Achieving Trustworthy Biomedical Data Solutions

Peter Washington
Department of Bioengineering, Stanford University
Stanford, CA, 94305, USA
Email: peterwashington@stanford.edu

Serena Yeung
Department of Biomedical Data Science, Stanford University
Stanford, CA, 94305, USA
Email: syyeung@stanford.edu

Bethany Percha
Department of Medicine, Icahn School of Medicine at Mount Sinai
New York, NY 10029, USA
Email: bethany.percha@mssm.edu

Nicholas Tatonetti
Department of Biomedical Informatics, Columbia University
New York, NY 10032, USA
Email: nick.tatonetti@columbia.edu

Jan Liphardt
Department of Bioengineering, Stanford University
Stanford, CA, 94305, USA
Email: jan.liphardt@stanford.edu

Dennis P. Wall
Departments of Pediatrics (Systems Medicine), Psychiatry and Behavioral Sciences, and Biomedical Data Science, Stanford University
Stanford, CA, 94305, USA
Email: dpwall@stanford.edu

Privacy and trust of biomedical solutions that capture and share data is an issue rising to the center of public attention and discourse. While large-scale academic, medical, and industrial research initiatives must collect increasing amounts of personal biomedical data from patient stakeholders, central to ensuring precision health becomes a reality, methods for providing sufficient privacy in biomedical databases and conveying a sense of trust to the user is equally crucial for the field of biocomputing to advance with the grace of those stakeholders. If the intended audience does not trust new precision health innovations, funding and support for these efforts will inevitably be limited. It is therefore crucial for the field to address these issues in a timely manner. Here we describe current research directions towards achieving trustworthy biomedical informatics solutions.

Keywords: privacy; trust; data security; biomedical systems; bioinformatics; artificial intelligence (AI); trustworthy AI

1. Introduction

The importance of trust in biomedical and healthcare technologies, especially consumer-facing artificial-intelligence (AI) software, cannot be overstated. Issues of privacy and trust with regard to large-scale data capture and analysis, particularly passive data capture by mobile devices and social media, have recently come to the forefront of public and academic discourse across multiple domains [1-4]. Such issues are especially important for healthcare, where solutions must prioritize patient privacy. At a minimum, biomedical tools in the United States must satisfy the Health Insurance Portability and Accountability Act of 1996 (HIPAA), which mandates a set of regulations regarding the privacy of patient health data [5]. While satisfying legal constraints is necessary, the true metric of achieving satisfactory patient trust will come from the patients themselves, who may request more stringent solutions.

In recent years, the biomedical research community has produced a wide array of research findings relating to trustworthy biomedical data, spanning multiple fields and subdomains. Work in these areas has included genomic data storage [6], privacy and sharing of protected health information (PHI) [7-9], cryptography solutions to sharing genetic data that allow public querying while protecting patient privacy [10], ethical considerations of new technologies and paradigms [11], and privacy-preserving machine learning methods [12-13]. However, the increasing prevalence of large-scale biomedical data collection capabilities and efforts (such as the continued decrease in sequencing costs), coupled with the explosion of applied machine learning systems and products, continually creates demand for innovations in trustworthy methods which can handle growing technological capabilities.

Here, we focus on four active themes in biomedical data science where the importance of trust in data has taken center stage: (1) preserving privacy and explaining the decisions of artificial intelligence algorithms, (2) sharing genomic and health records, (3) deploying digital health solutons, and (4) crowdsourcing healthcare. For each research theme, we describe several core methodological approaches (Figure 1) for building trustworthy biomedical data solutions which apply across the data science pipeline: (1) data transformation (e.g., dimension reduction and image modification), (2) access control (e.g., federated learning and cryptography), (3) data aggregation (e.g., aggregate queries and differential privacy), and (4) transparency (e.g., explainable AI). We discuss how these trust-enabling methodologies can and should be invoked and describe prior efforts. We conclude with a brief discussion of the bioethics literature.

2. Preserving Privacy and Explaining Decisions of Artificial Intelligence

AI in healthcare is increasingly rising in importance for solving challenges in the medicial workflow including clinicial decision support, preventing errors, and scaling redundant tasks. Privacy preservation and explainability are crucial when machine learning algorithms are deployed in these settings. We describe three common machine learning paradigms for attaining and preserving patient privacy when biomedical data are used to train algorithms: (1) transformation of the data, (2) federated learning, and (3) differential privacy. We also discuss efforts to attain explainable AI.

If the data can be transformed in such a way that the downstream model still yields high predictive performance, simply altering the data to obfuscate the identity of the subject may be the most desired option. For example, when using computer vision for use in activity recognition in hospital bedside settings [14-15], Yeung et al. leverage thermal [16] and depth [17] sensors to create

privacy-preserved video streams. Washington et al. simply place a face box over the patients' faces and pitch shift the audio when generating behavioral phenotypes of children with autism using machine learning and crowdsourcing [18], only minimally degrading performance compared to when using unaltered videos. Machine learning models should be trained and tested on the maximally private alteration of the data while maintaining acceptable performance.

Federated learning as a privacy enhancing technique has garnered widespread attention for achieving privacy in distributed mobile devices that may collect multimedia data streams. In federated machine learning, several distributed machines train models based on local data and share only model weights, which do not contain any protected information, on either the other distributed devices or a centralized server [19]. Federated learning has been applied to analyze data from electronic health records [20-22], recognize activity patterns based on data from wearable devices [23], and improve the interpretation of medical images [24].

A third commonly used privacy preserving technique is differential privacy. Differential privacy involves injecting random noise into the training dataset such that the identifiability of each individual record is destroyed while the aggregate properties of the dataset are preserved [25]. Examples of applying differential privacy to protect patient privacy in the biomedical domain include injecting noise into data from wearable sensors [26], genome wide association studies [27], and healthcare social networks [28]. This session includes a paper by Shi et al. that explores the tradeoffs between the performance of commonly used machine learning models and the level of privacy attained using differential privacy.

Another crucial property of trustworthy machine learning is explainability, including but not limited to interpretability. Some machine learning algorithms are inherently explainable. In classification with logistic regression, for example, the exact prediction can be calculated from the input values by plugging them into an equation. Making the coefficients associated with each variable transparent to the patient in a user-friendly manner would increase trust. However, with a large dataset of high complexity, explainable algorithms may not be sufficient, requiring more powerful yet less interpretable algorithms like neural networks. While components of certain neural networks can be interpreted, such as by visualizing the weights and activations of feature maps in the intermediate layers of a convolutional neural network, making neural networks explainable is an emerging active area of research [29]. Creating explainable AI has enabled increased reasoning about the decision making process behind stroke prediction algorithms [30], further understanding of changes in the skin microbiome [31], and elucidation of the reasoning of algorithms trained on electronic health record [32]. In some cases, explainable AI can lead to scientific discovery, for example by elucidating complex disease pathways in autism [33]. As explainable AI is becoming a popular research direction across computing research fields, we expect more translatable innovations in the coming years that safely embed AI in a variety of sectors of the healthcare ecosystem.

3. Sharing Genomic and Health Records

The genome is a core foundation of precision healthcare, and shared human DNA records are essential to advancements in human health. Millions of human genomes have been sequenced, either through direct-to-consumer DNA platforms (e.g., 23andme and Ancestry) or through a healthcare provider, with the number likely to exponentially increase as genomic sequencing becomes progressively more affordable and more speedy, improving at a rate faster than Moore's Law [34]. Genomic data are exceptionally sensitive, and increasingly so as advancements in bioinformatics

methods can uncover a patient's identity in a dataset with a small number of queries [35-39] through approaches like membership inference attack [40]. Addressing secure storage and sharing of genomic data to solve such issues is a key research challenge required to advance genomics-based precision health and medicine pipelines to the clinic [41]. Several methods for preserving genetic privacy have been published, including differential privacy-based approaches [42-44], perturbing the data with Bayesian statistics and Markov Chain Monte Carlo techniques [45], applying cryptographic protocols and frequency-based clinical genetics [10], and encrypting the data before offloading it to the cloud [46].

While the genome is a key data modality for precision health, it must be tightly tied to the phenotype, perhaps best embodied in electronic medical record (EMR) data. EMR can be mined to make data driven predictions about important biomedical issues such as the risk for diseases at the heart of immediate public health crises (i.e., COVID-19) [47-49], understudied and unknown adverse drug interactions [50-51], and psychiatric and behavioral conditions with a small number of behavioral biomarkers [52-56], including in underserved countries with differing laws and expectations about data sharing [57]. EMR are susceptible to attack, for example by inferring disease heritability from exposed pedigree information [58]. Previously explored solutions to addressing the sensitive nature of such records include only performing inference on common medical events while keeping the remainder private [59], reducing the dimensionality of the dataset [60-61], transforming the dataset with the use of generative adversarial networks [62], giving the patient control over who has access to the electronic health records [63], only allowing aggregate queries without revealing the underlying dataset [64], and deploying cryptography schemes such as symmetric key or asymmetric key encryption [65].

	Data Transformation	Access Control	Data Aggregation	Transparency
Preserving Privacy and Explaining Decisions of Artificial Intelligence	✓	✓	✓	✓
Sharing Genomic and Health Records	✓	✓	✓	✓
Deploying Digital Health Solutions	✓	✓	x	✓
Crowdsourcing Healthcare	x	✓	x	x

Figure 1. An opportunity space for innovation in methods for achieving trustworthy biomedical data solutions. We list the 4 most active areas where security and trust in the exchange of data is highest: private and explainable artificial intelligence; sharing and integration of genomic and medical records; construction and use of digital health tools; and crowdsourcing of healtcare management. In all 4, methodologies of data transformation, access control, data aggregation, and transparency can and should be deployed.

4. Deploying Digital Health Solutions

While EMR are traditionally generated in the clinic, digital health solutions are increasingly deployed to home settings [66-68]. As digital devices continue to receive FDA approval for medical use [69-70], it is inevitable, and exciting, that large portions of EMR data will be acquired through consumer devices such as smartphones and embedded hardware. Digital devices can longitudinally quantify patient symptoms when away from the clinic for conditions such as brain-mediated neurological and psychiatric disorders [71-72], cardiovascular disease [73-74], and infectious disease [75], among others. Examples of digital health solutions used in sensitive settings include therapeutic devices administered by clinicians [76], therapeutic tools administered in home settings [77-79], monitoring systems in hospital settings [80-81], dual-purpose interventions which explicitly collect patient health information to train machine learning models [82-84], pediatric healthcare interventions disguised to the child as a game [85-86], and wearable devices [87]. Many of these therapeutic and diagnostic devices collect potentially sensitive audio, image, and video streams for clinical use [88-91], and these data streams are often shared with clinicians or even crowdsourced with the consent of the patient. Furthermore, several digital therapies are used in home settings, and such rich data streams are filled with protected health information accompanied by potentially sensitive identifiable information such as the patient's face, images and video of the patient's home, and audio recordings of the patient or their family while using the device. It is therefore crucial to ensure patient privacy when these data leave the patient's device and are introduced into clinical workflows. Best practices discussed by Martínez-Pérez et al. include creating role-based access to data, making the privacy policy precise and clear to the user, transferring data with TLS using 256-bit encryption, erasing the data after it has been used for its intended purpose, and creating a data breach notification system [92].

Because consumer health technologies do not have direct oversight by clinicians, biased and deliberately inaccurate reporting by the target audience can be a risk. Therefore, it is particularly important to assess the quality of incoming data to garner the trust of healthcare providers and scientists, using those data for healthcare management and innovation. Algorithms that perform quality control to safeguard against biased or inaccurate reporting must go hand-in-hand with digital innovations. It is crucial for researchers to easily identify invalid or unintended data. For both consumers and scientists to gain confidence in the generalized applicability of digital tools, the data must be representative of the target population, making it pertinent to collect data that are balanced across race, ethnicity, geography, gender, and other relevant demographics.

5. Crowdsourcing Healthcare

Crowdsourcing is another approach used increasingly in clinical workflows [93-97]. Digital health and telemedical solutions that can scale through crowdsourcing approaches will become a norm for healthcare. The use of crowdsourcing in healthcare can be broadly partitioned into three categories: (1) crowdsourcing to achieve consensus on the presence or absence of medical conditions; (2) crowdsourced capture (whether active or passive, or a combination) of longitudinal data streams from from a large target cohort; (3) crowdsourcing the construction of training libraries of robustly labeled health data (e.g., radiological images), that enable progressive improvement of predictive models that can augment or replace decision points in the healthcare process.

Crowdsourcing appears in diverse healthcare settings and has been used for measurement of autism symptoms for diagnostic decision support [98-101], ranking adverse drug reactions [102], and COVID-19 contact tracing and surveillance [103-105]. Despite the strong clinical utility of crowdsourcing approaches, studies of trust and privacy for text, audio, image, and video streams rated on crowdsourcing platforms (e.g., Amazon Mechanical Turk [106-107] and Microworkers.com [108]) are lacking in the literature, especially with respect to biomedical research. As with digital consumer technologies, labeled data from crowdsourcing pipelines have the potential to suffer from low quality [109], requiring methods to filter crowd workers and into a trusted workforce of repeatedly high quality workers. This session includes a paper by Washington et al. which introduces quantitative metrics for evaluating crowd workers for their trustworthiness and reliability and provides behavioral metrics for identifying a valuable subset of crowd workers for inclusion in private clinical workflows. We hope that this study will inspire further work toward ensuring trustworthy crowd-powered telemedicine. Figure 1 highlights that research into trustworthy biomedical crowdsourcing is relatively light. In particular, privacy-preserved crowdsourced annotation of transformed data and on aggregate data is a currently unexplored yet fruitful research direction.

6. Considering the Bioethics

It is important to keep sight of the ethical considerations and formal bioethical perspectives with respect to biomedical innovations using trustworthy methods, or the lack thereof. Bioethical arguments are typically grounded in traditional ethical theories. Deontology is an ethical theory that considers actions as moral if they pass a series of conditions or rules [110]. A contrasting family of ethical theories, consequentialism, requires that moral actions maximize the public good and the utility of the action to all relevant stakeholders [110]. A third category, virtue ethics, states that moral actions should be a manifestation of a virtuous character trait [110]. While all ethical theories sound optimal in isolation, bioethical decisions may often satisfy one ethical theory while violating another. For example, heavy COVID-19 surveillance will maximize the good to all people (Utilitarianism, a type of consequentialism) while violating a core principle (deontological ethics) of the right to privacy. Bioethical analyses have been applied to genome sequencing for newborn screening [111-112], clinical machine learning [113-114], precision medicine [115-116], wearables and mobile health [117-118], and crowdsourcing [119-120].

This session includes a paper by Greenbaum et al. discussing the implications of expanded access programs with respect to COVID-19, a particularly timely topic. We hope that informaticians and scientists will interact more often with bioethicists to understand the societal implications of their work.

7. Anticipating the Future

Trustworthy biomedical data solutions will be crucial for realizing wide adoption of emerging technologies and methodologies for precision health. This session includes promising directions of exploration for the biomedical informatics research community. We have summarized some of the methods for building trust in key parts of the data analysis pipeline: data analysis(for artificial intelligence), data sharing (of genomic and health records), data capture (through digital devices), and data labeling (through crowdsourcing).

The study of trustworthy biomedical data science is in its infancy and ripe for innovations. We hope that this session will inspire further work in this important area, complementing the public's broader discussion of privacy and security considerations related to large-scale data collection and analysis. We anticipate that research that aims to improve the trustworthiness of biocomputing methods will become a major part of the PSB and a major focus for biomcomputing research in the coming years.

References

1. Bradshaw, Samantha, and Philip N. Howard. "Challenging truth and trust: A global inventory of organized social media manipulation." *The Computational Propaganda Project* 1 (2018).
2. Milne, Richard, Katherine I. Morley, Heidi Howard, Emilia Niemiec, Dianne Nicol, Christine Critchley, Barbara Prainsack et al. "Trust in genomic data sharing among members of the general public in the UK, USA, Canada and Australia." *Human genetics* 138, no. 11 (2019): 1237-1246.
3. Ovide, Shira. "Will More Data Make Us Healthier?" *The New York Times.* August 28, 2020.
4. Yan, Wei Qi. *Introduction to intelligent surveillance: Surveillance data capture, transmission, and analytics*. Springer, 2019.
5. Annas, George J. "HIPAA regulations—a new era of medical-record privacy?." (2003): 1486-1490.
6. Lin, Chi, Zihao Song, Houbing Song, Yanhong Zhou, Yi Wang, and Guowei Wu. "Differential privacy preserving in big data analytics for connected health." *Journal of medical systems* 40, no. 4 (2016): 97.
7. Lu, Rongxing, Xiaodong Lin, and Xuemin Shen. "SPOC: A secure and privacy-preserving opportunistic computing framework for mobile-healthcare emergency." *IEEE transactions on parallel and distributed systems* 24, no. 3 (2012): 614-624.
8. Drazen, Jeffrey M. "Sharing individual patient data from clinical trials." *New England Journal of Medicine* 372, no. 3 (2015): 201-202.
9. El Emam, Khaled, Sam Rodgers, and Bradley Malin. "Anonymising and sharing individual patient data." *bmj* 350 (2015): h1139.
10. Jagadeesh, Karthik A., David J. Wu, Johannes A. Birgmeier, Dan Boneh, and Gill Bejerano. "Deriving genomic diagnoses without revealing patient genomes." *Science* 357, no. 6352 (2017): 692-695.
11. Navarro, Robert. "An ethical framework for sharing patient data without consent." *Journal of Innovation in Health Informatics* 16, no. 4 (2008): 257-262.
12. Yeom, Samuel, Irene Giacomelli, Matt Fredrikson, and Somesh Jha. "Privacy risk in machine learning: Analyzing the connection to overfitting." In *2018 IEEE 31st Computer Security Foundations Symposium (CSF)*, pp. 268-282. IEEE, 2018.
13. Beaulieu-Jones, Brett K., Zhiwei Steven Wu, Chris Williams, Ran Lee, Sanjeev P. Bhavnani, James Brian Byrd, and Casey S. Greene. "Privacy-preserving generative deep neural networks support clinical data sharing." *Circulation: Cardiovascular Quality and Outcomes* 12, no. 7 (2019): e005122.
14. Singh, Amit, Albert Haque, Alexandre Alahi, Serena Yeung, Michelle Guo, Jill R. Glassman, William Beninati, Terry Platchek, Li Fei-Fei, and Arnold Milstein. "Automatic detection of hand hygiene using computer vision technology." *Journal of the American Medical Informatics Association* 27, no. 8 (2020): 1316-1320.
15. Yeung, Serena. "Visual Understanding of Human Activity: Towards Ambient Intelligence in AI-assisted Hospitals." PhD diss., Stanford University, 2018.
16. Yeung, Serena, N. Lance Downing, Li Fei-Fei, and Arnold Milstein. "Bedside computer vision-moving artificial intelligence from driver assistance to patient safety." *N Engl J Med* 378, no. 14 (2018): 1271-3.
17. Yeung, Serena, Francesca Rinaldo, Jeffrey Jopling, Bingbin Liu, Rishab Mehra, N. Lance Downing, Michelle Guo et al. "A computer vision system for deep learning-based detection of patient mobilization activities in the ICU." *NPJ digital medicine* 2, no. 1 (2019): 1-5.
18. Washington, Peter, Qandeel Tariq, Emilie Leblanc, Brianna Chrisman, Kaitlyn Dunlap, Aaron Kline et al. "Crowdsourced feature tagging for scalable autism diagnoses." *Under review.* 2020.
19. Yang, Qiang, Yang Liu, Tianjian Chen, and Yongxin Tong. "Federated machine learning: Concept and applications." *ACM Transactions on Intelligent Systems and Technology (TIST)* 10, no. 2 (2019): 1-19.

20. Brisimi, Theodora S., Ruidi Chen, Theofanie Mela, Alex Olshevsky, Ioannis Ch Paschalidis, and Wei Shi. "Federated learning of predictive models from federated electronic health records." *International journal of medical informatics* 112 (2018): 59-67.
21. Huang, Li, Yifeng Yin, Zeng Fu, Shifa Zhang, Hao Deng, and Dianbo Liu. "Loadaboost: Loss-based adaboost federated machine learning on medical data." *arXiv preprint arXiv:1811.12629* (2018).
22. Liu, Dianbo, Timothy Miller, Raheel Sayeed, and Kenneth D. Mandl. "Fadl: Federated-autonomous deep learning for distributed electronic health record." *arXiv preprint arXiv:1811.11400* (2018).
23. Chen, Yiqiang, Xin Qin, Jindong Wang, Chaohui Yu, and Wen Gao. "Fedhealth: A federated transfer learning framework for wearable healthcare." *IEEE Intelligent Systems* (2020).
24. Kaissis, Georgios A., Marcus R. Makowski, Daniel Rückert, and Rickmer F. Braren. "Secure, privacy-preserving and federated machine learning in medical imaging." *Nature Machine Intelligence* (2020): 1-7.
25. Dwork, Cynthia. "Differential privacy: A survey of results." In *International conference on theory and applications of models of computation*, pp. 1-19. Springer, Berlin, Heidelberg, 2008.
26. Lin, Zhen, Art B. Owen, and Russ B. Altman. "Genomic research and human subject privacy." (2004): 183-183.
27. Tramèr, Florian, Zhicong Huang, Jean-Pierre Hubaux, and Erman Ayday. "Differential privacy with bounded priors: reconciling utility and privacy in genome-wide association studies." In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pp. 1286-1297. 2015.
28. Phan, NhatHai, Yue Wang, Xintao Wu, and Dejing Dou. "Differential privacy preservation for deep auto-encoders: an application of human behavior prediction." In *Aaai*, vol. 16, pp. 1309-1316. 2016.
29. Gunning, David. "Explainable artificial intelligence (xai)." *Defense Advanced Research Projects Agency (DARPA), nd Web* 2 (2017): 2.
30. Prentzas, Nicoletta, Andrew Nicolaides, Efthyvoulos Kyriacou, Antonis Kakas, and Constantinos Pattichis. "Integrating Machine Learning with Symbolic Reasoning to Build an Explainable AI Model for Stroke Prediction." In *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, pp. 817-821. IEEE, 2019.
31. Carrieri, Anna Paola, Niina Haiminen, Sean Maudsley-Barton, Laura-Jayne Gardiner, Barry Murphy, Andrew Mayes, Sarah Paterson et al. "Explainable AI reveals key changes in skin microbiome associated with menopause, smoking, aging and skin hydration." *bioRxiv* (2020).
32. Lauritsen, Simon Meyer, Mads Kristensen, Mathias Vassard Olsen, Morten Skaarup Larsen, Katrine Meyer Lauritsen, Marianne Johansson Jørgensen, Jeppe Lange, and Bo Thiesson. "Explainable artificial intelligence model to predict acute critical illness from electronic health records." *Nature communications* 11, no. 1 (2020): 1-11.
33. Spencer, Matt, Saad Khan, Zohreh Talebizadeh, and Chi-Ren Shyu. "Explainable AI: Mining of Genotype Data Identifies Complex Disease Pathways—Autism Case Studies." *Application Of Omics, Ai And Blockchain In Bioinformatics Research* 21 (2019): 11.
34. Muir, Paul, Shantao Li, Shaoke Lou, Daifeng Wang, Daniel J. Spakowicz, Leonidas Salichos, Jing Zhang et al. "The real cost of sequencing: scaling computation to keep pace with data generation." *Genome biology* 17, no. 1 (2016): 1-9.
35. Al Aziz, Md Momin, Reza Ghasemi, Md Waliullah, and Noman Mohammed. "Aftermath of bustamante attack on genomic beacon service." *BMC medical genomics* 10, no. 2 (2017): 43.
36. Backes, Michael, Pascal Berrang, Mathias Humbert, and Praveen Manoharan. "Membership privacy in MicroRNA-based studies." In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 319-330. 2016.
37. Deznabi, Iman, Mohammad Mobayen, Nazanin Jafari, Oznur Tastan, and Erman Ayday. "An inference attack on genomic data using kinship, complex correlations, and phenotype information." *IEEE/ACM transactions on computational biology and bioinformatics* 15, no. 4 (2017): 1333-1343.
38. Humbert, Mathias, Erman Ayday, Jean-Pierre Hubaux, and Amalio Telenti. "Addressing the concerns of the lacks family: quantification of kin genomic privacy." In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pp. 1141-1152. 2013.
39. Shringarpure, Suyash S., and Carlos D. Bustamante. "Privacy risks from genomic data-sharing beacons." *The American Journal of Human Genetics* 97, no. 5 (2015): 631-646.

40. Shokri, Reza, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. "Membership inference attacks against machine learning models." In *2017 IEEE Symposium on Security and Privacy (SP)*, pp. 3-18. IEEE, 2017.
41. Altman, Russ B., Snehit Prabhu, Arend Sidow, Justin M. Zook, Rachel Goldfeder, David Litwack, Euan Ashley et al. "A research roadmap for next-generation sequencing informatics." *Science translational medicine* 8, no. 335 (2016): 335ps10-335ps10.
42. Almadhoun, Nour, Erman Ayday, and Özgür Ulusoy. "Differential privacy under dependent tuples—the case of genomic privacy." *Bioinformatics* 36, no. 6 (2020): 1696-1703.
43. He, Zaobo, Yingshu Li, and Jinbao Wang. "Differential privacy preserving genomic data releasing via factor graph." In *International Symposium on Bioinformatics Research and Applications*, pp. 350-355. Springer, Cham, 2017.
44. Raisaro, Jean Louis, Gwangbae Choi, Sylvain Pradervand, Raphael Colsenet, Nathalie Jacquemont, Nicolas Rosat, Vincent Mooser, and Jean-Pierre Hubaux. "Protecting privacy and security of genomic data in I2B2 with homomorphic encryption and differential privacy." *IEEE/ACM transactions on computational biology and bioinformatics* 15, no. 5 (2018): 1413-1426.
45. Simmons, Sean, Bonnie Berger, and Cenk S. Sahinalp. "Protecting Genomic Data Privacy with Probabilistic Modeling." In *PSB*, pp. 403-414. 2019.
46. Wang, Bing, Wei Song, Wenjing Lou, and Y. Thomas Hou. "Privacy-preserving pattern matching over encrypted genetic data in cloud computing." In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pp. 1-9. IEEE, 2017.
47. Vaid, Akhil, Sulaiman Somani, Adam J. Russak, Jessica K. De Freitas, Fayzan F. Chaudhry, Ishan Paranjpe, Kipp W. Johnson et al. "Machine Learning to Predict Mortality and Critical Events in COVID-19 Positive New York City Patients." *medRxiv* (2020).
48. Yoo, Edwin, Bethany Percha, Max Tomlinson, Victor Razuk, Stephanie Pan, Madeleine Basist, Pranai Tandon et al. "Development and calibration of a simple mortality risk score for hospitalized COVID-19 adults." *medRxiv* (2020).
49. Zietz, Michael, and Nicholas P. Tatonetti. "Testing the association between blood type and COVID-19 infection, intubation, and death." *MedRxiv* (2020).
50. Basile, Anna O., Alexandre Yahi, and Nicholas P. Tatonetti. "Artificial intelligence for drug toxicity and safety." *Trends in pharmacological sciences* 40, no. 9 (2019): 624-635.
51. Percha, Bethany, and Russ B. Altman. "Informatics confronts drug–drug interactions." *Trends in pharmacological sciences* 34, no. 3 (2013): 178-184.
52. Duda, M., N. Haber, J. Daniels, and D. P. Wall. "Crowdsourced validation of a machine-learning classification system for autism and ADHD." *Translational psychiatry* 7, no. 5 (2017): e1133-e1133.
53. Duda, M., J. A. Kosmicki, and D. P. Wall. "Testing the accuracy of an observation-based classifier for rapid detection of autism risk." *Translational psychiatry* 4, no. 8 (2014): e424-e424.
54. Lyalina, Svetlana, Bethany Percha, Paea LePendu, Srinivasan V. Iyer, Russ B. Altman, and Nigam H. Shah. "Identifying phenotypic signatures of neuropsychiatric disorders from electronic medical records." *Journal of the American Medical Informatics Association* 20, no. e2 (2013): e297-e305.
55. Tariq, Qandeel, Jena Daniels, Jessey Nicole Schwartz, Peter Washington, Haik Kalantarian, and Dennis Paul Wall. "Mobile detection of autism through machine learning on home video: A development and prospective validation study." *PLoS medicine* 15, no. 11 (2018): e1002705.
56. Wall, Dennis Paul, J. Kosmicki, T. F. Deluca, E. Harstad, and Vincent Alfred Fusaro. "Use of machine learning to shorten observation-based screening and diagnosis of autism." *Translational psychiatry* 2, no. 4 (2012): e100-e100.
57. Tariq, Qandeel, Scott Lanyon Fleming, Jessey Nicole Schwartz, Kaitlyn Dunlap, Conor Corbin, Peter Washington, Haik Kalantarian, Naila Z. Khan, Gary L. Darmstadt, and Dennis Paul Wall. "Detecting developmental delay and autism through machine learning models using home videos of Bangladeshi children: Development and validation study." *Journal of medical Internet research* 21, no. 4 (2019): e13822.
58. Polubriaginof, Fernanda CG, Rami Vanguri, Kayla Quinnies, Gillian M. Belbin, Alexandre Yahi, Hojjat Salmasian, Tal Lorberbaum et al. "Disease heritability inferred from familial relationships reported in medical records." *Cell* 173, no. 7 (2018): 1692-1704.
59. Tatonetti, Nicholas, Russ B. Altman, and Guy Haskin Fernald. "Signal detection algorithms to identify drug effects and drug interactions." U.S. Patent 9,305,267, issued April 5, 2016.

60. Johnson, Kipp W., Jessica K. De Freitas, Benjamin S. Glicksberg, Jason R. Bobe, and Joel T. Dudley. "Evaluation of patient re-identification using laboratory test orders and mitigation via latent space variables." In *PSB*, pp. 415-426. 2019.
61. Washington, Peter, Kelley Marie Paskov, Haik Kalantarian, Nathaniel Stockham, Catalin Voss, Aaron Kline, Ritik Patnaik et al. "Feature selection and dimension reduction of social autism data." In *Pac Symp Biocomput*, vol. 25, pp. 707-718. 2020.
62. Bae, Ho, Dahuin Jung, and Sungroh Yoon. "AnomiGAN: Generative adversarial networks for anonymizing private medical data." *arXiv preprint arXiv:1901.11313* (2019).
63. Demuynck, Liesje, and Bart De Decker. "Privacy-preserving electronic health records." In *IFIP International Conference on Communications and Multimedia Security*, pp. 150-159. Springer, Berlin, Heidelberg, 2005.
64. Luthria, Gaurav, and Qingbo Wang. "Implementing a Cloud Based Method for Protected Clinical Trial Data Sharing." In *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, vol. 25, pp. 647-658. NIH Public Access, 2020.
65. Fernández-Alemán, José Luis, Inmaculada Carrión Señor, Pedro Ángel Oliver Lozoya, and Ambrosio Toval. "Security and privacy in electronic health records: A systematic literature review." *Journal of biomedical informatics* 46, no. 3 (2013): 541-562.
66. Istepanian, Robert, Swamy Laxminarayan, and Constantinos S. Pattichis, eds. *M-health: Emerging mobile health systems*. Springer Science & Business Media, 2007.
67. Lupton, Deborah. *Digital health: critical and cross-disciplinary perspectives*. Routledge, 2017.
68. Murray, Elizabeth, Eric B. Hekler, Gerhard Andersson, Linda M. Collins, Aiden Doherty, Chris Hollis, Daniel E. Rivera, Robert West, and Jeremy C. Wyatt. "Evaluating digital health interventions: key questions and approaches." (2016): 843-851.
69. Maisel, William H. "Medical device regulation: an introduction for the practicing physician." *Annals of internal medicine* 140, no. 4 (2004): 296-302.
70. Zuckerman, Diana M., Paul Brown, and Steven E. Nissen. "Medical device recalls and the FDA approval process." *Archives of internal medicine* 171, no. 11 (2011): 1006-1011.
71. Stark, David E., Rajiv B. Kumar, Christopher A. Longhurst, and Dennis P. Wall. "The quantified brain: a framework for mobile device-based assessment of behavior and neurological function." *Applied clinical informatics* 7, no. 2 (2016): 290.
72. Torous, John, and Laura Weiss Roberts. "Needed innovation in digital health and smartphone applications for mental health: transparency and trust." *JAMA psychiatry* 74, no. 5 (2017): 437-438.
73. McConnell, Michael V., Mintu P. Turakhia, Robert A. Harrington, Abby C. King, and Euan A. Ashley. "Mobile health advances in physical activity, fitness, and atrial fibrillation: moving hearts." *Journal of the American College of Cardiology* 71, no. 23 (2018): 2691-2701.
74. McConnell, Michael V., Anna Shcherbina, Aleksandra Pavlovic, Julian R. Homburger, Rachel L. Goldfeder, Daryl Waggot, Mildred K. Cho et al. "Feasibility of obtaining measures of lifestyle from a smartphone app: the MyHeart Counts Cardiovascular Health Study." *JAMA cardiology* 2, no. 1 (2017): 67-76.
75. Ngwatu, Brian Kermu, Ntwali Placide Nsengiyumva, Olivia Oxlade, Benjamin Mappin-Kasirer, Nhat Linh Nguyen, Ernesto Jaramillo, Dennis Falzon, and Kevin Schwartzman. "The impact of digital health technologies on tuberculosis treatment: a systematic review." *European Respiratory Journal* 51, no. 1 (2018).
76. Washington, Peter, Catalin Voss, Nick Haber, Serena Tanaka, Jena Daniels, Carl Feinstein, Terry Winograd, and Dennis Wall. "A wearable social interaction aid for children with autism." In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2348-2354. 2016.
77. Kalantarian, Haik, Khaled Jedoui, Peter Washington, and Dennis P. Wall. "A mobile game for automatic emotion-labeling of images." *IEEE Transactions on Games* (2018).
78. Kline, Aaron, Catalin Voss, Peter Washington, Nick Haber, Hessey Schwartz, Qandeel Tariq, Terry Winograd, Carl Feinstein, and Dennis P. Wall. "Superpower glass." *GetMobile: Mobile Computing and Communications* 23, no. 2 (2019): 35-38.
79. Washington, Peter, Catalin Voss, Aaron Kline, Nick Haber, Jena Daniels, Azar Fazel, Titas De, Carl Feinstein, Terry Winograd, and Dennis Wall. "SuperpowerGlass: a wearable aid for the at-home therapy of children with autism." *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, no. 3 (2017): 1-22.

80. Daniels, Jena, Nick Haber, Catalin Voss, Jessey Schwartz, Serena Tamura, Azar Fazel, Aaron Kline et al. "Feasibility testing of a wearable behavioral aid for social learning in children with autism." *Applied clinical informatics* 9, no. 1 (2018): 129.
81. Waran, Vicknes, Nor Faizal Ahmad Bahuri, Vairavan Narayanan, Dharmendra Ganesan, and Khairul Azmi Abdul Kadir. "Video clip transfer of radiological images using a mobile telephone in emergency neurosurgical consultations (3G Multi-Media Messaging Service)." *British journal of neurosurgery* 26, no. 2 (2012): 199-201.
82. Kalantarian, Haik, Khaled Jedoui, Peter Washington, Qandeel Tariq, Kaiti Dunlap, Jessey Schwartz, and Dennis P. Wall. "Labeling images with facial emotion and the potential for pediatric healthcare." *Artificial intelligence in medicine* 98 (2019): 77-86.
83. Voss, Catalin, Peter Washington, Nick Haber, Aaron Kline, Jena Daniels, Azar Fazel, Titas De et al. "Superpower glass: delivering unobtrusive real-time social cues in wearable systems." In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pp. 1218-1226. 2016.
84. Kalantarian, Haik, Khaled Jedoui, Kaitlyn Dunlap, Jessey Schwartz, Peter Washington, Arman Husic, Qandeel Tariq, Michael Ning, Aaron Kline, and Dennis Paul Wall. "The Performance of Emotion Classifiers for Children With Parent-Reported Autism: Quantitative Feasibility Study." *JMIR Mental Health* 7, no. 4 (2020): e13174.
85. Daniels, Jena, Jessey N. Schwartz, Catalin Voss, Nick Haber, Azar Fazel, Aaron Kline, Peter Washington, Carl Feinstein, Terry Winograd, and Dennis P. Wall. "Exploratory study examining the at-home feasibility of a wearable tool for social-affective learning in children with autism." *NPJ digital medicine* 1, no. 1 (2018): 1-10.
86. Kalantarian, Haik, Peter Washington, Jessey Schwartz, Jena Daniels, Nick Haber, and Dennis Wall. "A gamified mobile system for crowdsourcing video for autism research." In *2018 IEEE international conference on healthcare informatics (ICHI)*, pp. 350-352. IEEE, 2018.
87. Voss, Catalin, Jessey Schwartz, Jena Daniels, Aaron Kline, Nick Haber, Peter Washington, Qandeel Tariq et al. "Effect of wearable digital intervention for improving socialization in children with autism spectrum disorder: a randomized clinical trial." *JAMA pediatrics* 173, no. 5 (2019): 446-454.
88. Brown, Stephen James. "Interactive video based remote health monitoring system." U.S. Patent 7,979,284, issued July 12, 2011.
89. Kalantarian, Haik, Peter Washington, Jessey Schwartz, Jena Daniels, Nick Haber, and Dennis P. Wall. "Guess what?." *Journal of Healthcare Informatics Research* 3, no. 1 (2019): 43-66.
90. Ramanujam, Bhargavi, Deepa Dash, and Manjari Tripathi. "Can home videos made on smartphones complement video-EEG in diagnosing psychogenic nonepileptic seizures?." *Seizure* 62 (2018): 95-98.
91. Rao, Sira P., Nikil S. Jayant, Max E. Stachura, Elena Astapova, and Anthony Pearson-Shaver. "Delivering diagnostic quality video over mobile wireless networks for telemedicine." *International Journal of Telemedicine and Applications* 2009 (2009).
92. Martínez-Pérez, Borja, Isabel De La Torre-Díez, and Miguel López-Coronado. "Privacy and security in mobile health apps: a review and recommendations." *Journal of medical systems* 39, no. 1 (2015): 181.
93. Celi, Leo Anthony, Andrea Ippolito, Robert A. Montgomery, Christopher Moses, and David J. Stone. "Crowdsourcing knowledge discovery and innovations in medicine." *Journal of medical Internet research* 16, no. 9 (2014): e216.
94. Créquit, Perrine, Ghizlène Mansouri, Mehdi Benchoufi, Alexandre Vivot, and Philippe Ravaud. "Mapping of crowdsourcing in health: systematic review." *Journal of medical Internet research* 20, no. 5 (2018): e187.
95. Ranard, Benjamin L., Yoonhee P. Ha, Zachary F. Meisel, David A. Asch, Shawndra S. Hill, Lance B. Becker, Anne K. Seymour, and Raina M. Merchant. "Crowdsourcing—harnessing the masses to advance health and medicine, a systematic review." *Journal of general internal medicine* 29, no. 1 (2014): 187-203.
96. Swan, Melanie. "Health 2050: The realization of personalized medicine through crowdsourcing, the quantified self, and the participatory biocitizen." *Journal of personalized medicine* 2, no. 3 (2012): 93-118.
97. Wazny, Kerri. "Applications of crowdsourcing in health: an overview." *Journal of global health* 8, no. 1 (2018).
98. David, Maude M., Brooke A. Babineau, and Dennis P. Wall. "Can we accelerate autism discoveries through crowdsourcing?." *Research in Autism Spectrum Disorders* 32 (2016): 80-83.

99. Washington, Peter, Haik Kalantarian, Qandeel Tariq, Jessey Schwartz, Kaitlyn Dunlap, Brianna Chrisman, Maya Varma et al. "Validity of online screening for autism: crowdsourcing study comparing paid and unpaid diagnostic tasks." *Journal of medical Internet research* 21, no. 5 (2019): e13668.
100. Washington, Peter, Emilie Leblanc, Kaitlyn Dunlap, Yordan Penev, Aaron Kline, Kelley Paskov, Min Woo Sun et al. "Precision Telemedicine through Crowdsourced Machine Learning: Testing Variability of Crowd Workers for Video-Based Autism Feature Recognition." *Journal of personalized medicine* 10, no. 3 (2020): 86.
101. Washington, Peter, Natalie Park, Parishkrita Srivastava, Catalin Voss, Aaron Kline, Maya Varma, Qandeel Tariq et al. "Data-driven diagnostics and the potential of mobile artificial intelligence for digital therapeutic phenotyping in computational psychiatry." *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* (2019).
102. Gottlieb, Assaf, Robert Hoehndorf, Michel Dumontier, and Russ B. Altman. "Ranking adverse drug reactions with crowdsourcing." *Journal of medical Internet research* 17, no. 3 (2015): e80.
103. Budd, Jobie, Benjamin S. Miller, Erin M. Manning, Vasileios Lampos, Mengdie Zhuang, Michael Edelstein, Geraint Rees et al. "Digital technologies in the public-health response to COVID-19." *Nature medicine* (2020): 1-10.
104. Hegde, Ajay, Ramesh Masthi, and Darshan Krishnappa. "Hyperlocal Postcode Based Crowdsourced Surveillance Systems in the COVID-19 Pandemic Response." *Frontiers in Public Health* 8 (2020): 286.
105. Sun, Kaiyuan, Jenny Chen, and Cécile Viboud. "Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study." *The Lancet Digital Health* (2020).
106. Kittur, Aniket, Ed H. Chi, and Bongwon Suh. "Crowdsourcing user studies with Mechanical Turk." In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 453-456. 2008.
107. Paolacci, Gabriele, Jesse Chandler, and Panagiotis G. Ipeirotis. "Running experiments on amazon mechanical turk." *Judgment and Decision making* 5, no. 5 (2010): 411-419.
108. Hirth, Matthias, Tobias Hoßfeld, and Phuoc Tran-Gia. "Anatomy of a crowdsourcing platform-using the example of microworkers. com." In *2011 Fifth international conference on innovative mobile and internet services in ubiquitous computing*, pp. 322-329. IEEE, 2011.
109. Raykar, Vikas C., and Shipeng Yu. "Eliminating spammers and ranking annotators for crowdsourced labeling tasks." *The Journal of Machine Learning Research* 13, no. 1 (2012): 491-518.
110. Beauchamp, Tom L., and James F. Childress. *Principles of biomedical ethics*. Oxford University Press, USA, 2001.
111. Botkin, Jeffrey R., and Erin Rothwell. "Whole genome sequencing and newborn screening." *Current genetic medicine reports* 4, no. 1 (2016): 1-6.
112. Howard, Heidi Carmen, Bartha Maria Knoppers, Martina C. Cornel, Ellen Wright Clayton, Karine Sénécal, and Pascal Borry. "Whole-genome sequencing in newborn screening? A statement on the continued importance of targeted approaches in newborn screening programmes." *European Journal of Human Genetics* 23, no. 12 (2015): 1593-1600.
113. Char, Danton S., Nigam H. Shah, and David Magnus. "Implementing machine learning in health care—addressing ethical challenges." *The New England journal of medicine* 378, no. 11 (2018): 981.
114. Cohen, I. Glenn, Ruben Amarasingham, Anand Shah, Bin Xie, and Bernard Lo. "The legal and ethical concerns that arise from using complex predictive analytics in health care." *Health affairs* 33, no. 7 (2014): 1139-1147.
115. Korngiebel, Diane M., Kenneth E. Thummel, and Wylie Burke. "Implementing precision medicine: the ethical challenges." *Trends in pharmacological sciences* 38, no. 1 (2017): 8-14.
116. Minari, Jusaku, Kyle B. Brothers, and Michael Morrison. "Tensions in ethics and policy created by National Precision Medicine Programs." *Human genomics* 12, no. 1 (2018): 1-10.
117. Kreitmair, Karola V., Mildred K. Cho, and David C. Magnus. "Consent and engagement, security, and authentic living using wearable and mobile health technology." *Nature biotechnology* 35, no. 7 (2017): 617-620.
118. Torous, John, and Laura Weiss Roberts. "The ethical use of mobile health technology in clinical psychiatry." *The Journal of nervous and mental disease* 205, no. 1 (2017): 4-8.
119. Fort, Karën, Gilles Adda, and K. Bretonnel Cohen. "Amazon mechanical turk: Gold mine or coal mine?." *Computational Linguistics* 37, no. 2 (2011): 413-420.

120. Kreitmair, Karola V., and David C. Magnus. "Citizen science and gamification." *Hastings center report* 49, no. 2 (2019): 40-46.