CHAPTER 37

THE ETHICS OF AI IN BIOMEDICAL RESEARCH, PATIENT CARE, AND PUBLIC HEALTH

ALESSANDRO BLASIMME
AND EFFY VAYENA

Introduction

In March 2019 the World Health Organization announced amid a number of key reforms, the establishment of a new department of Digital Health with the aim to harness "the power of digital health and innovation by supporting countries to assess, integrate, regulate and maximize the opportunities of digital technologies and artificial intelligence."[1] This commitment at the global level is in the same vein with several national plans announced over the last couple of years[2] as governments began to grabble with AI in health. Numerous examples of AI-enabled digital health applications are available today, some have received market authorization, and if the private investment in digital health is anything to go by, the pipeline of future digital health products is going to be full. Certainly, the so-called big data revolution has been instrumental to this development.

In this chapter we discuss ethical challenges linked to the use of AI in biomedical research, patient care, and public health. We then draw on a systemic oversight model

[1] See https://www.who.int/news-room/detail/06-03-2019-who-unveils-sweeping-reforms-in-drive-towards-triple-billion-targets (accessed April 4, 2019).

[2] Lynne E. Parker, "Creation of the National Artificial Intelligence Research and Development Strategic Plan," *AI Magazine* 39, no. 2 (2018); Corinne Cath et al., "Artificial Intelligence and the 'Good Society': The US, EU, and UK Approach," *Science and Engineering Ethics* 24, no. 2 (2018): 505–528; Sophie-Charlotte Fischer, "Artificial Intelligence: China's High-Tech Ambitions," *CSS Analyses in Security Policy* 220 (2018).

for the governance of AI innovation in the health sector[3] and discuss possible ways to address emerging ethical challenges in this rapidly evolving domain. Our aim is to lay the groundwork for an ethically responsible development of AI in the domains of health research, clinical practice, and public health.

AI in Biomedical Research

In the last decade, biomedical research has become a data-centric activity[4] enabled by novel material and experimental practices linked to data collection, distribution, and use.

In the burgeoning field of precision medicine,[5] for instance, "omic" data are now routinely being collected alongside clinical data, phenotypic data, and life-style and socioeconomic data to form bigger-than-ever research cohorts. Artificial intelligence is predicted to enable the simultaneous computation of such diverse arrays of data, thus contributing to the promise of precision medicine to bring about more targeted approaches to diagnosis and treatment of individual patients.[6] As far as translational medicine is concerned, artificial intelligence is being employed in drug discovery to screen libraries of potentially therapeutic molecules, to automate searches in the biomedical literature through natural language processing techniques, to predict experimental dosage, and so on.[7]

Machine learning is also deployed to generate predictive models that could help doctors in prognostic assessment and in personalizing therapy and rehabilitation for individual patients, for instance in the aftermath of a stroke.[8]

[3] Effy Vayena and Alessandro Blasimme, "Health Research with Big Data: Time for Systemic Oversight," *Journal of Law, Medicine & Ethics* 46, no. 1 (2018): 119–129; Alessandro Blasimme and Effy Vayena, "Towards Systemic Oversight in Digital Health: Implementation of the AFIRRM Principles," in *Cambridge Handbook of Health Research Regulation*, ed. Graeme Laurie (Cambridge University Press, forthcoming).

[4] Sabina Leonelli, *Data-Centric Biology: A Philosophical Study* (University of Chicago Press, 2016).

[5] Francis S. Collins and Harold Varmus, "A New Initiative on Precision Medicine," *New England Journal of Medicine* 372, no. 9 (February 26, 2015): 793–795; Alessandro Blasimme and Effy Vayena, "Becoming Partners, Retaining Autonomy: Ethical Considerations on the Development of Precision Medicine," *BMC Medical Ethics* 17 (2016): 67; Alessandro Blasimme and Effy Vayena, " 'Tailored-to-You': Public Engagement and the Political Legitimation of Precision Medicine," *Perspectives in Biology and Medicine* 59, no. 2 (2017): 172–188.

[6] Bertalan Mesko, "The Role of Artificial Intelligence in Precision Medicine," *Expert Review of Precision Medicine and Drug Development* 2, no. 5 (2017): 239–241; Jia Xu et al., "Translating Cancer Genomics into Precision Medicine with Artificial Intelligence: Applications, Challenges and Future Perspectives," *Human Genetics* 138, no. 2 (February 1, 2019): 109–124.

[7] Eric J. Topol, "High-Performance Medicine: The Convergence of Human and Artificial Intelligence," *Nature Medicine* 25, no. 1 (2019): 51.

[8] See https://precise4q.eu (accessed April 4, 2019).

Electronic health records (EHR) offer the opportunity to use real-world data to generate knowledge about the outcomes of a given medical procedure (be it a diagnosis, a prognosis, a therapy, or a rehabilitation plan).[9] AI can be employed to mine EHR to discover disease familiarity or people at risk for a given chronic disease and also to improve the organization of health systems by providing support in triage and patient management.[10] In a recent study, deep learning was employed to create predictive modeling with EHR to accurately gauge in-hospital mortality, readmission odds, length of stay, and final discharge diagnoses.[11] In another study, a machine learning algorithm identified cancer patients at high risk of thirty-day mortality before they start chemotherapy (both palliative and curative).[12] Such an algorithm can help decisions about chemotherapy initiation, enabling more rational allocation of resources.

Facial recognition technologies based on machine learning are also being developed to streamline patient identification, to detect genetic disorders that correspond to specific facial traits[13] or to diagnose mood disorders such as depression.[14] Recently, researchers validated a system that, based on human-computer interaction patterns using data from a smartphone app, is able to recognize what the authors of the study call digital biomarkers of cognitive function.[15] Lately, there is increasing interest in voice analysis algorithms for health-related purposes with research concentrating on mental health.[16]

The main concern raised by AI in the previously described context is the quality and representativeness of data used to train machine learning algorithms. In the existing medical data sets, adult males of Caucasian origin are strongly overrepresented.[17] This lack of diversity is likely to result in biased algorithms trained on biased data. Similarly, EHR data used to train algorithms may suffer from issues such as missing data and

[9] Institute of Medicine, *The Learning Healthcare System: Workshop Summary (IOM Roundtable on Evidence-Based Medicine)*, 2007, https://www.nap.edu/catalog/11903/the-learning-healthcare-system-workshop-summary-iom-roundtable-on-evidence.

[10] Pavel Hamet and Johanne Tremblay, "Artificial Intelligence in Medicine," *Metabolism* 69 (2017): S36–40.

[11] Alvin Rajkomar et al., "Scalable and Accurate Deep Learning with Electronic Health Records," *NPJ Digital Medicine* 1, no. 1 (2018): 18.

[12] Aymen A. Elfiky et al., "Development and Application of a Machine Learning Approach to Assess Short-Term Mortality Risk among Patients with Cancer Starting Chemotherapy," *JAMA Network Open* 1, no. 3 (2018): e180926–e180926.

[13] Yaron Gurovich et al., "Identifying Facial Phenotypes of Genetic Disorders Using Deep Learning," *Nature Medicine* 25, no. 1 (2019): 60.

[14] Yu Zhu et al., "Automated Depression Diagnosis Based on Deep Networks to Encode Facial Appearance and Dynamics," *IEEE Transactions on Affective Computing* 9, no. 4 (2018): 578–584; Albert Haque et al., "Measuring Depression Symptom Severity from Spoken Language and 3D Facial Expressions," ArXiv Preprint ArXiv:1811.08592 (2018).

[15] Paul Dagum, "Digital Biomarkers of Cognitive Function," *NPJ Digital Medicine* 1, no. 1 (2018): 10.

[16] Nicholas Cummins, Alice Baird, and Björn W. Schuller, "Speech Analysis for Health: Current State-of-the-Art and the Increasing Impact of Deep Learning," *Health Informatics and Translational Data Analytics* 151 (December 1, 2018): 41–54.

[17] Latrice G. Landry et al., "Lack of Diversity in Genomic Databases Is a Barrier to Translating Precision Medicine Research into Practice," *Health Affairs* 37, no. 5 (2018): 780–785.

misclassification.[18] For example, people of lower socioeconomic levels may be less represented in certain diagnostic categories, or may be overrepresented in categories of emergency care. Such patients may be more concentrated to an institution than to others making research results of potential medical relevance more meaningful to overrepresented populations than minorities or socially emarginated groups.

Another concern relates to the sufficiency of informed consent as an ethical safeguard in research involving algorithmic processing. The traditional concept of informed consent is already challenged in cases of data collected in more conventional research settings, as it is increasingly hard to predict who will be accessing the data in the future, for which purposes, and under which conditions.[19] The reuse of data and the linkage of disparate data sets makes even the notion of broad consent—a typical safeguard of autonomy when future uses of human data and samples are hard to anticipate—weak. In the case of AI, it is still not clear whether research participants shall be specifically informed about the intention to use AI algorithms and whether informed consent for automated processing of personal data should reflect a heightened level of protection and, for instance, offer the possibility to opt out.

The creation of large cohorts of deeply phenotyped participants raises doubts about the huge amounts of information that such initiatives put in the hands of governments or private organizations. The latter include healthcare organizations, big tech, and companies active in the field of smart technologies that stipulate agreements with national governments to collect and analyze data from millions of citizens. As a consequence, issues of data privacy and security loom large on the horizon of biomedical big data research.[20]

AI adds a layer of ethical complexity to this scenario in that it uses data to extract additional, fine-grained information about individuals. It is an ethical responsibility of researchers to securely protect this information from unauthorized access in order to avoid privacy-related harms to data subjects in the course of research projects. The unwanted leak of health-relevant information can lead to discriminative uses of such information in domains such as employment, education, and insurance. This problem applies both to information generated and stored by researchers and to information that researchers feed back to research participants as primary, secondary, or incidental findings. Return of research results enjoys widespread support as a way to show respect for the interests and the welfare of research participants.[21] In particular, precision medicine initiatives, such as the U.S. All of Us Research Program, endorse a model of empowerment

[18] Milena A. Gianfrancesco et al., "Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data," *JAMA Internal Medicine* 178, no. 11 (2018): 1544–1547.

[19] Effy Vayena and Alessandro Blasimme, "Biomedical Big Data: New Models of Control over Access, Use and Governance," *Journal of Bioethical Inquiry* 14, no. 4 (2017).

[20] Omer Tene and Jules Polonetsky, "Privacy in the Age of Big Data: A Time for Big Decisions," *Stanford Law Review Online* 64 (2011): 63.

[21] Susan M. Wolf, "Return of Individual Research Results and Incidental Findings: Facing the Challenges of Translational Science," *Annual Review of Genomics and Human Genetics* 14, no. 1 (2013): 557–577, https://doi.org/10.1146/annurev-genom-091212–153506.

that is premised on the release of medically relevant information to research participants. This model, while laudable, can have consequences, for instance, for those research data subjects who intend to buy a life insurance policy.[22]

The criteria that are being employed in the evaluation of research involving human data and human subjects (including clinical trials) have been developed in the postwar period and formalized in most countries since the late 1970s. Such criteria—for example, social or scientific value, scientific validity, fair selection of participants, acceptable risk-benefit ratio, informed consent, and consideration for participants' welfare and rights[23]—while being still valid at a formal level, do not adequately capture the specificities of research involving the use of AI to analyze vast amounts of personal data.[24] Consider the case of a recent study that utilizing deep neural networks analyzed the association of facial traits and self-declared sexual orientation in order to understand whether homosexuals have distinct facial characteristics.[25] Besides the technical validity of this study, its aim is highly dubitable from an ethical point of view because it lends support to stereotyped views about homosexuality—namely, the idea that male homosexuals are effeminate and that female homosexuals are manly. Moreover, while it is hard to imagine any socially beneficial use of such a study, it can be expected that stigmatization and discrimination would likely result from either intentional or unintentional misuses of its results. This study exemplifies how AI can power new forms of classification based on the association between biological, personal, behavioral, and social characteristics. The unprecedented classificatory power of AI can obviously produce both tangible and intangible harms.[26] Notably, this particular study was reviewed by an institutional review board, passed peer-review, and was eventually published. The heated controversy that followed its publication brought to light the difficulty in assessing societal-wide effects when reviewing research, as well as the lack of agreed-upon criteria on how to do such an assessment.

Another issue of ethical relevance in the context of health research has emerged from collaborations between corporations with advanced capabilities in AI and healthcare institutions in control of health data sets. While such collaborations can be mutually beneficial, several examples to date have raised more concern than enthusiasm. The case of Deep Mind accessing 1.6 million health records from the Royal Free London NHS in order to test a kidney safety app, ended with the Information Commissioner finding a number of shortcomings in the contractual agreements. The Italian government's

[22] Alessandro Blasimme, Effy Vayena, and Ine Van Hoyweghen, "Big Data, Precision Medicine and Private Insurance: A Delicate Balancing Act," *Big Data & Society* 6, no. 1 (2019): 2053951719830111.

[23] David Wendler and Ezekiel J. Emanuel, "What Makes Clinical Research Ethical?," *JAMA* 283, no. 20 (May 2000): 2701–2711.

[24] Marcello Ienca et al., "Considerations for Ethics Review of Big Data Health Research: A Scoping Review," *PLOS ONE* 13, no. 10 (2018): e0204937.

[25] Yilun Wang and Michal Kosinski, "Deep Neural Networks Are More Accurate than Humans at Detecting Sexual Orientation from Facial Images," *Journal of Personality and Social Psychology* 114, no. 2 (2018): 246.

[26] Vanessa K. Ing, "Spokeo, Inc. v. Robins: Determining What Makes an Intangible Harm Concrete," *Berkeley Technology Law Journal* 32 (2017): 503.

decision to grant an IBM research unit access to citizens' health records has been questioned by both data protection and fair competition officials.[27] Beyond the question of whether such data are used with adequate consent, or whether social benefit will be accrued from their use, the further question is how such benefit will be distributed. If for-profit entities have exclusive deals with national health data organization, how will this affect access and distribution of subsequent AI products? We are still in the early days of understanding the implications of such arrangements and of articulating fair agreements despite the fact that there is a litany of cases that seem to raise the questions.

AI in Patient Care

AI-driven diagnosis is certainly one of the most promising fields of application for AI in patient care. AI has largely demonstrated its ability to interpret various types of medical images, such as X-ray scans, magnetic resonance, and also photographic images of body parts (such as skin or eye fundus) and digitalized pathology slides. Image interpretation and visual pattern recognition are therefore among the major drivers in this space. An obviously limited list of examples includes the use of deep learning techniques to train algorithms to detect wrist fractures in X-ray scans;[28] to help cardiologists interpret magnetic resonance images;[29] and a machine learning software that detects diabetic retinopathy by automatically interpreting images from the back of the patient's eye.[30] These three applications received clearance for marketing from the U.S. Food and Drug Administration (FDA). Many more have appeared in the literature, including algorithms that can compute cardiovascular risk factors based on retinal images.[31] In all those studies, the performance of the algorithms was tested against the benchmark of certified specialists' assessments, revealing equal or superior outcomes for AI system as compared to human physicians. This criterion is widely used in research settings, but it is not yet established as a sufficient one for AI applications in clinical care. The issue of evidence standards has obvious implications in terms of safety and efficacy. As a consequence, a major issue with clear ethical implications is the reliability of the evidence in favor of AI clinical applications.

[27] See https://www.repubblica.it/economia/2017/12/05/news/dati_sanitari_alle_multinazionali_senza_consenso_passa_la_norma-183005262/ (accessed April 4, 2019). At the time of writing, the initiative is on hold.

[28] Food and Drug Administration, "FDA Permits Marketing of Artificial Intelligence Algorithm for Aiding Providers in Detecting Wrist Fractures," available at https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm608833.htm (accessed April 4, 2019).

[29] Bernard Marr, "First FDA Approval for Clinical Cloud-Based Deep Learning in Healthcare," *Forbes* (January 20, 2017), available at https://www.forbes.com/sites/bernardmarr/2017/01/20/first-fda-approval-for-clinical-cloud-based-deep-learning-in-healthcare/#6af6ceef161c.

[30] See https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357 (accessed April 4, 2019).

[31] Ryan Poplin et al., "Prediction of Cardiovascular Risk Factors from Retinal Fundus Photographs via Deep Learning," *Nature Biomedical Engineering* 2, no. 3 (2018): 158.

Some AI-driven diagnostic applications can also be operated directly by the patient on portable devices outside the clinical setting. One can imagine, for example, that smartphone apps could incorporate already existing AI-powered algorithms to inspect nevi and detect the presence of skin cancer.[32] Similarly, the first smart pill was approved by the FDA in 2017 and included an ingestible sensor that sends a signal to the patient's device once the pill is taken in order to help him or her adhere to a prescription.[33] Commentators have highlighted that, from a patient perspective, ethical issues for this type of devices include concerns for autonomy, privacy, and dependability in case of technical failures.[34]

Ethical issues in the use of AI for patient care depend on specific uses and applications. It is intuitively plausible to think that ethical stakes correlate with the severity of the condition at hand or with the degree of reliance on AI for serious medical tasks such as diagnosis or treatment. It would be wrong, however, to assume that automation in health system services is less likely to have ethically relevant implications. Consider the case of triage. AI-driven decisions such as which patient is treated first or which one is offered chemotherapy[35] should certainly follow cost-effectiveness considerations. But exclusive reliance on algorithms may rule out that necessary degree of flexibility that allows healthcare operators to calibrate objective criteria with the reality of each individual case.[36] For instance, a system that factors the risk of longer stays into decisions about hospital admission may discriminate against the most vulnerable patients, that is, arguably, those who are more in need of care. While it is premature to say that these unfair outcomes will be the case, such ethically relevant aspects of automating clinical workflow deserve careful scrutiny.

As to the use of AI for diagnostic purposes, the already mentioned problem of a biased training data set that lead to suboptimal performance for underrepresented social groups creates an ethical bottleneck. In the current ethical debate about AI in medicine, the issue of whether and why the use of AI should be disclosed to patients during informed consent procedures is still in its infancy. However, a bigger discussion is ongoing as to whether black-box algorithms—that is, algorithms whose self-learned rules are too complex to reconstruct and explain—should be used in medicine.[37] Some have called for a duty to transparency in order to dispel the opacity of black-box algorithms.[38] Others, however, have highlighted that more limited requirements are

[32] Andre Esteva et al., "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," *Nature* 542, no. 7639 (2017): 115.

[33] https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm584933.htm.

[34] Craig M. Klugman et al., "The Ethics of Smart Pills and Self-Acting Devices: Autonomy, Truth-Telling, and Trust at the Dawn of Digital Medicine," *American Journal of Bioethics* 18, no. 9 (2018): 38–47.

[35] Rajkomar, "Scalable and Accurate Deep Learning"; Elfiky, "Development and Application of a Machine Learning Approach."

[36] Effy Vayena, Alessandro Blasimme, and I. Glenn Cohen, "Machine Learning in Medicine: Addressing Ethical Challenges," *PLOS Medicine* 15, no. 11 (2018): e1002689.

[37] W. Nicholson II Price, "Black-Box Medicine," *Harvard Journal of Law & Technology* 28 (2015): 419.

[38] Sandra Wachter, Brent Mittelstadt, and Luciano Floridi, "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation," *International Data Privacy Law* 7, no. 2 (2017): 76–99.

sufficient to adequately protect the morally relevant interests of patients when machine learning algorithms are employed to provide care.[39]

An important issue concerns the shift of medical authority from human physicians to algorithms—the problem of the so-called "collective medical mind."[40] The risk here is that AI systems introduced as decision support tools become central nodes of medical decision-making. In this scenario, it is uncertain how the established principles of medical ethics (beneficence, nonmaleficence, respect for patients) can still be expected to play the central role in the patient-doctor relationship that they have—or at least can be expected to have—now. The mediation of AI-powered tools can fundamentally alter the doctor-patient relationship. AI, especially as it enables remote care or communication via robotic assistants, may create interpersonal distance between patients and their physicians. An incentive to use such tools could be the need to streamline patient care, but the downside of this phenomenon is that the patient becomes more isolated, with potentially negative repercussions on health outcomes. The same considerations can be made about AI-based home-assistance platforms. In principle, these systems can be extremely useful to, for instance, provie better care to elderly people with limited mobility. However, they can also increase their social isolation.

The easiness with which an AI system can keep track of a person's health and perform accurate diagnostic has been discussed as a potential source of overdiagnosis and nonactionable diagnoses. For instance, employing deep learning to infer cardiovascular risk factors from retinal fundus pictures[41] is warranted by the fact that it could lead to life-style adaptations that may actually improve a patients' condition. But the use of images of retinal structures as biomarkers of dementia[42] are more problematic in the absence of concluding evidence regarding the efficacy of interventions to delay or slow down dementia.[43]

Finally, the use of algorithms for mood detection promises to revolutionize mental health.[44] However, privacy issues acquire particular ethical relevance in this context. Tools like DeepMood, which allow the detection of mood based on mobile phone typing

[39] Andrew D. Selbst and Julia Powles, "Meaningful Information and the Right to Explanation," *International Data Privacy Law* 7, no. 4 (2017): 233–242; Agata Ferretti, Manuel Schneider, and Alessandro Blasimme, "Machine Learning in Medicine: Opening the New Data Protection Black Box," *European Data Protection Law Review* 4, no. 3 (2018): 320–332.

[40] Danton S. Char, Nigam H. Shah, and David Magnus, "Implementing Machine Learning in Health Care—Addressing Ethical Challenges," *New England Journal of Medicine* 378, no. 11 (March 15, 2018): 981–983, https://doi.org/10.1056/NEJMp1714229.

[41] Poplin et al., "Prediction of Cardiovascular Risk Factors from Retinal Fundus Photographs via Deep Learning."

[42] Unal Mutlu et al., "Association of Retinal Neurodegeneration on Optical Coherence Tomography with Dementia: A Population-Based Study," *JAMA Neurology* 75, no. 10 (2018): 1256–1263.

[43] Engineering National Academies of Sciences and Medicine, *Preventing Cognitive Decline and Dementia: A Way Forward* (National Academies Press, 2017).

[44] David C. Mohr, Heleen Riper, and Stephen M. Schueller, "A Solution-Focused Research Approach to Achieve an Implementable Revolution in Digital Mental Health," *JAMA Psychiatry* 75, no. 2 (2018): 113–114.

dynamics, are certainly promising.[45] Yet pervasive tracking of one's emotional state is at least intrusive and may affect the legitimate interest of any individual to keep control over information about his or her mood. Mood and mental health can now be digitally tracked through sensors that capture anything from breathing patterns, to galvanic skin response, from the tone of our voice, to sleep patterns, facial expressions, our whereabouts, and social media traces.[46] The possibility of being constantly monitorable as to our emotional states and mental health is certainly problematic from an ethical viewpoint as it sets the conditions for a form of granular psychological surveillance that is at odds with the values of pluralistic liberal societies. Even if these tools are employed in the context of a therapeutic relationship, their excessive use undermines a patient's capacity to remain autonomous and to maintain a sense of self-determination vis-à-vis his or her doctor.

AI in Public Health

Uses of algorithms in public health research and practice can have significant impact on population health.[47] Health is affected by several social parameters (e.g., income, education, dietary habits, environmental factors, community context) that are not confined in the healthcare systems. Understanding specific effects and interactions between health and various social conditions can lead to the development of more effective and efficient public health programs. Examples from AI-enabled multilevel modeling using sociomarkers have already demonstrated such potential.[48] A particular area of AI application in public health is disease surveillance. Surveillance systems monitor disease incidence, outbreaks, and health behaviors. Typically these systems are state-funded and state-operated. Their purpose is to monitor the health of populations and subsequently to support decision-making for allocation of resources and types of interventions necessary to improve health. As a data-driven activity, surveillance can benefit substantially from algorithmic uses. Algorithms can sort through variables that are relevant for specific health outcomes, they can recognize patterns and signals at a much faster pace, and they can be used to forecast epidemics and to model their trajectories. Such algorithms have been used to mine not only standard health data collected for surveillance by state institutions but also real-world data through social media. This seemingly unconventional

[45] Bokai Cao et al., "DeepMood: Modeling Mobile Phone Typing Dynamics for Mood Detection" (Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2017), 747–755.

[46] Paddy M. Barrett et al., "Digitising the Mind," *The Lancet* 389, no. 10082 (2017): 1877.

[47] Arash Shaban-Nejad, Martin Michalowski, and David L. Buckeridge, "Health Intelligence: How Artificial Intelligence Transforms Population and Personalized Health," *NPJ Digital Medicine* 1, no. 1 (October 2, 2018): 53.

[48] Eun Kyong Shin et al., "Sociomarkers and Biomarkers: Predictive Modeling in Identifying Pediatric Asthma Patients at Risk of Hospital Revisits," *NPJ Digital Medicine* 1, no. 1 (October 2, 2018): 50.

approach suffered an early blow when Google Flu Trend algorithms failed to show their promised predictive power.[49] Since then, however, AI-enabled analysis of social media data has produced several successful examples, including better prediction of epidemics[50] and detection of food poisoning cases.[51] The broader field of digital epidemiology is a rapidly evolving field focused on epidemiological models based on content posted online by social network users.[52] Forms of AI like natural language processing obviously play a crucial role for the further development of this field. Ethical challenges in this domain revolve mainly around consent. Many commentators have stressed that the terms of use for social media fall short of complying with the rigorous requirements for informed consent in the domain of health-related research.[53]

AI combined with mobile health applications also offers a new avenue for delivering public health intervention to populations. Of relevance here are expectations for health promotion to reach populations that are marginalized by targeting them with tailored interventions.[54] An area of contest in public health ethics has been the ethical legitimacy of nudging personal behavior for health-related purposes. AI will make this issue even more significant. Continuous surveillance, tailored nudging, and paternalistic interventions can generate an Orwellian form of individual control and constrained personal freedoms.[55] States and corporations with access to tools that can monitor and alter health-related behaviors can exercise significant power over large numbers of people to further their specific interests. While in a democratic and accountable state such policies can be vetted, be transparent, and revised as necessary, that is not necessarily the case everywhere nor is it the case when such behavioral manipulation occurs in arenas that are controlled entirely by institutions without public accountability.

There is significant enthusiasm for the use of AI in global health with funding agencies and international organizations investing already in public health activities in low- and middle-income countries. The World Health Organization has recently committed to promote AI to achieve universal health coverage, and many governments have been interested in taking stock of digital technologies to improve healthcare systems as they stated in a 2018 resolution on digital health that was adopted by the 71st World Health

[49] Declan Butler, "When Google Got Flu Wrong," *Nature News* 494, no. 7436 (2013): 155.

[50] Mohammed Ali Al-Garadi et al., "Using Online Social Networks to Track a Pandemic: A Systematic Review," *Journal of Biomedical Informatics* 62 (August 1, 2016): 1–11.

[51] Jenine K. Harris et al., "Using Twitter to Identify and Respond to Food Poisoning: The Food Safety STL Project," *Journal of Public Health Management and Practice: JPHMP* 23, no. 6 (December 2017): 577–580.

[52] Marcel Salathé et al., "Digital Epidemiology," *PLOS Computational Biology* 8, no. 7 (2012): e1002616, https://doi.org/10.1371/journal.pcbi.1002616; Antoine Flahault et al., "Precision Global Health in the Digital Age," *Swiss Medical Weekly* 147 (April 19, 2017): w14423, https://doi.org/smw.2017.14423.

[53] Jeffrey P. Kahn, Effy Vayena, and Anna C. Mastroianni, "Opinion: Learning as We Go: Lessons from the Publication of Facebook's Social-Computing Research," *Proceedings of the National Academy of Sciences* 111, no. 38 (September 23, 2014): 13677–13679.

[54] Brian Wahl et al., "Artificial Intelligence (AI) and Global Health: How Can AI Contribute to Health in Resource-Poor Settings?," *BMJ Global Health* 3, no. 4 (2018): e000798.

[55] Sarah Nettleton and Robin Bunton, "Sociological Critiques of Health Promotion," in *The Sociology of Health Promotion*, ed. Sarah Nettleton, Robin Bunton, and Roger Burrows (Routledge, 1995) 41–58.

Assembly.[56] This commitment increases the likelihood of AI entering rapidly the domain of health, adding urgency to the need of identifying and addressing the ethical tensions that AI generates.[57] The most pertinent are those related to the potential exacerbation of health disparities through biases that are perpetuated or reinforced by AI-enabled interventions. We discussed the problem of misrepresentation of certain populations in health-related data sets above. Several methods are currently under development to compensate for bias, but at the time the problem remains and requires attention.[58] Underserved populations present certain negative health outcomes due to well-known social deficits. Algorithms that produce decisions based on health outcomes alone, without factoring in their social causes, can result in significant harm and increased health inequalities. For example, if poor or less-educated people have performed worse after certain health interventions (due to poor access to care, working schedules, etc.), an algorithm can determine that people with these characteristics will always perform worse and recommend that they are not offered the intervention in the first place. This will exacerbate disparity in access to care and attainment of good health outcomes. More importantly, it will make such disparity less visible because the decision will bear the authoritative objectivity often attributed to numbers and that is typically expected from automated decision-making tools.

Addressing the Ethical Challenges

The novelty represented by AI, and machine learning in particular, might be on the verge of pushing medical research, patient care, and public health into as yet uncharted ethical territories. The impact of AI in these three domains is particularly challenging to anticipate, and it is hard to predict whether expected benefits will offset emerging risks. In this scenario neither a precautionary approach nor a wait-and-see attitude is compatible with the widely accepted need to ensure ethically sustainable, socially robust, and responsible innovation in this domain. A precautionary approach implies erring on the side of containing possible risks when evidence about how a given phenomenon will evolve is scarce and the stakes are high in terms of potential harms.[59] As far as the use of AI in medicine is concerned, a precautionary approach would likely result in disproportionate constraints that might undermine the development of promising technologies. On the other hand, a more permissive "wait-and-see" approach, while being more

[56] See http://apps.who.int/gb/ebwha/pdf_files/WHA71/A71_R7-en.pdf (accessed April 4, 2019).

[57] Effy Vayena and Lawrence Madoff, "Navigating the Ethics of Big Data in Public Health," in *The Oxford Handbook of Public Health Ethics*, ed. A. C. Mastroianni, J. P. Kahn, and N. E. Kass (Oxford University Press, 2019): 354–367.

[58] Robert Challen et al., "Artificial Intelligence, Bias and Clinical Safety," *BMJ Quality & Safety* 28, no. 3 (March 1, 2019): 231.

[59] Elizabeth Charlotte Fisher, Judith S, Jones, and René von Schomberg, *Implementing the Precautionary Principle: Perspectives and Prospects* (Edward Elgar, 2006).

favorable to the development and rapid uptake of AI-driven solutions, would necessarily have to rely on existing ethical safeguards. But such safeguards, as we have seen, fall short of covering the rapidly expanding catalog of ethical issues that AI poses in the domain of biomedicine. The collection, use, and reuse of increasingly large amounts of personal data, as we have seen, calls into question the adequacy of key components of the existing regulatory toolkit, such as evidence standards, ethics review, and informed consent.[60]

What is needed to ensure responsible AI innovation is a governance approach that coevolves with the field itself, incorporating new governance actors and experimenting with new oversight mechanisms to cope with ethical challenges as they arise from practice. Such a governance model should primarily drive attention to the ethically controversial aspects of AI-driven innovation in biomedicine in order to ensure that emerging risks do not pass unnoticed. A second aim of an ideal governance frame would by that of channeling innovation toward socially beneficial outcomes. Finally, good governance should promote public trust in and accountability of the innovation process. These objectives demand a specific *systemic* approach to governing a complex phenomenon whose outcomes are still largely unpredictable.

In the last two decades, scholarship on governance of controversial areas of science and innovation has given substantial consideration to so-called adaptive governance as a model to cope with uncertainty in public policy.[61] Adaptive governance centers around constant monitoring of both the phenomenon at stake and the policy measures deployed to control it. In practical terms, this model invites oversight and regulation to take stock of evidence as it becomes available and promoting social learning among a variety of different governance stakeholders.[62] Drawing on the broad frame of adaptive governance, we have proposed a governance model for data-driven innovation in biomedicine called "systemic oversight."[63] Systemic oversight is specifically designed to address what gives rise to ethical issues in the use of big data and AI in biomedicine, that is, as we have seen, novel data sources, novel data uses, increased capacity to draw connections between disparate data points, and uncertainty about downstream effects of such increased classificatory powers. The systemic oversight approach is based on six principles offering guidance as to the desirable features of oversight structures and processes in the domain of data-intense biomedicine: adaptivity, flexibility, inclusiveness, reflexivity, responsiveness, and monitoring (the first letters of the principles form the acronym AFIRRM).

[60] Effy Vayena et al., "Digital Health: Meeting the Ethical and Policy Challenges," *Swiss Medical Weekly* 148 (2018): w14571.

[61] Carl Folke et al., "Adaptive Governance of Social-Ecological Systems," *Annual Review of Environment and Resources* 30, no. 1 (2005): 441–473.

[62] Brian Chaffin, Hannah Gosnell, and Barbara A. Cosens, "A Decade of Adaptive Governance Scholarship: Synthesis and Future Directions," *Ecology and Society* 19, no. 3 (2014): 56.

[63] Vayena and Blasimme, "Health Research with Big Data"; Blasimme and Vayena, "Towards Systemic Oversight in Digital Health."

Adaptivity refers to the capacity of governance bodies and mechanisms to guarantee appropriate forms of oversight for new data sources and new data analytics that get incorporated in research, patient care, or public health activities. *Flexibility* is the capacity to treat different data types based both on their source *and* on their actual use, and it is premised on the consideration that data acquire specific ethical meaning in different contexts of use. *Inclusiveness* stresses the need to include all affected parties in deliberations and decision-making practices about the use of data and algorithms in specific ambits. This component refers in particular to communities and actors that are historically marginalized, vulnerable, or otherwise excluded from the circuits of power, such as minorities and patient constituencies. *Reflexivity* prescribes careful scrutiny and assessment of emerging risks in the short run as well as in the long run in terms of the downstream effects of big data and AI on interests, rights, and values, for example, in terms of fair access to healthcare services, discrimination, stigmatization, medicalization, overdiagnosis, and so on.

We saw earlier that AI is a powerful generator of health-relevant information and thus exposes research participants, patients, and data subjects in general to unwanted leaks of personal data and information. *Responsiveness* refers therefore to the need for adequate mechanisms to mitigate the effects of unauthorized access to personal health-related information. Finally, *monitoring* expresses the need to predispose regular scrutiny of data-related activities and their effects on health-related practices in order to anticipate the emergence on new vulnerabilities and undesirable outcomes.

The implementation of the AFIRRM frame will require consideration for the well-characterized obstacles to adaptive governance in other policy domains. Particular attention needs to be paid to the composition of oversight bodies. The demands of inclusiveness, for example, can only be appropriately fulfilled if diverse stakeholders share at least a common understanding of the intended advantages and potential risks of using AI in biomedicine. It is possible, for instance, that automating hospital services through AI-driven triage systems caters to the financial interests of hospitals (by rationalizing resource allocation), while failing to meet the expectations of severely ill patients in terms of access to care. As a consequence, the inclusion of patients' perspectives into decisions about the adoption of such systems both requires and fosters the existence of shared visions about fairness in access to health services. Along similar lines, oversight mechanisms on the use and effects of AI in clinical practice must escape purely technical considerations about the safety and efficacy of automated clinical decisions. Downstream effects on the patient-doctor relationship or on the right of patients to decide whether they are open or not to highly automated decisions need to be considered. To this aim, new review processes for clinical validation as well as novel communication and consent requirements will have to be established. The same applies in the research domain when researchers interested in using large amounts of phenotypic data need to negotiate the terms of use with data subjects, some of which may have value-laden views about the ethical legitimacy of certain types of research.

With the advent of AI, the agenda of academic disciplines like clinical research ethics, medical ethics, and public health ethics is rapidly adapting to incorporate new issues

and new controversies. Given its theoretical and thematic specificity, one may characterize this area as a separate subarea of study in applied ethics and call it "digital bioethics." Whether and how this scholarship will inform the emergence of new oversight tools remains to be seen. In the meantime, practical proposals, criteria, and best practices about the governance of AI-driven innovation in biomedicine are just starting to emerge. The U.K. National Institute for Clinical Excellence (NICE), the body advising the U.K. National Health Service (NHS) on matters related to health technology assessment, has just released guidance on clinical validation of digital health technologies (DHTs).[64] This guidance establishes evidence standards (grouped in four evidence tiers) according to the function that a given DHT is intended to perform. Such standards are going to be applied to DHTs harboring an AI component as well as to stand-alone AI software. In February 2019 the NHS released an updated version of its Code of Conduct for Data-driven Health and Care Technologies.[65] The principles proposed by this code include understanding users' needs, clearly defining the expected outcomes and benefits, lawful data processing, transparency, and evidence of safety and effectiveness (based on the NICE criteria). The NHS frame has been criticized for its lack of attention to the risk that AI in the healthcare space may widen social inequalities.[66] Still in the United Kingdom, the Wellcome Trust—a major funder of biomedical research in the country—has recently proposed a model called "dynamic oversight" for emerging science and technologies that partially resembles our own systemic oversight approach and the AFIRRM principles.[67]

In the United States, the American Medical Association released its policy on AI in 2018.[68] This document highlights the transformative potential of AI in the clinical domain and recommends that clinically validated AI should be aligned to best clinical practices, be transparent, be reproducible, be immune to data biases, and protect patients' privacy as well as the integrity of their personal information. In the United States, the FDA is the gatekeeper of AI-driven health innovation because it has statutory oversight power on medical devices and software as a medical device. In Europe, instead, the new 2017 Regulation on Medical Devices[69] relies on third parties (called notified bodies) issuing conformity certificates for medical devices. The FDA is piloting a precertification program to identify "manufacturers who have demonstrated a robust culture of quality and organizational excellence, and who are committed to monitoring

[64] See https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf (accessed April 4, 2019).

[65] See https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology (accessed April 4, 2019).

[66] Melanie Smallman, "Policies Designed for Drugs Won't Work for AI," *Nature* 567, no. 7746 (2019): 7.

[67] See https://wellcome.ac.uk/sites/default/files/blueprint-for-dynamic-oversight.pdf (accessed April 4, 2019).

[68] See https://www.ama-assn.org/system/files/2019-01/augmented-intelligence-policy-report.pdf (accessed April 4, 2019).

[69] See https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0745 (accessed April 4, 2019).

real-world performance of their products once they reach the U.S. market."[70] In April 2019, the FDA also released a proposed regulatory framework for AI and machine learning medical software addressing the specific issue of algorithms that keep on training themselves based on new data acquired during clinical use.[71]

Conclusions

The current proliferation of guidelines and codes of conduct demonstrates the need for ethical and technical points of reference for this rapidly evolving field. Considering the broad scope of potential applications for research, clinical use, and public health, it is likely that some specific uses of AI will not be covered by existing oversight mechanisms. But reliance on existing regulatory tools alone will likely fail to ensure adequate levels of public trust and accountability. For this reason, we have advanced the systemic oversight/AFIRRM approach as a governance blueprint. Looking at the nature of ethical issues illustrated in this chapter in light of the AFIRRM principles, it seems at least advisable that certain measures be implemented in the short term. In the research domain, ethical review committees will have to incorporate reflexive assessment of the scientific and social merits of AI-driven research and, to this aim will likely have to open their ranks to new professional figures such as social scientists. Research funders, on the other hand, can require monitoring and responsiveness mechanisms to be part of research plans and could set up multidisciplinary committees to periodically assess data from such activities in order to adjust their funding policies in the future. When AI-driven research amounts to large-scale projects claiming data from entire communities or populations, adequate forms of inclusion must be experimented with in order to ensure social learning across different epistemic communities—including lay publics and nonacademic actors.

In the domain of patient care, clinical validation is a crucial issue. Ad hoc evidence standards are a necessary condition for responsible clinical innovation, but they are not sufficient to cover the breath of potential ethical issues we saw in this area. Hospitals could equip themselves with "clinical AI oversight bodies" charged with the task of advising clinical administrators regarding the adoption of a given AI technology and monitoring its effects on patient journeys and patients' engagement throughout the continuum of care. Moreover, consent requirements will need to be adapted to the presence of highly automated data-processing, for instance, in the domain of diagnostics.

In the public health sphere, the new level of granularity enabled by AI in disease surveillance and health promotion will have to be negotiated at the level of targeted communities or it will result in a sense of disempowerment and, as a consequence, in a lack of public trust. The acceptable limits of data collection and algorithmic analysis, in other

[70] See https://www.fda.gov/MedicalDevices/DigitalHealth/UCM567265 (accessed April 4, 2019).
[71] See https://www.regulations.gov/document?D=FDA-2019-N-1185-0001 (accessed April 4, 2019).

words, will have to result from community-wide inclusive deliberation, especially as to who is collecting and processing data and for which exact purposes.

These are just a few examples of initiatives that, if adopted, will contribute to the development AI into a socially robust technology. It is clear that we are at the very beginning of a foreseen transformation. Should this transformation occur, its real effects may be different from those that we are able to anticipate now. This level of uncertainty, however, shall not deter societal stakeholders—including scientific and clinical institutions—from experimenting with governance arrangements aimed at reaping the benefits of AI for human knowledge and health, while at the same time paying sufficient attention to emerging ethical challenges.

Bibliography

Char, Danton S., Nigam H. Shah, and David Magnus. "Implementing Machine Learning in Health Care—Addressing Ethical Challenges." *New England Journal of Medicine* 378, no. 11 (March 15, 2018): 981–983.

He, Jianxing, Sally L. Baxter, Jie Xu, Jiming Xu, Xingtao Zhou, and Kang Zhang. "The Practical Implementation of Artificial Intelligence Technologies in Medicine." *Nature Medicine* 25, no. 1 (January 2019): 30.

Price, W. Nicholson II. "Black-Box Medicine." *Harvard Journal of Law & Technology* 28 (2015 2014): 419.

Smallman, Melanie. "Policies Designed for Drugs Won't Work for AI." *Nature* 567, no. 7746 (2019): 7.

Topol, Eric J. "High-Performance Medicine: The Convergence of Human and Artificial Intelligence." *Nature Medicine* 25, no. 1 (2019): 44.

Vayena, Effy and Alessandro Blasimme. "Health Research with Big Data: Time for Systemic Oversight." *Journal of Law, Medicine & Ethics* 46, no. 1 (2018): 119–129.

Vayena, Effy, Alessandro Blasimme, and I. Glenn Cohen. "Machine Learning in Medicine: Addressing Ethical Challenges." *PLOS Medicine* 15, no. 11 (2018): e1002689.

Yu, Kun-Hsing, Andrew L. Beam, and Isaac S. Kohane. "Artificial Intelligence in Healthcare." *Nature Biomedical Engineering* 2, no. 10 (October 1, 2018): 719–731.