



OPEN ACCESS

EDITED BY

Gokce Banu Laleci Erturkmen,
Software Research and Development
Consulting, Türkiye

REVIEWED BY

Bertrand De Meulder,
European Institute for Systems Biology and
Medicine (EISBM), France

*CORRESPONDENCE

Kaya Akyüz

✉ kaya.akyuez@bbmri-eric.eu

RECEIVED 11 November 2023

ACCEPTED 19 January 2024

PUBLISHED 31 January 2024

CITATION

Akyüz K, Cano Abadía M, Goisau M and
Mayrhofer MT (2024) Unlocking the potential
of big data and AI in medicine: insights from
biobanking.

Front. Med. 11:1336588.

doi: 10.3389/fmed.2024.1336588

COPYRIGHT

© 2024 Akyüz, Cano Abadía, Goisau and
Mayrhofer. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Unlocking the potential of big data and AI in medicine: insights from biobanking

Kaya Akyüz*, Mónica Cano Abadía, Melanie Goisau and
Michaela Th. Mayrhofer

Department of ELSI Services and Research, BBMRI-ERIC, Graz, Austria

Big data and artificial intelligence are key elements in the medical field as they are expected to improve accuracy and efficiency in diagnosis and treatment, particularly in identifying biomedically relevant patterns, facilitating progress towards individually tailored preventative and therapeutic interventions. These applications belong to current research practice that is data-intensive. While the combination of imaging, pathological, genomic, and clinical data is needed to train algorithms to realize the full potential of these technologies, biobanks often serve as crucial infrastructures for data-sharing and data flows. In this paper, we argue that the 'data turn' in the life sciences has increasingly re-structured major infrastructures, which often were created for biological samples and associated data, as predominantly data infrastructures. These have evolved and diversified over time in terms of tackling relevant issues such as harmonization and standardization, but also consent practices and risk assessment. In line with the datafication, an increased use of AI-based technologies marks the current developments at the forefront of the big data research in life science and medicine that engender new issues and concerns along with opportunities. At a time when secure health data environments, such as European Health Data Space, are in the making, we argue that such meta-infrastructures can benefit both from the experience and evolution of biobanking, but also the current state of affairs in AI in medicine, regarding good governance, the social aspects and practices, as well as critical thinking about data practices, which can contribute to trustworthiness of such meta-infrastructures.

KEYWORDS

biobanks, artificial intelligence, big data, European Health Data Space, infrastructures

1 Introduction

Life sciences knowledge production is increasingly structured by big data approaches, internationalization of research and closer coupling between research and applications, where biobanks comprise a major form of infrastructure in the current research ecosystems. For decades, biobanks have efficiently ensured access to biological samples and associated health data, which is being produced, collected and used in various ways, such as for medical research and public health databases as the two broad categories of population-based and clinical biobanks reflect (1). The historical development of the biobanks and their diversification over time contrast starkly with the current efforts for standardization, harmonization, integration, globalization and most significantly datafication. They have evolved from mere repositories to trusted infrastructures in sharing biomaterials and data (2), highlighting their crucial role in

data-intensive research. These efforts for facilitating the movement of data materialized into platforms, infrastructures and guiding principles to enable the exchange of data that is compliant with ethical, legal and societal considerations.

With artificial intelligence (AI), renewed discussions are taking place due to the idiosyncrasies of AI, the speed and consequences of the implementation of such technologies in biobanking and other domains (3, 4). Over the last decade, the development of national and transnational biobank networks or infrastructures have made such infrastructures instrumental to international research consortia (5–7). In addition, meta data infrastructures called health data spaces are developed that have the potential to significantly transform the life sciences, medicine and healthcare. Back in December 2020, the European Commission published the roadmap for the European Health Data Space (EHDS) initiative inviting public responses and presenting a first draft in May 2022. Currently discussed in the European Council and the European Parliament, the ambitious goal remains to complete the legislative process by the end of 2023 but no later than within the current Commission's mandate to ensure the implementation by 2025 (8). The EHDS will undoubtedly transform the health sector in Europe. It remains to be seen in which form it will be realized, especially as expectations are high across various stakeholder groups, such as patient advocacy groups, researchers from academia and industry as well as policy makers (9). At the same time, infrastructures such as biobanks have a wealth of experience regarding the collection and use of health data for research purposes in an ethically and legally compliant way (10). The perspective we present here builds on the observation that many biobanks are already going through a transformation in becoming bio(data)banks and are entangled in trials of various data practices that can inform both the debates around AI's use in life sciences and health research and emerging meta infrastructures considering developments, such as EU's upcoming Artificial Intelligence Act. Although there has been a provisional agreement as of December 9th, 2023, among negotiators from EU's Parliament and Council, the legal text will be implemented when the two institutions provide their approval and, if so, with its risk-based categorization and the accompanying requirements, the AI Act may have an impact on many aspects of AI's use in health research and applications, such as on data governance, explainability, requirements, practicing human-in-the-loop among others with potential effect also on the EHDS (11). In light of these recent developments, we argue that it is timely to look back at the practice of biobanking, especially the so-called data turn, and the current momentum in biobanking and medicine regarding AI and its implementation into research and technology, for insights on health data spaces and their development.

2 Data turn in life sciences: biobanks as data infrastructures

Biomedical research has become increasingly data-intensive and undergone a process of datafication (12). Central to this datafication are biobanks. As infrastructures, they can be characterized as vital entities in organizing practices, as embedded in other structures, social arrangements and technologies (13). In this capacity, biobanks support medical innovation, such as personalized medicine and genomic research, with scholars noting the molecularization and computerization sustaining both (14, 15).

The molecularization and data turn in the focus of biobank research in the last two decades deserves more attention. For instance, infrastructures have been created that gather genetic data from commercial and clinical sources, enabling population-based genetics research to be conducted (16). The outcome of such research, especially in genomics, raises hopes with a better understanding of the genetic bases of health conditions such as coronary artery disease, ideally based on diverse populations (17). However, the genomic data and infrastructures raise also concerns, especially regarding phenomena, such as sexual orientation, which received renewed attention in the search for a genetic basis (18) and also harbor emerging risks that are radically different than the previous ones due to intensive datafication, for instance, risks of genomic identifiability (19).

The existence of efforts towards standardization and interoperability in biobanking as reflected in the acronyms SPREC (20), BRISQ (21), MIABIS (22, 23) and others show the centrality of these notions for the data turn, but also harmonization regarding samples, technical infrastructures and practices. The relevant research contributes to developments such as specific algorithms for post-analytical use, which may bridge the differences between distinct types of blood samples originally stored for different uses (24, 25). Such developments are especially salient considering that biobanks are not independent of the broader infrastructures of medicine and healthcare. From disease categorization to defining and standardizing biomarkers at a time wearable devices, sensors and emerging forms of data are increasingly being embedded into entire ecosystems often in the digital (26), the existing samples and data with different conditions of collection, annotation, consent status and storage, as well as variations across institutions are still part of the picture. Biobanks are expanding with both typical samples and data (e.g., blood, BMI) and further kinds (e.g., epigenetic, microbiome, etc.) being integrated and standardized, expanding the data in both dimensions of volume and diversity.

In attempts towards datafication, practices around samples such as in pathology are also being transformed, exemplified by “digital pathology” where whole slide images that are once created may decrease the need to store samples or increase the findability by turning images into data collected (27). Scholars observe along a trend of consolidation emergence of virtual biobanks brings together resources from multiple biobanks (28, 29), though such cataloging examples also include efforts of broader research infrastructures, such as BBMRI-ERIC (30). Similarly, in the genomics world, efforts to standardize and make genomic data accessible such as summary statistics of genome-wide association studies is picking up pace (31, 32) as well as the development of trusted research environments despite critique (33) with specific tools, such as DataSHIELD (34).

3 AI in medicine and new beginnings for biobanking

Large amounts of data are needed to advance biomedical knowledge generation as well as big data analytics and new data-driven technologies in AI. While the history of AI in medicine goes back half a century with the initiation of computational tools and technical infrastructures as well as events devoted to the topic (35), it has gained pronounced attention and applicability in recent years in line with its intensive use in other domains. Medical AI is seen as a

promising innovation for uses such as screening, diagnosis, risk assessment, clinical decision-making, management planning, and precision medicine, with available tools ranging from chatbots to clinical decision support (36). The hope is that AI systems will reduce human bias and improve performance, as has been demonstrated in certain areas such as radiology (37), by improving accuracy in medical image analysis and easing the workload in screening (38), or for AI-driven polygenic risk scores (PRS) which may enable greater accuracy, performance and prediction (39). AI can also bring improvements when it comes to clinical measurements (40), interpretation of tests (41), decision making for intensive care unit admission (42), or embryo implantation (43), among others. However, it is important to note that AI is not a one-size-fits-all solution, and its benefits may not be realized in every application.

The development and implementation of medical AI involves numerous key challenges. First, AI is data hungry. Large amounts of data are needed to train AI and access to these data is challenging for technical, legal, and practical reasons, along with emerging issues regarding computational power and infrastructures and alternatives such as federated learning, which bring their own challenges and opportunities (44). One salient challenge in this respect relates to the tradeoff between data access and data privacy, the resolution of which necessitates bottom-up, democratic and engaging processes (3) in consideration of commitment for findable, accessible, interoperable and reusable data as often referred to with the acronym FAIR (45) and further FAIR principles (e.g., <https://www.go-fair.org/fair-principles/>). Second, despite the immense potential benefits, the risks revolve around perpetuation or even amplification of societal inequality and injustices due to potentially biased datasets as well as certain data practices (46). Third, practitioners require practical recommendations for applying AI (47). Furthermore, patients' preference for human agents or human supervision, possible strain between patients and treating physicians, especially in relation to privacy, data security and potential vulnerabilities related to AI tools need attention as do the implementation of guidelines and frameworks to ensure bioethical principles [e.g., (48)] are upheld and monitored (49). These call for engagement of multiple stakeholders in the resolution of ethical and legal issues, sharing similarities with biobanking, though at a different scale.

Biobanks, as key entities for providing access to large amounts of high-quality data, are central to the development of new data-based technologies such as AI. Similar to AI in medicine, the early developments in the use of AI in biobanking often focus on biobank participants' health conditions as reviewed elsewhere (50). These include developments such as, identifying and categorizing Alzheimer's disease patients (51), calculating risks scores for conditions such as age-related macular degeneration (52) or cardiovascular diseases (53), aiding in classification of disease subtypes (54) as well as providing predictions at individual level for COVID-19 (55, 56) or potential conditions due to therapeutic agents such as aromatase inhibitor-related arthralgia (57). However, biobanks are not merely support structures for healthcare or repositories for medical data. Biobanks have the potential to handle the data turn as they pursue data-driven practices in a standardized, industrialized manner (58). As research infrastructures, biobanks, may benefit from AI in the collection of biological samples and data, such as analysis of the scholarly literature for development of criteria for sampling, analysis, interpretation, data extraction, even engagements with

biobank participants, from consent process to research process; however, AI can also contribute to purely *managerial tasks* including storage space optimization or *upstream research processes*, such as suggesting samples and data for research proposals based on content and methods, as well as *downstream research evaluation*, assessing the "value" of samples and data based on the scholarly literature (59). AI's potential impact on biobanking may also include possible increases in the use of biobank samples and data, thus contributing to sustainability and speed of research as well as aiding biobanks in identification and recruitment of participants, training, annotation of samples and data, increasing interoperability, visibility, and access (60).

AI is central to the idea of "biobanks for the future" (61) though challenges in implementation of AI in biobanking range from difficulties aligning standards not only across data in the long run, but also samples, workflows, ethics management, legal and governance-related aspects, from transparency to informed consent (28) as well as justice, both epistemically and ethically (14). There are efforts such as workshops or collections of best practices to increase the "readiness" of these infrastructures for AI (60) with calls, checklists, tools and frameworks for ethical use of AI in medicine/biobanking (47, 62). New and alternative forms of governance are needed for a new form of biobanking that revolves around big data considering the increasing widening of the scope of data from social media to devices capturing bodily function, resulting in streams of data over *time* and analytical capacity over *space* (63). Biobanks' positioning at the in practice often gray intersection of healthcare and research can inform the discussions on health data spaces, in light of the recent developments.

4 Discussion

The ways in which risks are approached in biobanking and the normative arguments regarding how they should, such as future-proofing the governance of biobanks (64) and adaptive risk governance (65), suggest biobanking may be helpful in identifying key questions medical AI and health data spaces are facing from informed consent, representation in datasets, to risks associated with data protection and responsibility. While acknowledging the digital divide and its consequences, the increased ability of participants to follow and engage with biobanking and healthcare infrastructures are leading to reconfigurations of "traditional boundaries between the public domain (healthcare systems, medical research, and clinical practice) and the private one (patients and citizens)" which necessitate new approaches to fostering trust (63). Health data spaces bring such observations to a new level.

Trust and trustworthiness have become keywords that are often attached to how AI should be, with limited discussion of what this entails. Despite the burgeoning literature on ethics of AI in medicine, three areas relevant for trust are problematic (46): limited analytical accuracy and conceptual slippages, inadequate analysis of the contexts in which medical AI tools are embedded, and scarcity of interdisciplinary approaches. Considering trust central to societal functioning as "a fundamental principle for interpersonal interactions" (66), it cannot be considered unidirectional. Rather, it needs to be understood as a complex, situated, context-dependent, and relational concept that involves several trustor/trustee relationships, such as trust in persons (e.g., scientists who trust each other, patients who trust scientists and clinicians), technology, and institutions (67, 68). Trust or

more precisely trusting relationships are fragile and require continuous work, which means that they need to be actively established and sustained. In this sense, we see three main considerations from biobanking – a domain that should be built on trust – that can contribute to better medical AI and health data spaces.

Regulations may provide guidance, but good governance is an active process that comprises more than following regulations. Efforts towards regulating and guiding AI have been abundant with ‘AI Ethics’ becoming a buzzword (69, 70) along with the legal frameworks such as the proposed Artificial Intelligence Act of the EU (11). Considering international standards, overseeing organizations, national legislations, as well as practices, from engaging participants to consents, biobanks have accumulated over decades experiences related to intensified transnational data sharing, international collaborations, including public-private partnerships, access to and reuse of data, and efforts to harmonize data, ethical/legal standards and societal aspects. Hence, biobanking incorporates knowledge of the “ethics work” that is an integral part of data flows (71) and necessitates thinking critically about potential issues that go beyond individual institutions, such as identifiability risks in a datafied world both in regards to genomic (19) and medical imaging data (72). Thus, necessary good governance involves more than procedure-following.

Infrastructures are not merely technical, i.e., buildings, data repositories, but also social – involving practices. A recent study (73) with biobank professionals and experts indicates that expectations towards biobanks in view of data processing are going beyond their status as repositories. They see biobanks in a more active role when it comes to providing information and communicating and engaging with biobanks participants and point to the need to improve consent procedures and the role of biobanks in sharing samples and data with industry partners and different countries. Considering that participants are the origin of the data, as key stakeholders they should be involved in the development and governance, just as staff in biobanks should be included (74). Decades of biobanking show that *the concerns of citizens cannot be ignored*. In the case of AI in health, these not only relate to the general concerns regarding AI. On the contrary, as suggested by the PRS and AI, ethical, legal and societal issues necessitate a layered understanding due to increasing complexity bringing new relevance to concepts such as explainability and interpretability, both for the users and the broader society (39). Considering the drivers of AI in medicine, such as identification and management of potential patients that can be “high-risk” but also “high-cost” (75), the developments may not benefit individuals who may otherwise develop conditions that are harder to treat or identify and manage emerging outbreaks in real-time, and such AI tools may cause further burdens on the individuals. These necessitate societal debates and empowering citizens, including involving potential non-users, as part of bringing infrastructures to life (76).

Not only are data not always perfect due to inherent finite categorization of potentially infinite diversity, but their capacity to represent should always be continuously problematized. Against the biobanking professionals’ concerns, the tendency to see biobanks as data repositories and medicine as increasingly digital (27, 63) can result in a false sense of security in the imaginary of increasing data interoperability and connectedness at the peril of ignoring what D’Ignazio and Klein (77) rightly note the existence of “problems that cannot be represented—or addressed—by data alone” (p. 10). Risks accompany the opportunities in a datafied world. The existence of data

should not automatically lead to testing of any potential association and scholars have been trying to identify ways of coping with such issues of reproducibility, e.g., for PRS (78, 79). In this regard, the “curse of dimensionality” in biobanking due to multitude of secondary data even in cases of low sample sizes, can also be seen as an opportunity to think outside of the box to overcome issues even in smaller sample size situations (80). Furthermore, AI may also exacerbate the existing big data issues that are yet to be resolved. While the uses may relate to privacy with unintended access to data from patient implants, sensors and other devices that collect and transfer multiple forms of data, they may also lead to spurious correlations and false positives, tacit assumptions regarding individual behavior based on limited data, sampling issues due to replacement of traditional ways of data collection as well as resulting in injustices due to resource mismanagement and allocation, especially in case of public health issues (81). With health data spaces, these issues will likely need more attention.

Projectified ways of health infrastructuring often restrict the outcome in many ways, through visions and expectations for whom and which purposes the infrastructure is to be developed even in cases where the aim is to involve stakeholders in co-creation processes (76). In this paper we have shown the wealth of knowledge generated through the use of AI in medicine and the evolution of biobanking. We argue, when taken into account, these can positively impact the future European Health Data Space, but also similar establishments, giving power to the citizen, strengthening governance, breaking down potential silos and contributing to trustworthiness of such meta-infrastructures.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

KA: Conceptualization, Writing – original draft, Writing – review & editing. MA: Conceptualization, Writing – original draft, Writing – review & editing. MG: Conceptualization, Writing – original draft, Writing – review & editing. MM: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This publication was funded by BBMRI-ERIC in the context of the activities of BBMRI-ERIC’s ELSI Services and Research Unit.

Acknowledgments

Where authors are identified as personnel of the Biobanking and BioMolecular resources Research Infrastructure (BBMRI-ERIC), the authors alone are responsible for the views

expressed in this article and they do not necessarily represent the decisions, policy, or views of BBMRI-ERIC.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Parodi B. Biobanks: a definition In: D Mascalon, editor. *Ethics, law and governance of biobanking: National, European and international approaches*. Dordrecht: Springer (2015).
- De Souza YG, Greenspan JS. Biobanking past, present and future: responsibilities and benefits. *AIDS*. (2013) 27:303–12. doi: 10.1097/QAD.0b013e32835c1244
- Bak M, Madai VI, Fritzsche M-C, Mayrhofer MT, McLennan S. You Can't have AI both ways: balancing health data privacy and access fairly. *Front Genet*. (2022) 13:929453. doi: 10.3389/fgene.2022.929453
- Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: mapping the debate. *Big Data Soc*. (2016) 3:205395171667967. doi: 10.1177/2053951716679679
- Saunders G, Baudis M, Becker R, Beltran S, Bérout C, Birney E, et al. Leveraging European infrastructures to access 1 million human genomes by 2022. *Nat Rev Genet*. (2019) 20:693–701. doi: 10.1038/s41576-019-0156-9
- Mate S, Kampf M, Rödle W, Kraus S, Proynova R, Silander K, et al. Pan-European data harmonization for biobanks in ADOPT BBMRI-ERIC. *Appl Clin Inform*. (2019) 10:679–92. doi: 10.1055/s-0039-1695793
- Mayrhofer MT, Prainsack B. Being a member of the club: the transnational (self-) governance of networks of biobanks. *IJRAM*. (2009) 12:64–81. doi: 10.1504/IJRAM.2009.024130
- European Commission (2023). Directorate-general for health and food safety. European Health Data Space: European Commission. Available at: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en
- Marelli L, Stevens M, Sharon T, Van Hoyweghen I, Boeckhout M, Colussi I, et al. The European health data space: too big to succeed? *Health Policy*. (2023) 135:104861. doi: 10.1016/j.healthpol.2023.104861
- BBMRI-ERIC (2023). Public statement: Recommendations for the realisation of the EHDS (European health data spaces) for biobanking from the viewpoint of patient advocates and patient representatives: BBMRI-ERIC. Available at: <https://www.bbMRI-eric.eu/wp-content/uploads/BBMRI-ERIC-EHDS-Statement-310823.pdf>
- Council of the European Union. Artificial intelligence act: Council and parliament strike a deal on the first rules for AI in the world 2023. Available at: <https://www.council.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/>
- Ruckenstein M, Schüll ND. The Datafication of health. *Annu Rev Anthropol*. (2017) 46:261–78. doi: 10.1146/annurev-anthro-102116-041244
- Star SL, Ruhleder K. Steps toward an ecology of infrastructure: design and access for large information spaces. *Inf Syst Res*. (1996) 7:111–34. doi: 10.1287/isre.7.1.111
- Brault N, Aucouturier E. Ethical horizons of biobank-based artificial intelligence in biomedical research In: A Saxena and N Brault, editors. *Artificial intelligence and computational dynamics for biomedical research*. Berlin: De Gruyter (2022).
- Lemoine M. Neither from words, nor from visions: understanding p-medicine from innovative treatments. *Lato Sensu*. (2018) 4:12–23. doi: 10.20416/lrsrps.v4i2.793
- Hoeyer K, Bauer S, Pickersgill M. Datafication and accountability in public health: introduction to a special issue. *Soc Stud Sci*. (2019) 49:459–75. doi: 10.1177/0306312719860202
- Tcheandjieu C, Zhu X, Hilliard AT, Clarke SL, Napolioni V, Ma S, et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nat Med*. (2022) 28:1679–92. doi: 10.1038/s41591-022-01891-3
- Goisauf M, Akyüz K, Martin GM. Moving back to the future of big data-driven research: reflecting on the social in genomics. *Humanit. Soc. Sci*. (2020) 7:55. doi: 10.1057/s41599-020-00544-5
- Akyüz K, Goisauf M, Chassang G, Kozera L, Mežinska S, Tzortzatou-Nanopoulou O, et al. Post-identifiability in changing sociotechnological genomic data environments. *BioSocieties*. (2023):1–28. doi: 10.1057/s41292-023-00299-7 [Online ahead of print].
- Lehmann S, Guadagni F, Moore H, Ashton G, Barnes M, Benson E, et al. Standard Preanalytical coding for biospecimens: review and implementation of the sample PREanalytical code (SPREC). *Biopreserv Biobank*. (2012) 10:366–74. doi: 10.1089/bio.2012.0012

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Moore HM, Kelly A, Jewell SD, McShane LM, Clark DP, Greenspan R, et al. Biospecimen reporting for improved study quality. *Biopreserv Biobank*. (2011) 9:57–70. doi: 10.1089/bio.2010.0036
- Eklund N, Andrianarisoa NH, van Enckevort E, Anton G, Debucquoy A, Müller H, et al. Extending the minimum information about Biobank data sharing terminology to describe samples, sample donors, and events. *Biopreserv Biobank*. (2020) 18:155–64. doi: 10.1089/bio.2019.0129
- Merino-Martinez R, Norlin L, van Enckevort D, Anton G, Schuffenhauer S, Silander K, et al. Toward global biobank integration by implementation of the minimum information about Biobank data sharing (MIABIS 2.0 Core). *Biopreserv Biobank*. (2016) 14:298–306. doi: 10.1089/bio.2015.0070
- Zhuang Y-J, Mangwiro Y, Wake M, Saffery R, Greaves RF. Multi-omics analysis from archival neonatal dried blood spots: limitations and opportunities. *Clin Chem Lab Med*. (2022) 60:1318–41. doi: 10.1515/cclm-2022-0311
- Kaushal A, Zhang H, Karmaus WJJ, Ray M, Torres MA, Smith AK, et al. Comparison of different cell type correction methods for genome-scale epigenetics studies. *BMC Bioinformatics*. (2017) 18:216. doi: 10.1186/s12859-017-1611-2
- Califf RM. Biomarker definitions and their applications. *Exp Biol Med*. (2018) 243:213–21. doi: 10.1177/1535370217750088
- Bonizzi G, Zattoni L, Fusco N. Biobanking in the digital pathology era. *Oncol Res*. (2021) 29:229–33. doi: 10.32604/or.2022.024892
- Kozlakidis Z. Biobanks and biobank-based artificial intelligence (AI) implementation through an international Lens In: A Holzinger, R Goebel, M Mengel and H Müller, editors. *Artificial intelligence and machine learning for digital pathology: State-of-the-art and future challenges*. Cham: Springer International Publishing (2020)
- Vande Look K, Van der Stock E, Debucquoy A, Emmerechts K, Van Damme N, Marbaix E. The Belgian virtual Tumorbank: a tool for translational Cancer research. *Front Med*. (2019) 6:6. doi: 10.3389/fmed.2019.00120
- Holub P, Swertz M, Reihls R, van Enckevort D, Müller H, Litton J-E. BBMRI-ERIC directory: 515 biobanks with over 60 million biological samples. *Biopreserv Biobank*. (2016) 14:559–62. doi: 10.1089/bio.2016.0088
- Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Mangone C, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. (2018) 47:D1005–12. doi: 10.1093/nar/gky1120
- Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. *Nucleic Acids Res*. (2023) 51:D977–85. doi: 10.1093/nar/gkac1010
- Mackenzie G, Richard M, Paige F, Mark S. Trust and the Goldacre review: why trusted research environments are not about trust. *J Med Ethics*. (2022) 49:670–3. doi: 10.1136/jme-2022-108435
- Marcon Y, Bishop T, Avraam D, Escriba-Montagut X, Ryser-Welch P, Wheeler S, et al. Orchestrating privacy-protected big data analyses of data from different resources with R and DataSHIELD. *PLoS Comput Biol*. (2021) 17:e1008880. doi: 10.1371/journal.pcbi.1008880
- Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. *Gastrointest Endosc*. (2020) 92:807–12. doi: 10.1016/j.gie.2020.06.040
- Chen M, Decary M. Artificial intelligence in healthcare: an essential guide for health leaders. *Health Manage Forum*. (2019) 33:10–8. doi: 10.1177/0840470419873123
- McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature*. (2020) 577:89–94. doi: 10.1038/s41586-019-1799-6
- Mudgal KS, Das N. The ethical adoption of artificial intelligence in radiology. *BJR Open*. (2019) 2:20190020. doi: 10.1259/bjro.20190020
- Fritzsche M-C, Akyüz K, Cano Abadía M, McLennan S, Marttinen P, Mayrhofer MT, et al. Ethical layering in AI-driven polygenic risk scores—new complexities, new challenges. *Front Genet*. (2023) 14:1098439. doi: 10.3389/fgene.2023.1098439
- Niel O, Bastard P, Boussard C, Hogan J, Kwon T, Deschênes G. Artificial intelligence outperforms experienced nephrologists to assess dry weight in pediatric patients on chronic hemodialysis. *Pediatr Nephrol*. (2018) 33:1799–803. doi: 10.1007/s00467-018-4015-2

41. Topalovic M, Das N, Burch P-R, Daenen M, Derom E, Haenebalcke C, et al. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. *Eur Respir J*. (2019) 53:1801660. doi: 10.1183/13993003.01660-2018
42. Carrano FM, Wang B, Sherman SE, Makarov DV, Berman RS, Newman E, et al. Artificial intelligence outperforms clinical judgment in triage for postoperative ICU care: prospective preliminary results. *J Am Coll Surg*. (2019) 229:S141–2. doi: 10.1016/j.jamcollsurg.2019.08.312
43. Hariton E, Dimitriadis I, Kanakasabapathy MK, Thirumalaraju P, Gupta R, Pooniwalla R, et al. A deep learning framework outperforms embryologists in selecting day 5 euploid blastocysts with the highest implantation potential. *Fertil Steril*. (2019) 112:e77–8. doi: 10.1016/j.fertnstert.2019.07.324
44. Rahman A, Hossain MS, Muhammad G, Kundu D, Debnath T, Rahman M, et al. Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues. *Clust Comput*. (2023) 26:2271–311. doi: 10.1007/s10586-022-03658-4
45. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*. (2016) 3:160018. doi: 10.1038/sdata.2016.18
46. Goisau M, Cano AM. Ethics of AI in radiology: a review of ethical and societal implications. *Front Big Data*. (2022) 5:850383. doi: 10.3389/fdata.2022.850383
47. Muller H, Mayrhofer M, Veen EV, Holzinger A. The ten commandments of ethical medical AI. *Computer*. (2021) 54:119–23. doi: 10.1109/MC.2021.3074263
48. Beauchamp TL, Childress JF. Principles of biomedical ethics: marking its fortieth anniversary. *AJOB*. (2019) 19:9–12. doi: 10.1080/15265161.2019.1665402
49. Prakash S, Balaji JN, Joshi A, Surapaneni KM. Ethical conundrums in the application of artificial intelligence (AI) in healthcare — a scoping review of reviews. *J Pers Med*. (2022) 12:1914. doi: 10.3390/jpm12111914
50. Battineni G, Hossain MA, Chintalapudi N, Amenta F. A survey on the role of artificial intelligence in biobanking studies: a systematic review. *Diagnostics*. (2022) 12:1179. doi: 10.3390/diagnostics12051179
51. Tian J, Smith G, Guo H, Liu B, Pan Z, Wang Z, et al. Modular machine learning for Alzheimer's disease classification from retinal vasculature. *Sci Rep*. (2021) 11:238. doi: 10.1038/s41598-020-80312-2
52. Yan Q, Jiang Y, Huang H, Swaroop A, Chew EY, Weeks DE, et al. Genome-wide association studies-based machine learning for prediction of age-related macular degeneration risk. *Translational vision. Sci Technol*. (2021) 10:29. doi: 10.1167/tvst.10.2.29
53. Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF, van der Schaar M. Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK biobank participants. *PLoS One*. (2019) 14:e0213653. doi: 10.1371/journal.pone.0213653
54. Schulz M-A, Chapman-Rounds M, Verma M, Bzdok D, Georgatzis K. Inferring disease subtypes from clusters in explanation space. *Sci Rep*. (2020) 10:12900. doi: 10.1038/s41598-020-68858-7
55. Dabbah MA, Reed AB, Booth ATC, Yassaee A, Despotovic A, Klasmer B, et al. Machine learning approach to dynamic risk modeling of mortality in COVID-19: a UK biobank study. *Sci Rep*. (2021) 11:16936. doi: 10.1038/s41598-021-95136-x
56. Jimenez-Solem E, Petersen TS, Hansen C, Hansen C, Lioma C, Igel C, et al. Developing and validating COVID-19 adverse outcome risk prediction models from a bi-national European cohort of 5594 patients. *Sci Rep*. (2021) 11:3246. doi: 10.1038/s41598-021-81844-x
57. Reinbolt RE, Sonis S, Timmers CD, Fernández-Martínez JL, Cernea A, de Andrés-Galiana EJ, et al. Genomic risk prediction of aromatase inhibitor-related arthralgia in patients with breast cancer using a novel machine-learning algorithm. *Cancer Med*. (2018) 7:240–53. doi: 10.1002/cam4.1256
58. Mayrhofer MT. About the new significance and the contingent meaning of biological material and data in biobanks. *Hist Phil Life Sci*. (2013) 35:449–67.
59. Lee J. Artificial intelligence in the future biobanking: current issues in the biobank and future possibilities of artificial intelligence. *Biomed J Sci Tech Res*. (2018) 7:5937–9. doi: 10.26717/BJSTR.2018.07.001511
60. Grossman GH, Henderson MK. Readiness for artificial intelligence in biobanking. *Biopreserv Biobank*. (2023) 21:119–20. doi: 10.1089/bio.2023.29121.editorial
61. Garcia DL. ISBER President's message: ISBER's 20th anniversary—celebrating the journey. *Biopreserv Biobank*. (2019) 17:375–6. doi: 10.1089/bio.2019.29056.dlg
62. Kargl M, Plass M, Müller H. A literature review on ethics for AI in biomedical research and biobanking. *Yearb Med Inform*. (2022) 31:152–60. doi: 10.1055/s-0042-1742516
63. Tozzo P, Delicati A, Marcante B, Caenazzo L. Digital biobanking and big data as a new research tool: a position paper. *Healthcare*. (2023) 11:1825. doi: 10.3390/healthcare11131825
64. Gille F, Vayena E, Blasimme A. Future-proofing biobanks' governance. *Eur J Hum Genet*. (2020) 28:989–96. doi: 10.1038/s41431-020-0646-4
65. Akyüz K, Chassang G, Goisau M, Kozera Ł, Mezinska S, Tzortzidou O, et al. Biobanking and risk assessment: a comprehensive typology of risks for an adaptive risk governance. *Life Sci Soc Policy*. (2021) 17:1–28. doi: 10.1186/s40504-021-00117-7
66. Ryan M, Stahl BC. Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *J Inf Commun Ethics Soc*. (2020) 19:61–86. doi: 10.1108/JICES-12-2019-0138
67. Bijker EM, Sauerwein RW, Bijker WE. Controlled human malaria infection trials: how tandems of trust and control construct scientific knowledge. *Soc Stud Sci*. (2016) 46:56–86. doi: 10.1177/0306312715619784
68. Wyatt S, Harris A, Adams S, Kelly SE. Illness online: self-reported data and questions of Trust in Medical and Social Research. *Theory Cult Soc*. (2013) 30:131–50. doi: 10.1177/0263276413485900
69. Floridi L, Cows J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Mind Mach*. (2018) 28:689–707. doi: 10.1007/s11023-018-9482-5
70. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell*. (2019) 1:389–99. doi: 10.1038/s42256-019-0088-2
71. Hoeyer K, Tupasela A, Rasmussen MB. Ethics policies and ethics work in cross-national genetic research and data sharing. *Sci Technol Hum Values*. (2017) 42:381–404. doi: 10.1177/0162243916674321
72. Lotan E, Tschider C, Sodickson DK, Caplan AL, Bruno M, Zhang B, et al. Medical imaging and privacy in the era of artificial intelligence: myth, fallacy, and the future. *J Am Coll Radiol*. (2020) 17:1159–62. doi: 10.1016/j.jacr.2020.04.007
73. Goisau M, Martin G, Bentzen HB, Budin-Ljøsne I, Ursin L, Durnová A, et al. Data in question: a survey of European biobank professionals on ethical, legal and societal challenges of biobank research. *PLoS One*. (2019) 14:e0221496. doi: 10.1371/journal.pone.0221496
74. Akyüz K, Goisau M, Martin GM, Mayrhofer MT, Antoniou S, Charalambidou G, et al. Risk mapping for better governance in biobanking: The case of biobank.Cy. In press.
75. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff*. (2014) 33:1123–31. doi: 10.1377/hlthaff.2014.0041
76. Felt U, Öchsner S, Rae R, Osipova E. Doing co-creation: power and critique in the development of a European health data infrastructure. *J Responsible Innov*. (2023) 10:2235931. doi: 10.1080/23299460.2023.2235931
77. D'Ignazio C, Klein LF. *Data Feminism*. Cambridge: MIT Press (2020).
78. Lambert SA, Gil L, Jupp S, Ritchie SC, Xu Y, Buniello A, et al. The polygenic score catalog as an open database for reproducibility and systematic evaluation. *Nat Genet*. (2021) 53:420–5. doi: 10.1038/s41588-021-00783-5
79. Wand H, Lambert SA, Tamburro C, Iacocca MA, O'Sullivan JW, Sillari C, et al. Improving reporting standards for polygenic scores in risk prediction studies. *Nature*. (2021) 591:211–9. doi: 10.1038/s41586-021-03243-6
80. Narita A, Ueki M, Tamiya G. Artificial intelligence powered statistical genetics in biobanks. *J Hum Genet*. (2021) 66:61–5. doi: 10.1038/s10038-020-0822-y
81. Strang KD, Sun Z. Hidden big data analytics issues in the healthcare industry. *Health Informatics J*. (2019) 26:981–98. doi: 10.1177/1460458219854603