disparity is likely to die sooner and consequently is a poor candidate for an aggressive experimental chemotherapy. The algorithm is not wrong, but access to a more advanced therapy might be restricted because of these pre-existing disparities (Magnus 2019). Our framework can help to identify this problem (which, importantly, is located not only in the predictive algorithm per se but also, or even more so, in the way that the output is used to make decisions, which is to say how this ML-HCA is implemented). Identification is a step towards problem solving, but subsequently formulating and implementing solutions will require far more extensive involvement by ethicists and health disparities researchers in ML-HCA efforts (McLennan 2020, Hardt 2020, Karnik 2020, Fiske 2020).

Regarding the so-called zero-sum problem (Becker 2018), attempting to correct an ML-HCA’s resource allocation to improve the outcome for a worse-off group risks reducing—or does reduce—the beneficial outcome for the initially better-off group (Goodman 2018, Obermeyer 2019). While our framework can identify this problem, how to address this problem is outside of the intended scope of the framework. We also did not address the primacy of this problem relative to other ethical concerns (Karnik 2020, Hardt 2020). Our framework raises, however, the possibility of such explicit prioritization among identified ethical concerns. Certainly, distributive justice and anti-bias are issues not just for the black box algorithm design itself, but for the preceding and subsequent stages of ML-HCA creation as they are societal problems magnified by the ML-HCA.

Exceptional Ethical Problems

Many commentators highlighted the novelty of ML-HCAs, and how they are in turn giving rise to ethical considerations that have novel elements. Examples include neuro-diagnostic interpretation (Overton 2020), continuous learning (Bæroe 2020), the possible fiduciary role of designers (Lewis 2020), and privacy and safety with non-traditional and expanded sources of patient digital information (Skorburg 2020). We agree that a framework to identify ethical concerns with ML-HCAs needs to be able to do both: to recognize unexceptional, existing concerns that appear in a new context and to also recognize when antecedents do not encompass the new problem. We caution, however, against too much appeal to the exceptional. The perception of too much novelty with ML-HCAs, and concern specifically about this novelty, may itself be a form of ML exceptionalism (Findley 2020), which in turn can blind us to non-novel concerns.

For example, although an ML-HCA may generate a stigmatizing neurologic diagnosis, previous ethical scholarship on stigmatizing diagnoses may be relevant. Although debates such as the choice of continuously learning or locked ML-HCA approaches are important, these debates are informed by previous scholarship on clinical trials, accuracy and harm.

Although prior ethical literature has examined issues raised by bedside use of prognostic scoring systems, much of this literature is centered on the appropriateness of using population probabilities over individual clinician bedside judgment to make care decisions (Luce 1994). Other ethics scholarship has examined concerns around the human-computer interface including the value aims and assumptions in computer-system design (Manders-Huits 2011).

What is especially novel about ML-HCA systems is that, unlike previous generations of medical technology based on complex data but where the interpretation of the output from these technologies required a human to make a diagnosis or recommend a clinical action, some ML-HCAs generate conclusions or recommended actions. As such, these applications encompass all of the ethical concerns of complex statistical and computational tools and also the ethical challenges that weigh on making a diagnosis or clinical decision.

Responsibility and Unintended Consequences

An important consideration with ML-HCAs is who is responsible for identifying and, ultimately addressing, ethical concerns. In the case of autonomous ML-HCAs the responsibility should rest with the system's designer since, by definition, such an application is intended to function autonomously (Abramoff 2020). As Zawati et al. (Zawati 2020) point out, for assistive ML systems the accountability for ethical concerns is much less clear. For many clinicians, contradicting an algorithm's recommendation will likely require them to produce far more and better data than the algorithm's recommendation was based on as they face employment hierarchies (Gerhards 2020, O'Neil 2017). As McCradden et al. discuss (McCradden 2020), ethics research oversight is, at best, only in a partial position to evaluate ML-HCA systems, as the majority seem to be deploying under QI initiatives and are being constructed within private industry, with many of these ML-HCAs targeting process automation and administrative applications (Faggella 2019, Davenport 2019, Char 2018, Challen 2019). While examining ML-HCAs using experimentally controlled studies with pre-specified end-points is clearly the ideal (Abramoff 2020), whether the multi-billion-dollar ML-HCA market will easily submit to such a slowdown in product release and whether controlled studies will be used to evaluate process automation and administrative applications remains to be seen, though some initiatives are starting to address this (CCOI 2020). Certainly, the ethics community needs to be unified in their call for such scrutiny.

A related concern is around unintended uses or consequences of an ML-HCA created for one context but adapted, untested and potentially unmonitored, to another (Swirsky 2020). Our framework identifies this as a potential problem with subsequent implementations. Addressing this problem will require ongoing monitoring, but what body will be responsible for this monitoring still needs to be considered (McCradden 2020, Abramoff 2020, F.D.A. 2018, Zawati 2020). To do this, various types of stakeholders should be involved, but these different stakeholders will likely have variable knowledge and skill to identify the wide range of ethical concerns that ML-HCAs can raise; consequently, ethicists may need to be embedded into design and development teams to facilitate the timely identification of ethical concerns (Fiske 2020). Whether certain stakeholder groups (such as patients) will need to

have their views prioritized relative to other stakeholder groups is an important consideration and may vary by context (Findley 2020).

Adaptability and Agility

Potential unintended uses of an ML-HCA also raise the question of the adaptability (or “agility”) of any framework to identify ethical concerns with an emerging technology (Nabi 2020, Swirsky 2020). In addition to a step-wise development process for a new ML-HCA, the pipeline can be used to situate an already existing ML-HCA at a particular point in the developmental pipeline, and then to identify and examine ethical considerations at that point and in previous steps that might have relevant bearing (Nabi 2020). As ML-HCA developers and users poised at a particular point in the pipeline are actually part of a complex ecosystem of care (Gerhards 2020), the ethical operating characteristics that arise from previous decisions about how the ML-HCA has been constructed, and subsequent ethical characteristics will still have relevance.

This Particular Framework is an Evolving Tool

We thank the commentators for underscoring the importance of having a systematic overview of the ethical considerations with ML-HCAs. Hopefully, at its best, this pipeline framework offers groups of diverse stakeholders a “bigger picture” of ML-HCAs that can, with dialogue, help to forge a shared mental model of the range of relevant questions and ethical considerations that should guide design, evaluation, and implementation decisions. Our collective goal is to identify and address ethical concerns with ML-HCAs before they lead to harm or injustice. To this end, we are grateful for all comments, suggestions, and criticisms by everyone who applies and adapts the framework to systematically scrutinize the ethics of ML-HCAs. We hope this pipeline framework approach will aid such examinations of actual ML-HCAs and the identification of ethical concerns and subsequent development of solutions.

References

- Abraham MD, Tobey D, Char DS. Lessons Learned About Autonomous AI: Finding a Safe, Efficacious, and Ethical Path Through the Development Process. *Am J Ophthalmol*. 2020 06;214:134–142. [PubMed: 32171769]
- All of Us Research program, National Institutes of Health. “Ethical Considerations in the All of Us Research Program.” Accessed online 10 September 2020 at https://www.bioethics.nih.gov/courses/pdf/2018/session5_blizinsky.pdf
- Bærøe Kristine, Jansen Maarten, Kerasidou Angeliki. 2020 The ethics of machine learning in healthcare: Is exceptionalism required after all? *The American Journal of Bioethics*.
- Becker S, Gamble M. Healthcare as a Zero-Sum Game: 7 key points. *Becker’s Hospital Review*, 12 19, 2018 Accessed online 10 September 2020 at <https://www.beckershospitalreview.com/hospital-management-administration/healthcare-as-a-zero-sum-game-7-key-points.html>
- Botkin JR, Goldenberg AJ, Rothwell E, Anderson RA, Lewis MH. Retention and research use of residual newborn screening bloodspots. *Pediatrics*. 2013;131(1):120–127. doi:10.1542/peds.2012-0852 [PubMed: 23209103]
- Boudette N Fatal Tesla Crash Raises New Questions About Autopilot System. *New York Times* 3 31, 2018 Accessed online 11 September 2020 at <https://www.nytimes.com/2018/03/31/business/tesla-crash-autopilot-musk.html>

- Brey Philip A. E. 2012 “Anticipatory Ethics for Emerging Technologies.” *NanoEthics* 6 (1): 1–13. 10.1007/s11569-012-0141-7.
- Char Danton S., Shah Nigam H., and Magnus David. 2018 “Implementing Machine Learning in Health Care - Addressing Ethical Challenges.” *The New England Journal of Medicine* 378 (11): 981–83. 10.1056/NEJMp1714229. [PubMed: 29539284]
- Collaborative Community on Ophthalmic Imaging (CCOI) 2020 Conference, “The Future of Artificial Intelligence-Enabled Ophthalmic Image Interpretation: Accelerating Innovation and Implementation Pathways.” September 3–4 2020, <https://www.cc-oi.org/conference-program>
- Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J.* 2019;6(2):94–98. doi:10.7861/futurehosp.6-2-94
- Efrati A Uber Finds Deadly Accident Likely Caused By Software Set To Ignore Objects on Road. *The Information* 5 7, 2018 Accessed online 11 September 2020 at <https://www.theinformation.com/articles/uber-finds-deadly-accident-likely-caused-by-software-set-to-ignore-objects-on-road>
- Faggella D Machine Learning Healthcare Applications – 2018 and Beyond. *Emerj.* 5 19 2019, accessed online at <https://emerj.com/ai-sector-overviews/machine-learning-healthcare-applications/>
- Findley Jess; Woods Andrew; Robertson Christopher; Slepian Marv. 2020 Keeping the Patient at the Center of Machine Learning in Healthcare. *The American Journal of Bioethics.*
- Fiske Amelia; Tigard Daniel; Müller Ruth; Haddadin Sami; Buyx Alena; McLennan Stuart. 2020 Embedded ethics could help implement the pipeline model framework for machine learning healthcare applications. *The American Journal of Bioethics.*
- Fröhlich H, Balling R, Beerenwinkel N et al. From hype to reality: data science enabling personalized medicine. *BMC Med* 16, 150 (2018). 10.1186/s12916-018-1122-7 [PubMed: 30145981]
- Gerhards Helene; Weber Karsten; Bittner Uta; Fangerau Heiner. 2020 Machine Learning Healthcare Applications (ML-HCAs) are no stand-alone systems but part of an ecosystem – A broader ethical and health technology assessment approach is needed. *The American Journal of Bioethics.*
- Goodman Steven N., Goel Sharad, and Cullen Mark R.. 2018 “Machine Learning, Health Disparities, and Causal Reasoning.” *Annals of Internal Medicine* 169 (12): 883–84. 10.7326/M18-3297 [PubMed: 30508423]
- Hardt Michaela; Chin Marshall H. 2020 It is Time for Bioethicists to Enter the Arena of Machine Learning Ethics. *The American Journal of Bioethics.*
- Ho Calvin Wai-Loon. 2020 Deepening the Normative Evaluation of Machine Learning Healthcare Application by Complementing Ethical Considerations with Regulatory Governance. *The American Journal of Bioethics.*
- Human-Centered Artificial Intelligence, Stanford University Accessed online 24 May 2020 at <https://hai.stanford.edu/>
- Karnik Niranjana; Afshar Majid; Churpek Matthew; Nunez-Smith Marcella. 2020 Structural Disparities in Data Science: A Prolegomenon for the Future of Machine Learning. *The American Journal of Bioethics.*
- Kraft Stephanie. 2020 Respect and trustworthiness in the patient-provider-machine relationship: Applying a relational lens to machine learning healthcare applications. *The American Journal of Bioethics.*
- Lewis Anna. 2020 Where bioethics meets machine ethics. *The American Journal of Bioethics.*
- Lo B, Parham L. Ethical issues in stem cell research. *Endocr Rev.* 2009;30(3):204–213. doi:10.1210/er.2008-0031 [PubMed: 19366754]
- Luce JM, Wachter RM. The ethical appropriateness of using prognostic scoring systems in clinical management. *Crit Care Clin.* 1994 1;10(1):229–41. [PubMed: 8118731]
- Manders-Huits N What values in design? The challenge of incorporating moral values into design. *Sci Eng Ethics.* 2011;17(2):271–287. doi:10.1007/s11948-010-9198-2 [PubMed: 20224927]
- Magnus D Personal interview. 9 26, 2019
- McCradden M, Joshi S, Mazwi M, Anderson JA. Ethical limitations of algorithmic fairness solutions in health care machine learning. *Lancet Digital Health.* 2020; 2(5): 221–223

- McCradden Melissa; Anderson James; Zlotnik Shaul Randi. 2020 Accountability in the machine learning pipeline: the critical role of research ethics. *The American Journal of Bioethics*.
- McLennan Stuart; Lee Meredith; Fiske Amelia; Celi Leo. 2020 AI ethics is not a panacea. *The American Journal of Bioethics*.
- Nabi Junaid. 2020 Addressing the “wicked” problems in machine learning applications – Time for bioethical agility. *The American Journal of Bioethics*.
- Nicas J, Glanz J, Gelles D. “In Test of Boeing Jet, Pilots Had 40 Seconds to Fix Error.” *New York Times*, 3 25, 2019 Accessed online 17 April 2019 at <https://www.nytimes.com/2019/03/25/business/boeing-simulation-error.html>
- Nuffield Council on Bioethics. Artificial Intelligence (AI) in Healthcare and Research. Nuffield Council on Bioethics, London 2018 Accessed online 17 April 2019 at <http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf>
- Obermeyer Ziad, and Emanuel Ezekiel J.. 2016 “Predicting the Future — Big Data, Machine Learning, and Clinical Medicine.” *New England Journal of Medicine* 375 (13): 1216–19. 10.1056/NEJMp1606181. [PubMed: 27682033]
- Obermeyer Ziad, Powers Brian, Vogeli Christine, and Mullainathan Sendhil. 2019 “Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations.” *Science* (New York, N.Y.) 366 (6464): 447–53.
- O’Neil C The Ivory Tower Can’t Keep Ignoring Tech. *New York Times*, 11 14, 2017, accessed online at <https://www.nytimes.com/2017/11/14/opinion/academia-tech-algorithms.html>, December 2, 2017
- Overton Dawson. 2020 An Evaluation of the Pipeline Framework for Ethical Considerations in Machine Learning Healthcare Applications: The Case of Prediction from Functional Neuroimaging Data. *The American Journal of Bioethics*.
- Petrie-Flom Center for Health Law Policy, The Project on Precision Medicine, Artificial Intelligence, and the Law (PMAIL). Accessed online 09 9 2020 at <https://petrieflom.law.harvard.edu/research/precision-medicine-artificial-intelligence-and-law>
- Rajkomar Alvin, Hardt Michaela, Howell Michael D., Corrado Greg, and Chin Marshall H.. 2018 “Ensuring Fairness in Machine Learning to Advance Health Equity.” *Annals of Internal Medicine* 169 (12): 866 10.7326/M18-1990. [PubMed: 30508424]
- Skorburg Joshua. 2020 What counts as ‘clinical data’ in machine learning healthcare applications? *The American Journal of Bioethics*.
- Shilton Katie. 2018 “Values and Ethics in Human-Computer Interaction.” *Foundations and Trends in Human-Computer Interaction* 12 (2): 107–71. 10.1561/11000000073.
- Stolberg SG. “Studies Using Fetal Tissue Face Battle for Funding.” *New York Times*, 8 19, 2020 Accessed online 11 September 2020 at <https://www.nytimes.com/2020/08/18/us/politics/trump-fetal-tissue-research.html>
- Swirsky Eric; Gu Carol; Boyd Andrew. 2020 An Ethical Framework to Nowhere. *The American Journal of Bioethics*.
- United States Food and Drug Administration (FDA) News Release. FDA Permits Marketing of Artificial Intelligence-based Device to Detect Certain Diabetes-related Eye Problems. <https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm604357.htm> (2018).
- Zawati Ma’n; Lang Michael. 2020 “What’s in the Box? Or, the Uncertain Accountability of Machine Learning Applications in Healthcare”. *The American Journal of Bioethics*.