

PERSPECTIVE OPEN



An intriguing vision for transatlantic collaborative health data use and artificial intelligence development

Daniel C. Baumgart ¹✉

Our traditional approach to diagnosis, prognosis, and treatment, can no longer process and transform the enormous volume of information into therapeutic success, innovative discovery, and health economic performance. Precision health, i.e., the right treatment, for the right person, at the right time in the right place, is enabled through a learning health system, in which medicine and multidisciplinary science, economic viability, diverse culture, and empowered patient's preferences are digitally integrated and conceptually aligned for continuous improvement and maintenance of health, wellbeing, and equity. Artificial intelligence (AI) has been successfully evaluated in risk stratification, accurate diagnosis, and treatment allocation, and to prevent health disparities. There is one caveat though: dependable AI models need to be trained on population-representative, large and deep data sets by multidisciplinary and multinational teams to avoid developer, statistical and social bias. Such applications and models can neither be created nor validated with data at the country, let alone institutional level and require a new dimension of collaboration, a cultural change with the establishment of trust in a precompetitive space. The Data for Health (#DFH23) conference in Berlin and the Follow-Up Workshop at Harvard University in Boston hosted a representative group of stakeholders in society, academia, industry, and government. With the momentum #DFH23 created, the European Health Data Space (EHDS) as a solid and safe foundation for consented collaborative health data use and the G7 Hiroshima AI process in place, we call on citizens and their governments to fully support digital transformation of medicine, research and innovation including AI.

npj Digital Medicine (2024)7:19; <https://doi.org/10.1038/s41746-024-01005-y>

INTRODUCTION

Traditionally, patients present to healthcare systems with specific symptoms or through a preventive healthcare program. Healthcare providers then capture information from the patient's medical history, perform a physical exam, review clinical data, order additional investigations, before eventually arriving at a diagnosis, prognosis and therapy based on their own medical knowledge, clinical experience and a thought and decision process. As the case evolves, decisions are critically revisited and adjusted. Over time healthcare professionals *learn* from their intervention outcomes and increasingly gather experience *informing* their future decisions.

Medical knowledge and data are growing exponentially. The body of medical knowledge and clinical experience a healthcare professional can acquire and apply depends on the individual's medical education and training, motivation, and support provided within their health system. It can multiply and grow faster when working in teams and across disciplines. Ultimately however, effective and efficient medical knowledge acquisition and translation are a function of capacity and time limited by the individual's life span.

Thus, our traditional approach to diagnosis, prognosis, and treatment, can no longer process and transform the enormous volume of information into therapeutic success, innovative discovery, and health economic performance. The digital revolution in medicine began two decades ago¹. Advancement of information and communication technology, powerful computer science methodology and digital innovation can help accelerate, enhance and refine medical knowledge acquisition and translation.

Precision health, i.e., the right targeted treatment², for the right person, at the right time in the right place, is enabled through a *learning health system*, in which medicine and multidisciplinary science, economic viability, diverse culture, and empowered patient's preferences are *digitally integrated* and *conceptually aligned* for *continuous improvement and maintenance of health, wellbeing, and equity*.

Healthcare-embedded research with real-world-evidence³ generation may overcome some of the limitations of traditional clinical trials. Artificial intelligence (AI)⁴ including machine learning⁵ has been successfully evaluated in risk stratification, accurate diagnosis and treatment allocation, and to prevent health disparities⁶. Large language models make unstructured data accessible⁷. AI can limit harm in drug development⁸, by deploying computational tools to derive biological insights from large amounts of data, and exploit *"in silico"* computer-assisted structure and binding prediction^{9–11}, synthesis^{12–15}. and testing well before any trial participant is ever exposed.

Healthcare facilities can become safer¹⁶ and smarter¹⁷ through AI enabled patient^{18,19} and staff monitoring²⁰, policy enforcement²¹, provider competency assessment^{22,23}, and resource allocation^{24,25}. Wearables promote continuous, citizen centered, preventive care²⁶. Public health threats can be taken on earlier and more effectively^{27,28}.

There is one caveat though: dependable AI models need to be trained on population-representative, large and deep data sets by multidisciplinary and multinational teams to avoid developer, statistical and social bias²⁹. Moreover, complex diseases can only be understood through *integration*³⁰ of basic research and health-related and -defining information (e.g. environmental-, socio-economic-, infrastructural-, and behavioral-, and social data). Such

¹Precision Health Signature Area, College of Health Sciences, College of Natural and Applied Sciences all at University of Alberta, Edmonton, Alberta, Canada.

✉email: baumgart@ualberta.ca

applications and models *can neither be created nor validated with data at country*, let alone *institutional level* and require a new dimension of collaboration. Federated learning, i.e., use of *decentralized data sources* to build models and transfer learning, enables multiple institutions to collaborate *without changing the physical location of data*. While federated learning still requires a central coordinator, swarm learning³¹ circumvents it by uniting edge computing³² and blockchain³³ enabled asynchronous convergence.

However, other applications require true collaborative data use as envisioned by the European Commission's European Health Data Space (EHDS)³⁴. The EHDS is an ecosystem comprised of rules, standards and practices, infrastructures, and a governance framework empowering individuals through increased digital access to their personal health data, at the national and EU level, support their *free* movement, and foster a genuine single market for electronic health record systems, relevant medical devices, and health AI. It will provide a consistent, trustworthy, and efficient set-up for the use of health data for research, innovation, policy-making and regulatory activities.

Collaborative health data use and AI deployment at scale are disruptive advances that potentially expose us to new risks (i.e., security, data manipulation, misuse, failure to recognize model limitations, and moral ethical dilemmas) that need to be carefully weighed against the above-outlined benefits. Balancing both in medicine and healthcare is perhaps one of the most daunting tasks, as it involves the most sensitive and vulnerable aspects of our lives, but also one where we can have the greatest positive impact at an individual and societal level.

Moral dilemmas for instance arise, when choosing between two AI-enabled accurate, yet equally unfavorable decisions and their consequences³⁵. An AI-facilitated therapy recommendation can only address the data-driven aspect of personalized medicine, whereas the individual patient's choice to accept or reject a therapy recommendation (i.e. choosing a potentially less effective therapy for personal reasons) remains unchanged. Ultimately, the role of both AI algorithms and healthcare professionals remains to help patients (and families) make their own, hopefully increasingly better-informed decisions instead of making those decisions for them.

Ethical challenges do not only result from data access, data ownership and training data quality-related bias. Ethical conflicts anticipated to occur with algorithms may in fact not be rooted in AI itself, but rather exposed by it. Today's physicians and biomedical researchers find themselves in a conflicted situation to serve their patients, their employers and payors, when other components of the healthcare system may have different interests (i.e. effort & expenditures vs. revenue). These interests determine the goals of AI algorithm development⁵.

The German Federal Ministry of Health in collaboration with the German Aerospace Center (DLR) organized a high-level Data for Health conference³⁶ in Berlin, Germany in June and a Follow-Up Workshop³⁷ in September at Harvard University in Boston, MA, USA (#DFH23) with a representative group of stakeholders in society, academia, industry, and government along with journalists and both vocal supporters and critics for open and productive discussions. It takes that joint perspective to drive the cultural change needed.

#DFH23 explored various ways of taking the EHDS concept across the Atlantic. Three G7 countries, Germany, the United States and Canada lead the way. The German Federal Ministry of Health is committed to pass legislation to enable collaborative health data use and AI, deploy a national electronic health record aligned with the EHDS, and paperless communication in acute and long-term care. In July the European Commission adopted its adequacy decision for the EU-US Data Privacy Framework, enabling European entities to transfer personal data to participating entities in the US. Government consultations are underway to expand longstanding collaboration in AI with Canada³⁸. #DFH23

developed harmonized contractual language and multi-lingual consent documents as a start, realizing many other legal and political challenges remain. Figure 1

Digital inclusion³⁹ with engagement of all stakeholders, the reduction of barriers for those with disabilities and the elderly, are critically important for success and have become a social determinant of health. Translation into daily practice critically depends on overcoming the knowledge asymmetry through education and training. To build trust, several model projects with explicit citizen and patient participation demonstrating tangible benefits of collaborative data use and AI to promote healthy living and well-being, advance cancer detection and care, streamline data access, verify AI readiness, accelerate knowledge translation into everyday healthcare were developed and agreed in bar camps, workshops and discussed on the podium at #DFH23.

As the UNs WHO and ITU establish a benchmarking process for AI in health, regulatory bodies need to adapt their workflows⁴⁰. The US FDA decided to regulate AI as a device in line with the International Medical Device Regulators Forum's commitment to develop a harmonized approach to the management of AI⁴¹. Transparency (explainable AI [XAI]), decision autonomy level, decision effect size, and opportunity of human oversight and override will be critical for approval. Post marketing surveillance and liability of developers and users are complex tasks since medical, unlike industrial AI models, are not static over their lifecycle^{42,43}.

Rising concerns of cyber security with the exploitation of personal data⁴⁴ have led the EU to foster citizens' right to informational self-determination through the General Data Protection Regulation⁴⁵. Narrow GDPR interpretation may however fail citizens by depriving them of the profound benefits of collaborative data use and AI. #DFH23 attendees propose a legislative and enforcement effort to *prioritize regulating data use* not just protection. Effective legal deterrents rather than unreasonable measures to avoid criminal decryption and re-identification are urgently needed.

In October, one month after the #DFH23 Follow-Up Workshop³⁷ at Harvard University the entire G7, i.e. Canada, France, Germany, Italy, Japan, the United Kingdom, the United States as well the EU passed an agreement on International Guiding Principles on Artificial Intelligence (AI)⁴⁶ and a voluntary Code of Conduct⁴⁷ for AI developers referred to as the Hiroshima AI process.

Based on what we know today, it is unlikely that there will be true competition between humans and machines in healthcare compared with the manufacturing or service industries. Although service robots have been experimentally deployed to hospitals and publicly shared places during the COVID-19 pandemic to clean, take temperatures and deliver food, minimize person contacts, this cannot be the primary goal to address manpower challenges or achieve cost-control in healthcare⁴⁸. Human healthcare professionals and machines can supplement each other, as both have unique skills and strengths. In fact, it is conceivable in our competitive, increasingly performance-driven healthcare systems, clinicians will have more time again to spend with their patients, a critical element of empathetic healthcare that has been missed lately due to the economic pressures imposed by hospital operators, payors and governments aiming to lower costs.

With the momentum #DFH23 created, the EHDS as a solid and safe foundation for consented national, EU-wide and transatlantic collaborative health data use as well as the G7 Hiroshima AI process in place, we call on citizens and their governments to fully support digital transformation of medicine, research and innovation including AI implementation. It's time to rethink healthcare. Our citizens and patients deserve it.



Transatlantic Collaborative Health Data Use and AI Development Inspired by the European Health Data Space

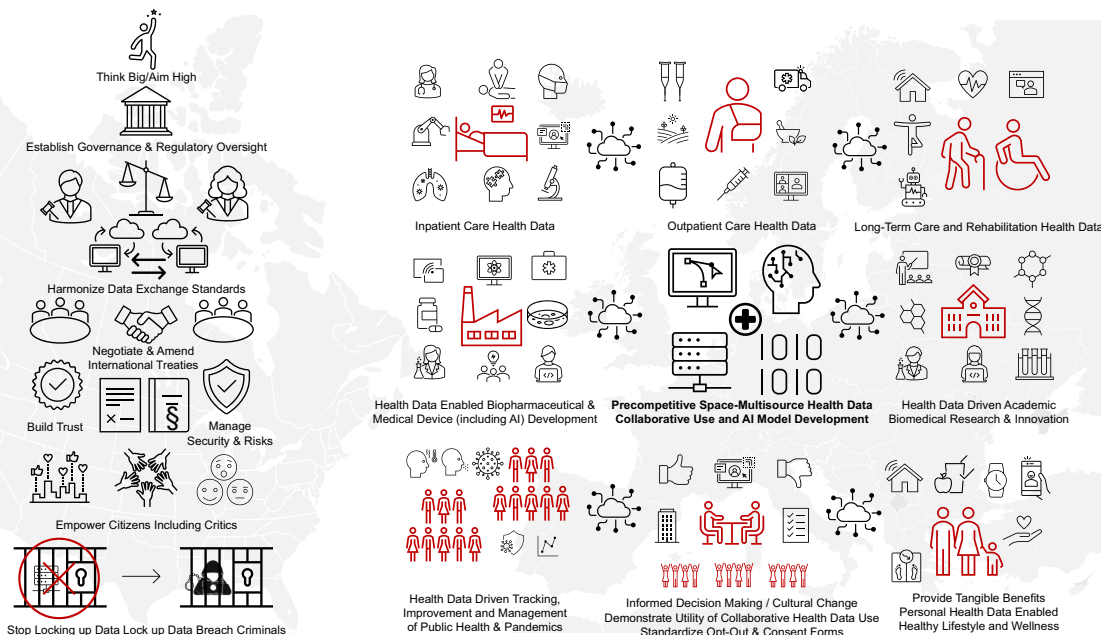


Fig. 1 EHDS inspired transatlantic collaborative health data use and artificial intelligence development. Collaborative health data enabled innovation requires cultural change with establishment of trust among all stakeholders in a precompetitive space. Trustworthy communication of goals, risks, benefits and challenges, continuous dialogue with citizens and patients (i.e. joint AI readiness assessment with patients), prejudice free consent and opt-out mechanisms, harmonization of technical and administrative standards (i.e. an international patient summary, harmonized transatlantic patient consent forms) will enable interoperability, aligned regulatory processes (i.e. development of regulatory fact sheets), model use cases and exemplary projects (i.e. a cancer genome tracker, a registry for cancer immune therapy) and a streamlined exchange platform for citizens, researchers, developers and innovators (i.e. a stake holder council, online R&D “tool box”).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

This manuscript does not include any additional data that could be made available.

Received: 14 November 2023; Accepted: 3 January 2024;

Published online: 23 January 2024

REFERENCES

- Baumgart, D. C. Personal digital assistants in health care: experienced clinicians in the palm of your hand? *Lancet* **366**, 1210–1222 (2005).
- Baumgart, D. C. & Le Berre, C. Newer Biologic and Small-Molecule Therapies for Inflammatory Bowel Disease. *N. Engl. J. Med* **385**, 1302–1315 (2021).
- Stern, A. D. et al. Advancing digital health applications: priorities for innovation in real-world evidence generation. *Lancet Digit Health* **4**, e200–e206 (2022).
- Haug, C. J. & Drazen, J. M. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. *N. Engl. J. Med* **388**, 1201–1208 (2023).
- Rajkomar, A., Dean, J. & Kohane, I. Machine Learning in Medicine. *N. Engl. J. Med* **380**, 1347–1358 (2019).
- Pierson, E., Cutler, D. M., Leskovec, J., Mullainathan, S. & Obermeyer, Z. An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nat. Med* **27**, 136–140 (2021).
- Shah, N. H., Entwistle, D. & Pfeffer, M. A. Creation and Adoption of Large Language Models in Medicine. *JAMA* **330**, 866–869, (2023).
- Freedman, D. H. Hunting for New Drugs with AI. *Nature* **576**, S49–S53 (2019).
- Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, <https://doi.org/10.1038/s41586-021-03819-2> (2021).
- Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, eabj8754, <https://doi.org/10.1126/science.abj8754> (2021).
- Lyu, J. et al. Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).
- Mikulak-Klucznik, B. et al. Computational planning of the synthesis of complex natural products. *Nature* **588**, 83–88 (2020).
- Steiner, S. et al. Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, <https://doi.org/10.1126/science.aav2211> (2019).
- Davies, I. W. The digitization of organic synthesis. *Nature* **570**, 175–181 (2019).
- Campos, K. R. et al. The importance of synthetic chemistry in the pharmaceutical industry. *Science* **363**, <https://doi.org/10.1126/science.aat0805> (2019).
- Yeung, S., Downing, N. L., Fei-Fei, L. & Milstein, A. Bedside Computer Vision - Moving Artificial Intelligence from Driver Assistance to Patient Safety. *N. Engl. J. Med* **378**, 1271–1273 (2018).
- Rajkomar, A. et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* **1**, 18 (2018).
- Chung, H. U. et al. Binodal, wireless epidermal electronic systems with in-sensor analytics for neonatal intensive care. *Science* **363**, <https://doi.org/10.1126/science.aau0780> (2019).
- Zhao, M., Hoti, K., Wang, H., Raghu, A. & Katabi, D. Assessment of medication self-administration using artificial intelligence. *Nat. Med* **27**, 727–735 (2021).
- Singh, A. et al. Automatic detection of hand hygiene using computer vision technology. *J. Am. Med. Inf. Assoc.* **27**, 1316–1320 (2020).
- Qin, B. & Li, D. Identifying Facemask-Wearing Condition Using Image Super-Resolution with Classification Network to Prevent COVID-19. *Sensors (Basel)* **20**, <https://doi.org/10.3390/s20185236> (2020).
- Dias, R. D., Gupta, A. & Yule, S. J. Using Machine Learning to Assess Physician Competence: A Systematic Review. *Acad. Med* **94**, 427–439 (2019).
- Khalid, S., Goldenberg, M., Grantcharov, T., Taati, B. & Rudzicz, F. Evaluation of Deep Learning Models for Identifying Surgical Actions and Measuring Performance. *JAMA Netw. Open* **3**, e201664 (2020).

24. Jilani, T. et al. Short and Long term predictions of Hospital emergency department attendances. *Int J. Med Inf.* **129**, 167–174 (2019).
25. Kutafina, E., Bechtold, L., Kabino, K. & Jonas, S. M. Recursive neural networks in hospital bed occupancy forecasting. *BMC Med Inf. Decis. Mak.* **19**, 39 (2019).
26. Sim, I. Mobile Devices and Health. *N. Engl. J. Med.* **381**, 956–968 (2019).
27. Ferretti, L. et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, <https://doi.org/10.1126/science.abb6936> (2020).
28. Baumgart, D. C. Digital advantage in the COVID-19 response: perspective from Canada's largest integrated digitalized healthcare system. *NPJ Digit Med* **3**, 114 (2020).
29. Kaushal, A., Altman, R. & Langlotz, C. Geographic Distribution of US Cohorts Used to Train Deep Learning Algorithms. *JAMA* **324**, 1212–1213 (2020).
30. Rajewsky, N. et al. LifeTime and improving European healthcare through cell-based interceptive medicine. *Nature* **587**, 377–386 (2020).
31. Warnat-Herresthal, S. et al. Swarm Learning for decentralized and confidential clinical machine learning. *Nature* **594**, 265–270 (2021).
32. Chen, J. & Ran, X. Deep Learning With Edge Computing: A Review. *Proc. IEEE* **107**, 1655–1674 (2019).
33. Leeming, G., Ainsworth, J. & Clifton, D. A. Blockchain in health care: hype, trust, and digital health. *Lancet* **393**, 2476–2477 (2019).
34. European Commission. European Health Data Space (EHDS), https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en (2023).
35. Awad, E. et al. The Moral Machine experiment. *Nature* **563**, 59–64 (2018).
36. Data for Health Conference 2023, Berlin, Germany. (2023). <https://projektraeger.dlr.de/media/events/dfh23/index.html>.
37. Data for Health Workshop 2023, Boston, MA, USA. (2023). https://projektraeger.dlr.de/media/events/dfh23/boston_workshop.html.
38. German Federal Foreign Office. Joint statement following the 2023 meeting of the German-Canadian High Level Steering Group on Bilateral Cooperation (HLSG), <https://www.auswaertiges-amt.de/en/newsroom/news/-/2601174> (2023).
39. Sieck, C. J. et al. Digital inclusion as a social determinant of health. *NPJ Digit Med* **4**, 52 (2021).
40. Wiegand, T. et al. WHO and ITU establish benchmarking process for artificial intelligence in health. *Lancet* **394**, 9–11 (2019).
41. International Medical Device Regulators Forum (IMDRF). Strategic Plan 2021 – 2025, <http://www.imdrf.org/docs/imdrf/final/procedural/imdrf-proc-n39-strategic-plan-200925.pdf> (2023).
42. Hwang, T. J., Kesselheim, A. S. & Vokinger, K. N. Lifecycle Regulation of Artificial Intelligence- and Machine Learning-Based Software Devices in Medicine. *JAMA* **322**, 2285–2286 (2019).
43. Finlayson, S. G. et al. The Clinician and Dataset Shift in Artificial Intelligence. *N. Engl. J. Med.* **385**, 283–286 (2021).
44. Wachter, R. M. & Cassel, C. K. Sharing Health Care Data With Digital Giants: Overcoming Obstacles and Reaping Benefits While Protecting Patients. *JAMA* **323**, 507–508 (2020).
45. Haug, C. J. Turning the Tables - The New European General Data Protection Regulation. *N. Engl. J. Med.* **379**, 207–209 (2018).
46. The Group of Seven (G7), Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems, <https://ec.europa.eu/newsroom/dae/redirection/document/99643> (2023).
47. The Group of Seven (G7), Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems, <https://ec.europa.eu/newsroom/dae/redirection/document/99641> (2023).
48. O'Meara, S. The robot recruits in China's health-care system. *Nature* **582**, S49–S51 (2020).

ACKNOWLEDGEMENTS

D.C.B. is supported by research grants from the University of Alberta, the Canadian Institute for Advanced Research (CIFAR), the Canadian Institutes of Health Research (CIHR), the National Sciences and Engineering Council of Canada (NSERC), Mathematics of Information Technology and Complex Systems (Mitacs), the Alberta Machine Intelligence Institute (Amii), the University Hospital Foundation (UHF) and Alberta Innovates. This work is part of the International Collaborative Research and Training Experience (NSERC CREATE) "From Data to Decision (FD2D) - Digital Transformation and Artificial Intelligence from Data Value Chain to Human Value" (<https://fd2d.org>) led by D.C.B.

AUTHOR CONTRIBUTIONS

This is to confirm that D.C.B. is the only author of the manuscript.

COMPETING INTERESTS

The author declares no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-024-01005-y>.

Correspondence and requests for materials should be addressed to Daniel C. Baumgart.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024