Artificial Intelligence In Medicine 167 (2025) 103192

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Toward responsible artificial intelligence in medicine: Reflections from the Australian epilepsy project

Mangor Pedersen [a,b,*], Heath R. Pardoe [b], Anton de Weger [b], Donna Hutchison [b], David F. Abbott [b,e], Karin Verspoor [c,d], Graeme D. Jackson [b,e,f], for the Australian Epilepsy Project Investigators

[a] Department of Psychology and Neuroscience, Auckland University of Technology (AUT), Auckland, New Zealand
[b] The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Australia
[c] School of Computing Technologies, RMIT University, Melbourne, Australia
[d] School of Computing and Information Systems, The University of Melbourne, Australia
[e] Department of Medicine, Austin Health, The University of Melbourne, Australia
[f] Department of Neurology, Austin Health, Melbourne, Australia

ARTICLE INFO

Keywords:
AI
Ethics
Artificial Intelligence
Epilepsy
Australian Epilepsy Project
AEP

ABSTRACT

Artificial intelligence (AI) is a multidisciplinary scientific field that uses machines to solve real-world problems and predict outcomes. Despite the current enthusiasm about AI's potential as a clinical support tool, there is also a growing awareness and concern about the potentially harmful effects of AI. Because AI will likely impact expert-based decision-making in medicine, it is critical to consider the issues that AI raises in medical research. This paper outlines the AI guidelines of the Australian Epilepsy Project. This large-scale platform aims to democratise specialist care in epilepsy and use AI for clinical decision support based on prospective multimodal datasets (MRI, genetic, clinical, and cognitive data) from thousands of people with epilepsy. As AI develops rapidly, we focus on key areas of medical AI identified in the literature, including *Trust, Responsibility* and *Safety*. We believe AI is changing medicine, and we believe it is imperative to advance and update our AI guidelines adaptably while preparing for an era of augmented-intelligence-based medicine.

1. Background

Artificial intelligence (AI) uses computing storage, processing, and knowledge in combination with interactions with environments to understand human intelligence [1]. AI is an exciting, rapidly evolving, and multidisciplinary field of science focused on problem-solving by machines. AI will most likely substantially impact expert-based decision-making in medicine and increasingly realise the challenge of deep phenotyping, which means understanding each individual's uniqueness through detailed characterization based on multimodal data [2]. To process complex data meaningfully, clinicians and AI scientists will benefit from working together to build and validate new models and demonstrate their utility to improve health outcomes [3].

Amongst the contemporary excitement around AI's potential, as well as its commercial success (the expected worth of the medical AI market is projected to be US$41.7 billion before 2027[1]) and research success (see the recent rise in medical AI publications in Fig. 1), there is an increasing awareness and apprehension about the potentially harmful effects of AI. As new developments in AI are continuously emerging, there are concerns about the potential detrimental impacts of AI, mainly how AI solutions are developed, shared, monitored, and regulated in medicine [4–16]. Given that AI will likely change medical practice, we must consider the possible implications and risks when developing, implementing, and governing AI algorithms in clinical settings.

2. What is the Australian Epilepsy Project?

Epilepsy is more than just seizures. Reduced educational and occupational attainment, decreased independence, lower quality of life, stigma, a higher risk of injury, cognitive deficiencies, mental illness, and

---

* Corresponding author at: 90 Akoranga Drive, Northcote, Auckland 0627, New Zealand.
E-mail address: mangor.pedersen@aut.ac.nz (M. Pedersen).

[1] https://www.reportlinker.com/p06325450/Global-AI-in-Healthcare-Market-Analysis-By-Component-By-Algorithm-By-Application-By-End-User-By-Region-Size-Forecast-with-Impact-Analysis-of-COVID-19-and-Forecast-up-to.html?utm_source=GNW

https://doi.org/10.1016/j.artmed.2025.103192
Received 27 April 2023; Received in revised form 19 March 2025; Accepted 3 June 2025

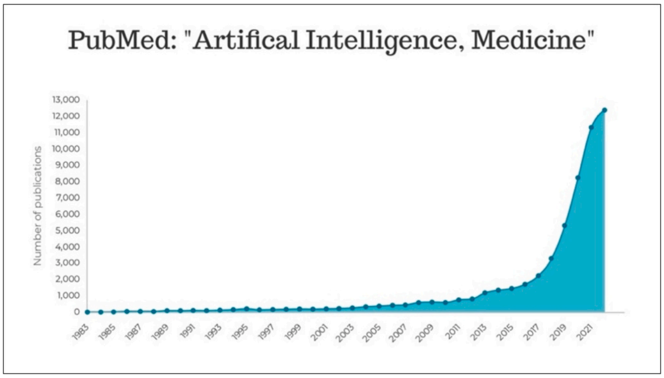

Fig. 1. In this figure, we display the number of publications in PubMed (https://pubmed.ncbi.nlm.nih.gov/) for the keywords: Artificial Intelligence, Medicine. It shows a rapid increase in medical publications referring to AI from 2017, at a time of significant scaling of AI models and the computational feasibility of deep learning models. Note that this is a rough approach to estimating the number of studies, not including papers relating to AI (e.g., using terminologies such as machine learning and natural language processing).

suicide are all connected with it [17–22]. *The Australian Epilepsy Project* [23–26] aims to give epilepsy specialists the information they need to make a difference in people's lives, facilitate access to currently limited care, and answer questions that are presently unanswerable. Will this person continue to suffer seizures? Which drug works the best? Can brain surgery be a successful solution? These are fundamental concerns that current medical practice frequently cannot answer.

The Australian Epilepsy Project has several aims that AI can address by integrating multimodal data, including advanced brain imaging, cognitive, and genetics data (see Fig. 2). The clinical purposes of the Australian Epilepsy Project include predicting whether a person who has recently suffered their first seizure is likely to experience a second seizure (epilepsy is diagnosed after more than one seizure). After people have been diagnosed with epilepsy, more critical clinical decisions exist. Typically, the first decision for the treating physician following diagnosis is the most appropriate antiseizure medication. In this context, AI could help decide which antiseizure medication is most useful for each patient by providing an individual-specific treatment plan. There is also a subset of patients where seizures arise from a circumscribed part of the brain, and in one third of people with epilepsy medication will not effectively control their seizures [27]. In these people, AI and advanced MRI images can be used to identify subtle epileptogenic lesions not otherwise seen in hospital-based brain imaging.

The Australian Epilepsy Project is a *prospective* and *clinically-led* study with a target sample size of *thousands of people* living with epilepsy from *all over Australia*. Participants in the Australian Epilepsy Project will also be tracked two years after study inclusion to provide reliable and clinically valid outcome data for AI classification, such as medical and surgical outcomes. These outcomes (i.e., labels for AI models) include whether people develop epilepsy after the first seizure, surgery outcome, control of seizures on medication as well as cognitive/psychiatric status. In addition to using AI to create predictions directly related to participant outcomes, the Australian Epilepsy Project intends to use AI in many aspects of the team's workflow and planning.

3. Why we developed a responsible AI framework in the Australian Epilepsy Project

We have developed a responsible AI framework in the Australian Epilepsy Project to ensure that AI implementation is both interpretable and clinically applicable. Over the last few years there have been extensive discussions around ethics and responsible AI. Our contribution to this discussion is to share the feasible implementation of responsible AI in the context of our medical research, aiming to improve diagnosis and treatment of epilepsy. The approach to responsible AI in the Australian Epilepsy project emphasises the critical importance of fostering trust in AI technologies, using work by Jobin et al. [28] and Sand et al. [15] as foundational references. Jobin et al. introduce the main principles for AI, influencing the initial AI principles of transparency, responsibility, non-maleficence, justice, fairness, and sustainability within the project. Sand et al. highlight physicians' responsibility to understand AI models used for clinical decision support. These guidelines empower human decision-making when developing, using, and interpreting AI models and highlight that domain knowledge is imperative to successfully implementing AI in medicine. This argument leads us to augmented intelligence [29,30], an emerging field in AI where people and computers symbiotically work together for the best possible and explainable AI outcomes. Augmented intelligence leverages the best aspects of human knowledge and computer algorithms. By incorporating the work of Jobin et al. and Sand et al., this article highlights the focus on AI trust, responsibility and safety as core focuses of AI in the Australian Epilepsy Project. This includes the responsibility of those who create and implement AI algorithms, as well as policymakers and regulators [31]. It also implies the design of human-in-the-loop uses of AI in medicine, where AI is deployed to support clinician decision-making rather than *replace* or *emulate* it [32].

The Australian population is ethnically diverse, which will likely benefit the generalisability of AI models derived from the Australian Epilepsy Project. The project will also collect demographic, gender, race, and other data often associated with biases in large datasets. It will use this information to reduce potential acquisition bias when training and using AI models. A priority for the Australian Epilepsy Project is to democratise clinical access for people with epilepsy. This means bringing people to research centers with novel technologies and capabilities, such as the ability to acquire advanced imaging including high-resolution neuroanatomy, functional MRI and diffusion imaging scans to image the brain's white matter pathways. Other data collection can now be done in people's homes, including saliva samples for genetic tests which can be sent by mail, and tele-neuropsychology, delivered online [26].

In the next three sub-sections we are going to cover the main aspects of conducting AI in the Australian Epilepsy Project, namely *trust, responsibility and safety* as we are moving to AI augmented-intelligence-based medicine.

4. Trust in AI

One of the key issues for AI is that scientists and clinicians must trust the AI model and its outputs. AI-based clinical decisions should be validated, explainable, and reproducible [33–35]. Understanding how AI systems arrive at their decisions or recommendations likely requires a shift into a multidisciplinary area of inquiry. Many AI models, including deep learning, are often conceived as a 'black box', and their outputs or results can be challenging to interpret [36]. For some applications, such as text search, the solution can be straightforward: ensure the AI system provides evidence for its output in the form of citations to original sources, to enable the user to check the response. In the Australian Epilepsy Project, the clinical trials team and clinicians are trialling application of a large language model to search and answer questions from the often lengthy medical records of participants. In our implementation, to facilitate efficient human validation, the AI answer to a posed question is accompanied alongside by highlighted passages of text from the original medical record that were the most relevant in determining the AI generated response.

For other applications, such as image processing, appropriate trust can be more complex. In epilepsy, identification of epileptogenic lesions increases the likelihood of obtaining a cure for many people with focal epilepsy via brain surgery [37]. Contemporary AI solutions such as deep learning have been explicitly developed for image-based applications and are well suited for detecting neuroanatomical abnormalities

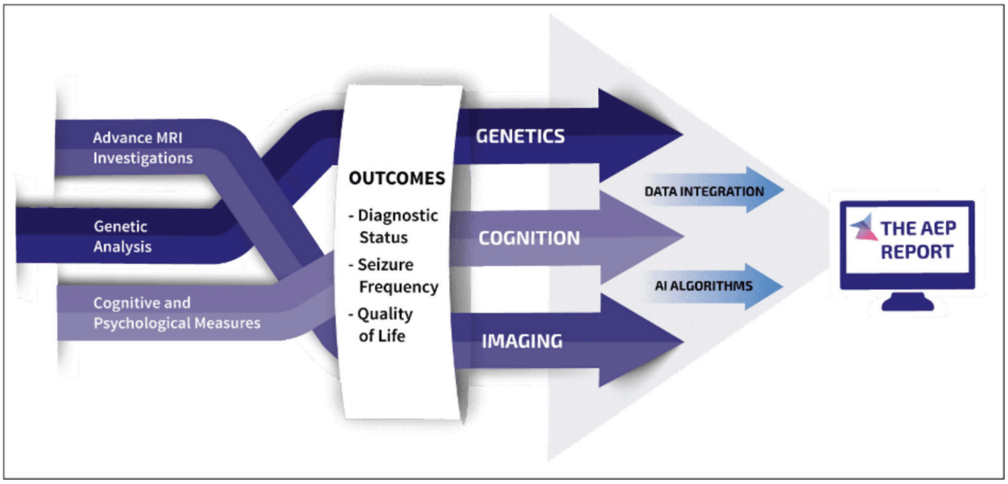

Fig. 2. In this figure, we outline the structure of the Australian Epilepsy project, including the data collection and long-term outcomes integral to using AI to improve the lives of people with epilepsy.

[38,39]. An example of a transparent and explainable lesion detection framework is proposed by Spitzer et al. [40], namely the Multi-centre Epilepsy Lesion Detection (MELD) framework. The MELD framework relies on detailed structural information about the brain to obtain a probability of a seizure focus based on structural brain imaging modalities commonly acquired in hospitals and research institutions. In addition to lesion probability, MELD provides interpretable outputs regarding the brain's cortical thickness, curvature, and white and grey matter imaging contrast explaining the model's output. Spitzer et al. [40] validated this approach with 1015 participants (epilepsy and controls) using a split-half training and testing paradigm. Patients with a visible brain lesion –focal cortical dysplasia– were detected with 85 % accuracy. The sensitivity and specificity of the MELD approach were lower when the structural features of the seizure focus was unclear. The Australian Epilepsy Project aims to implement MELD in the patient-specific clinical report that is sent to the referring neurologist. This report contains detailed clinical, genetic, and imaging findings that are intended for clinical support. In the Australian Epilepsy Project report, an outline of MELD findings is provided, in addition to a lay summary of what the MELD model does, how it is computed, and how it should be interpreted regarding false positives and negatives. Initial experience suggests that identified MELD target lesions may provide a starting point for clinicians to have an in-depth and zoom-in look at specific brain regions, so the manual lesion detection task is no longer a 'finding a needle in the haystack', especially for subtle lesions such as cortical dysplasia.

Another area in need of improvement in AI is research reporting, in order to trust research findings. Several reporting checklists encourage detailed reporting of AI methodologies, facilitating transparency and replicability. One of the most common reporting frameworks in the field is the Transparent Reporting of studies on prediction models for Individual Prognosis Or Diagnosis (TRIPOD [41]). A systematic review by Nagendram et al. [42] summarised adherence to the TRIPOD framework in AI studies. The authors found that in most studies reviewed, less than half of the data items in the TRIPOD framework were reported in peer-reviewed publications. These results show that reporting AI standards are inadequate, but a potential issue with the TRIPOD reporting is that it was made in an 'older era' of prediction modelling. The team is working on a TRIPOD-AI checklist for newer AI models [43]. In the meantime, other AI-specific reporting checklists have also been created [44–46]. One of the newer checklists is the Data, Optimization, Model, and Evaluation (DOME) framework [47]. The DOME framework is developed explicitly with contemporary AI and prediction models in mind. It includes a comprehensive list of items in AI research design, train/test/validation dataset split, and model parameter selection. Datasheets documenting key characteristics of datasets improve transparency and accountability of models built [48].

In addition to reporting checklists, registering publications is a way to increase transparency in research, and to foster trust. In a registered study, a manuscript submitted in two stages (before and after results are obtained) is pre-accepted for journal publication [49]. Emerging evidence suggests that study bias is reduced in pre-registered reports. Evidence indicates that five times more studies report no statistically significant findings if manuscripts are registered compared to manuscripts that are not registered. This practice also allows for the publication of negative findings [50], providing a more reliable and comprehensive picture of relevant evidence. This is in line with the benefits of clinical trial registration [51] and peer-reviewed publication of systematic review protocols via, e.g., PROSPERO [52] or Cochrane (https://www.cochrane.org/), including to avoid selective outcome reporting [53,54].

Non-transparent models are a concerning trend amongst large companies. For example, in recent times, Open-AI's large language model, GPT, had not been released for open scientific exploration, nor had adequate details been provided about the resources and models on which it is based to enable meaningful evaluation. This contradicts the idea of safe and transparent AI. There is already substantial fear of AI in the community [55]. Sanderson [56] has emphasized the critical need for developing adaptive guidelines that govern how AI technologies are used and developed. Legal frameworks will likely struggle to keep up with the pace of AI development. The European Union (EU) has proposed the 'AI-act' as a legal AI framework, as a unified regulatory and legal framework for AI. It will likely change the landscape of AI, ensuring that AI serves people and is a positive contribution to society (https://artificialintelligenceact.eu/). An open letter signed by some key opinion leaders called for a six-month pause on giant AI experiments that are yielding "unpredictable black-box models with emergent capabilities", urging AI developers to "work with policymakers and regulators to dramatically accelerate the development of robust AI governance systems" (see https://futureoflife.org/open-letter/pause-giant-ai-experiments/). While the AI developments planned by the Australian Epilepsy Project are limited in scope, the need for oversight of the use of resultant AI models is well recognized, including the need for appropriate education of clinicians wishing to take advantage of AI-powered decision support tools. AI-based education and research are interconnected and mutually reinforcing, enhancing the interpretability and applicability of AI models and cultivating a workforce capable of critically assessing and refining AI-driven clinical decision-making. The

Australian Epilepsy Project actively encourage, facilitate and coordinate AI training.

Given the project's patient-centric focus, transparency and inclusivity are prioritized. Patients and their families, who have often endured years of hopelessness, form a crucial part of the intended beneficiaries. The project seeks to improve clinical practice and foster trust by developing and utilizing AI-based clinical support tools. Effective communication of this trust from clinicians to patients and their families is paramount. Also, a recent epistemic view of AI illusions highlights that AI users often overestimate their understanding and tend to believe that the models explore a broader range of hypotheses than they do [57]. As AI developers and users, this reminds us to be cautious of overestimating their knowledge of AI models. Purpose-driven research, trustworthiness, and a transdisciplinary approach are essential in moving medical AI beyond the imaging domain and into the clinical setting. The Australian Epilepsy Project aims to share its experience in creating robust AI models for clinical care. As other clinical platforms endeavour to implement AI, the project's insights can serve as a valuable resource. The ultimate goal is to improve clinical care by integrating trustworthy AI technologies that benefit patients and healthcare practices.

Australian Epilepsy Project commitment to AI trust:

- Encourage using explainable AI algorithms (e.g., MELD lesion detection [40]).
- Encourage using AI reporting checklists and registered reports to enhance transparency and reproducibility.
- The project is developing online teaching courses for clinicians and researchers to increase competency and confidence in AI models. These courses cover various aspects, from the basics of AI to advanced models specific to epilepsy-related questions.

5. Responsibility in AI

When it comes to responsibility in AI, we are talking about key topics such as accountability, liability, and acting with integrity [58–60]. In medicine, we work with people's most private information and potentially influence life-changing decisions. The argument of "who is responsible" for AI models and their decision-making is not trivial. Consider a scenario where a novel (and nominally validated) deep learning algorithm has detected a potential brain 'lesion' in a patient with refractory focal epilepsy. The decision in this scenario was to use the AI-based lesion detection result as a target for brain surgery which is often curative in epilepsy [61]. But, in this case, the surgery on the brain area targeted by the AI algorithm was unsuccessful, and the patient still experienced recurrent seizures. Who is responsible for this outcome? The medical professional or the AI algorithm who 'blindly' identified a brain region that, in hindsight, was the wrong target for surgery? Ultimately, treating physicians are responsible for patients, but it highlights an important consideration as AI-based solutions become more integrated into clinical workflows. Therefore, a pertinent current issue in AI is to determine how much we can trust AI to correctly advise us of the ground truth.

Study bias is a common and intrinsic issue in research studies. For instance, clinical trial participation is often lower amongst women, people of colour, and older people [62]. Genome-wide association studies have a strong Eurocentric bias, with 79 % of GWAS participants of European descent. Polygenic risk scores derived from this data have far lower predictive value than non-European populations [63]. Even in mega-sized community studies such as the UK biobank project, there is evidence of study bias. The UK biobank study aims to provide multimodal data from 500,000 people representing a 'snapshot' of the aging population in the UK between 49 and 70 years. Nevertheless, in the UK biobank study, the subjects are generally older, more often female, and have a higher socioeconomic status than the general population in the UK [64]. Subjects in the UK biobank study were also less likely to be obese, to smoke, to drink alcohol daily, and to have fewer self-reported health conditions when compared to the general population. The UK Biobank is an excellent initiative and an invaluable resource for extensive population research. Still, given the 'volunteer bias' evidence, it may not represent the sampling population.

Biases in data acquisition also mean that AI algorithms will inevitably be biased. Medical AI exhibits bias, too, often with insufficient generalisation and skewed samples regarding gender, age, racial and ethnic backgrounds, hospitals, and methods related to data collection [65], contributing to algorithmic bias [66]. Wang and colleagues demonstrated reduced bias if MRI-based AI models include multisource data, including demographic, clinical, genetic, and cognitive scores. The U.S. Food and Drug Administration (FDA) recently presented a draft document requiring overt plans for having diversity in clinical trials (https://www.fda.gov/media/106965/download), which signifies a step toward less biased and community-representative studies. This FDA document outlines how a researcher should describe and explain the rationale behind the anticipated enrolment of individuals from underrepresented groups. Similar principles apply to ensure that large-scale health datasets are sufficiently diverse and representative [66].

Although the hope is that AI will ultimately play a significant role in reducing the effects of climate change and other environmental impacts [67], currently, AI leaves a large carbon footprint on the planet [68–70]. For example, ~15 % of Google's energy usage was dedicated to AI processing over the last 3 years – and the large language model used to train the GPT-3 model (the model behind ChatGPT), with approximately 175 billion parameters, used 552.1 tons of CO_2-equivalent emissions to train [71]. It is not only training AI algorithms that use large amounts of energy. The majority of Facebook's carbon emission is used for AI model inference –i.e., model adaptation to new data– of their AI algorithms [72] and lifecycle of the Amazon Echo incurs a massive toll on nonrenewable materials, labor, and data [73]. In this context, it is also important to consider that environmental impacts disproportionately affect marginalised communities [74].

There are several ways to reduce carbon emissions in AI. One of these includes accessing computing resources in regions with cleaner energy. Accessing computing resources in regions powered by more renewable energy sources can reduce emissions by up to 30 times [75]. Another way to reduce carbon emission footprints is to explore whether a research question can be answered with smaller AI models (as large AI models require the most energy [71]). Reporting the sensitivity of models to hyperparameters also allows for understanding training resource requirements [68]. It is furthermore good practice to use frequent checkpoints when training AI models. Frequent checkpoints pause the training and allow the researcher to estimate the model's performance. With checkpointing, errors or sub-optimal performance can be detected early in the training of models, potentially saving time and energy. It is worth noting that checkpointing can require a large amount of storage, and consequent energy consumption, depending on the amount of information to evaluate. A way to track the carbon footprint of AI models is the CodeCarbon Python package (https://github.com/mlco2/codecarbon) [76,77]. CodeCarbon calculates the electricity consumption from GPU, CPU, and RAM, and then summarizes the carbon footprint statistics for the AI model.

Australian Epilepsy Project commitment to AI responsibility:

- AI-based decision-making is based on human knowledge
- Democratise clinical access for people with epilepsy in Australia.
- Aim to create diverse, representative, and multimodal datasets that comply with standards for transparency of data diversity, such as those from the STANDING together initiative [78].
- Encourage tracking the carbon footprint of AI models (e.g., CodeCarbon Python package [76,77])

6. Safety in AI

A recent example of why governance and regulations are needed in AI comes from a recent report from Urbina et al. [79]. The authors, part of a company developing drug compounds to combat human disease, were invited to a conference hosted by the Swiss government. This conference scrutinizes potential harm and adverse outcomes from AI technologies. Based on this conference, Urbina et al. inverted their original 'good' AI algorithm using chemical compounds to a new model designed to do 'harm'. By transforming the positive algorithm into a harmful one, the authors created 40,000 compounds that could be used in chemical warfare in less than six hours of data training. Many compounds were deemed potentially more dangerous than current warfare agents. The authors acknowledge that they had never previously considered that their AI models with chemical agents could be used to cause harm. This example is an extreme scenario, as working with chemical compounds is potentially risky. Nevertheless, this study highlights the need to consider the known and potentially unknown implications of AI and modelling, especially in medicine, as these AI models influence people's lives and well-being. Indeed, even with relatively few AI tools actually demonstrating impact in clinical practice [80,81], hundreds of safety events have already been reported [82]. As identified by the Australian National Policy Roadmap for AI in Healthcare [83], AI must be safe, effective, and do no harm.

Sharing of raw data can be a double-edged sword. It enables transparency, which is good. But it also increases data safety and protection risks, especially disclosing people's identities. Although most people are happy for their data to be publicly shared [84], strategies are needed to ensure safe data sharing, as the re-identification of individuals is a risk when openly releasing data. For example, if a person's whole genome is freely available, re-identifying their identity is possible [85]. It is also possible to re-identify individuals based on MRI of people's faces. Without de-facing MRI images, automated face recognition was 97 % accurate in matching participants' real photographs to the correct MRI, and software packages exist to remove MRI features of people's faces before publicly releasing data [86]. The Australian Epilepsy Project is committed to publicly releasing data to the best standards in the field. When sharing data, we will consider the intended use of the data and the scope of the disclosure to determine how the data will be shared.

Another way to safely release data to the public is to generate synthetic data. This is because synthetic data has no identity and can cause similar features to the original, and real, data. Latent Diffusion Models, such as stable diffusion [87], have overtaken Generative Adversarial Networks [88] in the last year as the primary method for synthetic data generation. Pinaya et al. [89] recently showed that latent diffusion models could generate realistic MRI data using stable diffusion algorithms and released 100,000 synthetic MRI datasets with varying ages and ventricular volumes that can boost existing datasets, enabling large-scale AI approaches [90].

Overall, the safe development and use of data-driven clinical AI requires strong governance of both data and models, encompassing fairness, transparency, trustworthiness and accountability, as well as oversight from stakeholders including patient representatives, clinicians, and technical experts [91]. Algorithmic sovereignty [92] and data sovereignty, including those for indigenous populations, should be considered in any medical implementation of AI. Examples from non-AI patient data sharing and linkage platforms such as BioGrid Australia [93] can be leveraged to inform how this is approached and extended into the AI context.

Australian Epilepsy Project commitment to AI safety:

- Sharing data to the best standards, with minimal probability of re-identification (e.g., operate within The Five Safes Framework [94])
- Commitment to data and model governance to ensure that AI development and implementation decisions are carefully considered.

7. Flexibility and adaptability in the fast-changing field of medical AI

The metrics that we use to judge and evaluate AI systems need careful consideration. While model accuracy, calibration, and robustness are common intrinsic evaluation metrics, other dimensions of AI ethics goals, including transparency, efficiency, fairness, and bias, should be quantified and reported. Various frameworks for AI model reporting have been proposed, such as model cards [95] to document the benchmark performance of models under various conditions. Proposed conditions include many directly relevant to medicine (e.g. cultural, demographic, or phenotypic groups). Medical applications of AI may require new metrics, and possibly additional dimensions – e.g. for generative language models, researchers have proposed that toxicity should be added to the list [96]. An additional concern for metric-based evaluation of AI models in a biomedical context is to balance the need for short-term decisions about optimal model performance with the need for improved long-term health outcomes in patients. Here the adage known as Goodhart's law may be relevant, which is often phrased as "when a measure becomes a target, it ceases to be a good measure". Care must be taken to avoid optimizing AI models for short-term metrics, such as performance on an imaging-based lesion detection task, which may not reflect the ultimate patient-centred goal of achieving improved quality of life. We must continuously mature our approach to evaluation.

The Australian Epilepsy Project acknowledges the gradual adoption of novel technologies within clinical settings. With a foundation of three decades of extensive clinical and imaging research, we have observed a slow uptake of new technology in healthcare. This observation is significant as the AI landscape has rapidly evolved worldwide. Consequently, the Australian Epilepsy Project remains focused on its clinical endpoint while embracing the advancements and improvements in AI prediction models. The project aims to develop a flexible AI framework that enables algorithm selection and fine-tuning to align with clinical needs. While acknowledging the challenges of comparing different brain conditions, the Australian Epilepsy Project strives to serve as a blueprint for future platform initiatives across various psychiatric and neurological disorders. We aim to build systems that operate at scale, and the underlying technologies utilized in the Australian Epilepsy Project (imaging, tele-neuropsychology, genetics) could equally be employed for other neurological disorders and comorbidities, as epilepsy is associated with a range of comorbidities, including depression, anxiety and suicidality (Kanner et al. [97]). While AI solutions may not be universally applicable, certain aspects related to study operations, such as extracting categorical data from medical records and neuroimaging quality assurance processes, may transfer across domains.

AI is changing fast, and it sometimes seems futile to predict the next 5 and 10 years of AI development and its impact on medicine. Therefore, we need to have a flexible and adaptable approach to responsible AI that can help address the concerns of various stakeholders. Viewpoints on responsible AI from patients, healthcare providers, regulators, and technology developers are needed to generate adaptive frameworks that support safe AI to benefit health and society. The Australian Epilepsy Project is a patient-focused project led by scientists and clinicians, and the clinical expertise plays an integral role in AI model development and use. In a field where most advances reside in private sectors and large companies (~70 % of all AI PhD currently enters industry-based jobs, and nearly all large-AI models are trained by companies [98]), we believe we represent a voice that should be heard in the medical AI community to ensure responsible implementation of AI in a clinical setting. We hope our approach toward augmented intelligence in the Australian Epilepsy Project can serve as a blueprint for other projects that leverage human expertise to generate augmented intelligence-based medicine. We envisage an environment in the Australian Epilepsy Project where AI provides predictions and suggestions, and humans make the decisions.

CRediT authorship contribution statement

Mangor Pedersen: Writing – original draft, Investigation, Formal analysis, Conceptualization. **Heath R. Pardoe:** Writing – review & editing, Investigation, Conceptualization. **Anton de Weger:** Writing – review & editing, Investigation. **Donna Hutchison:** Writing – review & editing, Resources, Methodology, Investigation. **David F. Abbott:** Writing – review & editing, Investigation, Funding acquisition, Conceptualization. **Karin Verspoor:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Graeme D. Jackson:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgements

This work was supported by an Australian Government Medical Research Future Fund Frontier Health Grant (RFRHPSI000008). MP acknowledges funding from Health Research Council (HRC), New Zealand, Emerging Researcher Grant. D.F.A. acknowledges fellowship funding from the Australian National Imaging Facility.

The Florey Institute of Neuroscience and Mental Health acknowledges the strong support from the Victorian Government and, in particular, the funding from the Operational Infrastructure Support Grant. The authors acknowledge the facilities and scientific and technical assistance of the National Imaging Facility, a National Collaborative Research Infrastructure Strategy (NCRIS) capability, at the Florey Institute of Neuroscience and Mental Health.

We wish to thank the anonymous reviewer that greatly helped shaping this work into a framework toward responsible AI in medicine.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.artmed.2025.103192.

References

[1] McCarthy J. What is artificial intelligence?. 2004.
[2] Weng C, Shah NH, Hripcsak G. Deep phenotyping: embracing complexity and temporality—towards scalability, portability, and interoperability. J Biomed Inform 2020;105:103433. https://doi.org/10.1016/j.jbi.2020.103433.
[3] Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med 2022;28:1773–84. https://doi.org/10.1038/s41591-022-01981-2.
[4] Braun M, Hummel P, Beck S, Dabrock P. Primer on an ethics of AI-based decision support systems in the clinic. J Med Ethics 2021;47:e3. https://doi.org/10.1136/medethics-2019-105860.
[5] Geis JR, Brady AP, Wu CC, Spencer J, Ranschaert E, Jaremko JL, et al. Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. Radiology 2019;293:436–40. https://doi.org/10.1148/radiol.2019191586.
[6] Goldsmith J, Burton E. Why teaching ethics to AI practitioners is important. Proc AAAI Conf Artif Intell 2017;31. https://doi.org/10.1609/aaai.v31i1.11139.
[7] Gundersen T, Bærøe K. The future ethics of artificial intelligence in medicine: making sense of collaborative models. Sci Eng Ethics 2022;28:17. https://doi.org/10.1007/s11948-022-00369-2.
[8] Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism 2017;69:S36–40. https://doi.org/10.1016/j.metabol.2017.01.011.
[9] Martinho A, Kroesen M, Chorus C. A healthy debate: exploring the views of medical doctors on the ethics of artificial intelligence. Artif Intell Med 2021;121:102190. https://doi.org/10.1016/j.artmed.2021.102190.
[10] McLennan S, Fiske A, Tigard D, Müller R, Haddadin S, Buyx A. Embedded ethics: a proposal for integrating ethics into the development of medical AI. BMC Med Ethics 2022;23:6. https://doi.org/10.1186/s12910-022-00746-3.
[11] Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. Soc Sci Med 2020;260:113172. https://doi.org/10.1016/j.socscimed.2020.113172.
[12] Muller H, Mayrhofer MT, Van Veen E-B, Holzinger A. The ten commandments of ethical medical AI. Computer 2021;54:119–23. https://doi.org/10.1109/MC.2021.3074263.
[13] Naik N, Hameed BMZ, Shetty DK, Swain D, Shah M, Paul R, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility? Front Surg 2022;9:862322. https://doi.org/10.3389/fsurg.2022.862322.
[14] Saheb T, Saheb T, Carpenter DO. Mapping research strands of ethics of artificial intelligence in healthcare: a bibliometric and content analysis. Comput Biol Med 2021;135:104660. https://doi.org/10.1016/j.compbiomed.2021.104660.
[15] Sand M, Durán JM, Jongsma KR. Responsibility beyond design: physicians' requirements for ethical medical AI. Bioethics 2022;36:162–9. https://doi.org/10.1111/bioe.12887.
[16] Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. PLoS Med 2018;15:e1002689. https://doi.org/10.1371/journal.pmed.1002689.
[17] Baker GA. Depression and suicide in adolescents with epilepsy. Neurology 2006;66 (Suppl. 3):S5–12. https://doi.org/10.1212/WNL.66.66_suppl_3.S5.
[18] Berg AT. Epilepsy, cognition, and behavior: the clinical picture. Epilepsia 2011;52:7–12. https://doi.org/10.1111/j.1528-1167.2010.02905.x.
[19] Elger CE, Helmstaedter C, Kurthen M. Chronic epilepsy and cognition. Lancet Neurol 2004;3:663–72. https://doi.org/10.1016/S1474-4422(04)00906-8.
[20] Kerr MP. The impact of epilepsy on patients' lives. Acta Neurol Scand 2012;126:1–9. https://doi.org/10.1111/ane.12014.
[21] Motamedi G, Meador K. Epilepsy and cognition. Epilepsy Behav 2003;4:25–38. https://doi.org/10.1016/j.yebeh.2003.07.004.
[22] Smeets VMJ, van Lierop BAG, Vanhoutvin JPG, Aldenkamp AP, Nijhuis FJN. Epilepsy and employment: literature review. Epilepsy Behav 2007;10:354–62. https://doi.org/10.1016/j.yebeh.2007.02.006.
[23] Foster E, Chen Z, Zomer E, Rychkova M, Carney P, O'Brien TJ, et al. The costs of epilepsy in Australia: a productivity-based analysis. Neurology 2020;95:e3221–31. https://doi.org/10.1212/WNL.0000000000010862.
[24] Pedersen M, Verspoor K, Abbott DF, Jenkinson M, Sinclair B, Law M, et al. Deep learning for reliable detection of epileptogenic lesions. Augmenting neurological disorder prediction and rehabilitation using artificial intelligence. Elsevier; 2022. p. 163–75.
[25] Pedersen M, Verspoor K, Jenkinson M, Law M, Abbott DF, Jackson GD. Artificial intelligence for clinical decision support in neurology. Brain Commun 2020;2. https://doi.org/10.1093/braincomms/fcaa096.
[26] Tailby C, Collins AJ, Vaughan DN, Abbott DF, O'Shea M, Helmstaedter C, et al. Teleneuropsychology in the time of COVID-19: the experience of the Australian epilepsy project. Seizure 2020;83:89–97. https://doi.org/10.1016/j.seizure.2020.10.005.
[27] Kwan P, Brodie MJ. Early identification of refractory epilepsy. N Engl J Med 2000;342:314–9. https://doi.org/10.1056/NEJM200002033420503.
[28] Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell 2019;1:389–99. https://doi.org/10.1038/s42256-019-0088-2.
[29] Gennatas ED, Friedman JH, Ungar LH, Pirracchio R, Eaton E, Reichmann LG, et al. Expert-augmented machine learning. PNAS 2020;117:4571–7. https://doi.org/10.1073/pnas.1906831117.
[30] Zheng N, Liu Z, Ren P, Ma Y, Chen S, Yu S, et al. Hybrid-augmented intelligence: collaboration and cognition. Front Inf Technol Electron Eng 2017;18:153–79. https://doi.org/10.1631/FITEE.1700053.
[31] Bazoukis G, Hall J, Loscalzo J, Antman EM, Fuster V, Armoundas AA. The inclusion of augmented intelligence in medicine: a framework for successful implementation. Cell Rep Med 2022;3:100485. https://doi.org/10.1016/j.xcrm.2021.100485.
[32] Lederman A, Lederman R, Verspoor K. Tasks as needs: reframing the paradigm of clinical natural language processing research for real-world decision support. J Am Med Inform Assoc 2022;29:1810–7. https://doi.org/10.1093/jamia/ocac121.
[33] Ehsan U, Liao QV, Muller M, Riedl MO, Weisz JD. Expanding explainability: towards social transparency in AI systems. Proceedings of the 2021 CHI conference on human factors in computing systems. New York, NY, USA: Association for Computing Machinery; 2021. p. 1–19. https://doi.org/10.1145/3411764.3445188.
[34] Felzmann H, Fosch-Villaronga E, Lutz C, Tamò-Larrieux A. Towards transparency by Design for Artificial Intelligence. Sci Eng Ethics 2020;26:3333–61. https://doi.org/10.1007/s11948-020-00276-4.
[35] Larsson S, Heintz F. Transparency in artificial intelligence. Internet Policy Rev 2020;9:1–16. https://doi.org/10.14763/2020.2.1469.
[36] Castelvecchi D. Can we open the black box of AI? Nat News 2016;538:20. https://doi.org/10.1038/538020a.
[37] Baud MO, Perneger T, Rácz A, Pensel MC, Elger C, Rydenhag B, et al. European trends in epilepsy surgery. Neurology 2018;91:e96–106. https://doi.org/10.1212/WNL.0000000000005776.
[38] Sinclair B, Cahill V, Seah J, Kitchen A, Vivash LE, Chen Z, et al. Machine learning approaches for imaging-based prognostication of the outcome of surgery for mesial temporal lobe epilepsy. Epilepsia 2022;63:1081–92. https://doi.org/10.1111/epi.17217.
[39] Suzuki K. Overview of deep learning in medical imaging. Radiol Phys Technol 2017;10:257–73. https://doi.org/10.1007/s12194-017-0406-5.
[40] Spitzer H, Ripart M, Whitaker K, D'Arco F, Mankad K, Chen AA, et al. Interpretable surface-based detection of focal cortical dysplasias: a Multi-centre Epilepsy Lesion Detection study. Brain 2022;145:3859–71. https://doi.org/10.1093/brain/awac224.
[41] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med 2015;13:1. https://doi.org/10.1186/s12916-014-0241-z.
[42] Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting

standards, and claims of deep learning studies. BMJ 2020;368:m689. https://doi.org/10.1136/bmj.m689.

[43] Collins GS, Dhiman P, Navarro CLA, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021;11:e048008. https://doi.org/10.1136/bmjopen-2020-048008.

[44] Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell 2020;2:e200029. https://doi.org/10.1148/ryai.2020200029.

[45] Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med 2020;26:1320–4. https://doi.org/10.1038/s41591-020-1041-y.

[46] Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ 2022;377:e070904. https://doi.org/10.1136/bmj-2022-070904.

[47] Walsh I, Fishman D, Garcia-Gasulla D, Titma T, Pollastri G, Harrow J, et al. DOME: recommendations for supervised machine learning validation in biology. Nat Methods 2021;18:1122–7. https://doi.org/10.1038/s41592-021-01205-4.

[48] Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, HD III, et al. Datasheets for datasets. Commun ACM 2021;64:86–92. https://doi.org/10.1145/3458723.

[49] Chambers CD, Tzavella L. The past, present and future of registered reports. Nat Hum Behav 2022;6:29–42. https://doi.org/10.1038/s41562-021-01193-7.

[50] Allen C, Mehler DMA. Open science challenges, benefits and tips in early career and beyond. PLoS Biol 2019;17:e3000246. https://doi.org/10.1371/journal.pbio.3000246.

[51] Levin LA, Palmer JG. Institutional review boards should require clinical trial registration. Arch Intern Med 2007;167:1576–80. https://doi.org/10.1001/archinte.167.15.1576.

[52] Booth A, Clarke M, Ghersi D, Moher D, Petticrew M, Stewart L. An international registry of systematic-review protocols. Lancet 2011;377:108–9. https://doi.org/10.1016/S0140-6736(10)60903-8.

[53] Stewart L, Moher D, Shekelle P. Why prospective registration of systematic reviews makes sense. Syst Rev 2012;1:7. https://doi.org/10.1186/2046-4053-1-7.

[54] Ge L, Tian J, Li Y, Pan J, Li G, Wei D, et al. Association between prospective registration and overall reporting and methodological quality of systematic reviews: a meta-epidemiological study. J Clin Epidemiol 2018;93:45–55. https://doi.org/10.1016/j.jclinepi.2017.10.012.

[55] Cugurullo F, Acheampong RA. Fear of AI: an inquiry into the adoption of autonomous cars in spite of fear, and a theoretical framework for the study of artificial intelligence technology acceptance. AI & Soc 2023. https://doi.org/10.1007/s00146-022-01598-6.

[56] Sanderson K. GPT-4 is here: what scientists think. Nature 2023. https://doi.org/10.1038/d41586-023-00816-5.

[57] Messeri L, Crockett MJ. Artificial intelligence and illusions of understanding in scientific research. Nature 2024;627:49–58. https://doi.org/10.1038/s41586-024-07146-0.

[58] Coeckelbergh M. Artificial intelligence, responsibility attribution, and a relational justification of explainability. Sci Eng Ethics 2020;26:2051–68. https://doi.org/10.1007/s11948-019-00146-8.

[59] Constantinescu M, Voinea C, Uszkai R, Vică C. Understanding responsibility in responsible AI. Dianoetic virtues and the hard problem of context. Ethics Inf Technol 2021;23:803–14. https://doi.org/10.1007/s10676-021-09616-9.

[60] Tigard DW. Responsible AI and moral responsibility: a common appreciation. AI Ethics 2021;1:113–7. https://doi.org/10.1007/s43681-020-00009-0.

[61] Kral T, Clusmann H, Blümcke I, Fimmers R, Ostertun B, Kurthen M, et al. Outcome of epilepsy surgery in focal cortical dysplasia. J Neurol Neurosurg Psychiatry 2003;74:183–8. https://doi.org/10.1136/jnnp.74.2.183.

[62] Chastain DB, Osae SP, Henao-Martínez AF, Franco-Paredes C, Chastain JS, Young HN. Racial disproportionality in Covid clinical trials. N Engl J Med 2020;383:e59. https://doi.org/10.1056/NEJMp2021971.

[63] Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 2019;51:584–91. https://doi.org/10.1038/s41588-019-0379-x.

[64] van Alten S, Domingue BW, Galama T, Marees AT. Reweighting the UK biobank to reflect its underlying sampling population substantially reduces pervasive selection bias due to volunteering2022.05.16.22275048; 2022. https://doi.org/10.1101/2022.05.16.22275048.

[65] Wang R, Chaudhari P, Davatzikos C. Bias in machine learning models can be significantly mitigated by careful training: evidence from neuroimaging studies. Proc Natl Acad Sci 2023;120:e2211613120. https://doi.org/10.1073/pnas.2211613120.

[66] Arora A, Alderman JE, Palmer J, Ganapathi S, Laws E, McCradden MD, et al. The value of standards for health datasets in artificial intelligence-based applications. Nat Med 2023;29:2929–38. https://doi.org/10.1038/s41591-023-02608-w.

[67] Cowls J, Tsamados A, Taddeo M, Floridi L. The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations. AI & Soc 2023;38:283–307. https://doi.org/10.1007/s00146-021-01294-x.

[68] Strubell E, Ganesh A, McCallum A. Energy and policy considerations for deep learning in NLP. Proceedings of the 57th annual meeting of the association for computational linguistics. Florence, Italy: Association for Computational Linguistics; 2019. p. 3645–50. https://doi.org/10.18653/v1/P19-1355.

[69] Dhar P. The carbon impact of artificial intelligence. Nat Mach Intell 2020;2:423–5. https://doi.org/10.1038/s42256-020-0219-9.

[70] Schwartz R, Dodge J, Smith NA, Etzioni O. Green AI. Commun ACM 2020;63:54–63. https://doi.org/10.1145/3381831.

[71] Patterson D, Gonzalez J, Hölzle U, Le Q, Liang C, Munguia L-M, et al. The carbon footprint of machine learning training will plateau, then shrink. Computer 2022;55:18–28. https://doi.org/10.1109/MC.2022.3148714.

[72] Wu C-J, Raghavendra R, Gupta U, Acun B, Ardalani N, Maeng K, et al. Sustainable AI: environmental implications, challenges and opportunities. 2021. https://doi.org/10.48550/arXiv.2111.00364.

[73] Anatomy of an AI System. Anatomy of an AI system n.d. http://www.anatomyof.ai (accessed April 12, 2023).

[74] Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: can language models be too big?. Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. New York, NY, USA: Association for Computing Machinery; 2021. p. 610–23. https://doi.org/10.1145/3442188.3445922.

[75] Henderson P, Hu J, Romoff J, Brunskill E, Jurafsky D, Pineau J. Towards the systematic reporting of the energy and carbon footprints of machine learning. J Mach Learn Res 2020;21:248:10039–248:10081.

[76] Lottick K, Susai S, Friedler SA, Wilson JP. Energy usage reports: environmental awareness as part of algorithmic accountability. 2019. https://doi.org/10.48550/arXiv.1911.08354.

[77] Lacoste A, Luccioni A, Schmidt V, Dandres T. Quantifying the carbon emissions of machine learning. 2019. https://doi.org/10.48550/arXiv.1910.09700.

[78] Ganapathi S, Palmer J, Alderman JE, Calvert M, Espinoza C, Gath J, et al. Tackling bias in AI health datasets through the STANDING Together initiative. Nat Med 2022;28:2232–3. https://doi.org/10.1038/s41591-022-01987-w.

[79] Urbina F, Lentzos F, Invernizzi C, Ekins S. Dual use of artificial-intelligence-powered drug discovery. Nat Mach Intell 2022;4:189–91. https://doi.org/10.1038/s42256-022-00465-9.

[80] Susanto AP, Lyell D, Widyantoro B, Berkovsky S, Magrabi F. How well do AI-enabled decision support systems perform in clinical settings? Stud Health Technol Inform 2024;310:279–83. https://doi.org/10.3233/SHTI230971.

[81] Marwaha JS, Kvedar JC. Crossing the chasm from model performance to clinical impact: the need to improve implementation and evaluation of AI. NPJ Digit Med 2022;5:25. https://doi.org/10.1038/s41746-022-00572-2.

[82] Lyell D, Wang Y, Coiera E, Magrabi F. More than algorithms: an analysis of safety events involving ML-enabled medical devices reported to the FDA. J Am Med Inform Assoc 2023;30:1227–36. https://doi.org/10.1093/jamia/ocad065.

[83] Coiera E, Magrabi F, Hansen D, Verspoor K. A national policy roadmap for artificial intelligence in healthcare. 2023.

[84] Mello MM, Lieou V, Goodman SN. Clinical trial participants' views of the risks and benefits of data sharing. N Engl J Med 2018;378:2202–11. https://doi.org/10.1056/NEJMsa1713258.

[85] Erlich Y, Shor T, Pe'er I, Carmi S. Identity inference of genomic data using long-range familial searches. Science 2018;362:690–4. https://doi.org/10.1126/science.aau4832.

[86] Schwarz CG, Kremers WK, Wiste HJ, Gunter JL, Vemuri P, Spychalla AJ, et al. Changing the face of neuroimaging research: comparing a new MRI de-facing technique with popular alternatives. NeuroImage 2021;231:117845. https://doi.org/10.1016/j.neuroimage.2021.117845.

[87] Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. 2022. https://doi.org/10.48550/arXiv.2112.10752.

[88] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. 2014. https://doi.org/10.48550/arXiv.1406.2661.

[89] Pinaya WHL, Tudosiu P-D, Dafflon J, da Costa PF, Fernandez V, Nachev P, et al. Brain imaging generation with latent diffusion models. arXivOrg 2022. https://doi.org/10.48550/arXiv.2209.07162.

[90] Marquand AF, Rezek I, Buitelaar J, Beckmann CF. Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. Biol Psychiatry 2016;80:552–61. https://doi.org/10.1016/j.biopsych.2015.12.023.

[91] Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc 2020;27:491–7. https://doi.org/10.1093/jamia/ocz192.

[92] Coiera EW, Verspoor K, Hansen DP. We need to chat about artificial intelligence. Med J Aust 2023;219:98–100. https://doi.org/10.5694/mja2.51992.

[93] Merriel RB, Gibbs P, O'Brien TJ, Hibbert M. BioGrid Australia facilitates collaborative medical and bioinformatics research across hospitals and medical research institutes by linking data from diverse disease and data types. Hum Mutat 2011;32:517–25. https://doi.org/10.1002/humu.21437.

[94] Ritchie F. Five Safes: designing data access for research. 2016. https://doi.org/10.13140/RG.2.1.3661.1604.

[95] Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al. Model cards for model reporting. Proceedings of the conference on fairness, accountability, and transparency. New York, NY, USA: Association for Computing Machinery; 2019. p. 220–9. https://doi.org/10.1145/3287560.3287596.

[96] Liang P, Bommasani R, Lee T, Tsipras D, Soylu D, Yasunaga M. Holistic evaluation of language models. 2022.

[97] Kanner AM. Can the development of mood and anxiety disorders during pregnancy and the postpartum period be anticipated in women with epilepsy? Neurology 2022;99:637–8. https://doi.org/10.1212/WNL.0000000000201171.

[98] Ahmed N, Wahed M, Thompson NC. The growing influence of industry in AI research. Science 2023;379:884–6. https://doi.org/10.1126/science.ade2420.