Original article

The ethical, legal and social implications of using artificial intelligence systems in breast cancer care

Stacy M. Carter [a, *], Wendy Rogers [b], Khin Than Win [c], Helen Frazer [d], Bernadette Richards [e], Nehmat Houssami [f]

[a] Australian Centre for Health Engagement, Evidence and Values (ACHEEV), School of Health and Society, Faculty of Social Science, University of Wollongong, Northfields Avenue, New South Wales, 2522, Australia
[b] Department of Philosophy and Department of Clinical Medicine, Macquarie University, Balaclava Road, North Ryde, New South Wales, 2109, Australia
[c] Centre for Persuasive Technology and Society, School of Computing and Information Technology, Faculty of Engineering and Information Technology, University of Wollongong, Northfields Avenue, New South Wales, 2522, Australia
[d] Screening and Assessment Service, St Vincent's BreastScreen, 1st Floor Healy Wing, 41 Victoria Parade, Fitzroy, Victoria, 3065, Australia
[e] Adelaide Law School, Faculty of the Professions, University of Adelaide, Adelaide, South Australia, 5005, Australia
[f] Sydney School of Public Health, Faculty of Medicine and Health, Fisher Road, The University of Sydney, New South Wales, 2006, Australia

ARTICLE INFO

ABSTRACT

Breast cancer care is a leading area for development of artificial intelligence (AI), with applications including screening and diagnosis, risk calculation, prognostication and clinical decision-support, management planning, and precision medicine. We review the ethical, legal and social implications of these developments. We consider the values encoded in algorithms, the need to evaluate outcomes, and issues of bias and transferability, data ownership, confidentiality and consent, and legal, moral and professional responsibility. We consider potential effects for patients, including on trust in healthcare, and provide some social science explanations for the apparent rush to implement AI solutions. We conclude by anticipating future directions for AI in breast cancer care. Stakeholders in healthcare AI should acknowledge that their enterprise is an ethical, legal and social challenge, not just a technical challenge. Taking these challenges seriously will require broad engagement, imposition of conditions on implementation, and pre-emptive systems of oversight to ensure that development does not run ahead of evaluation and deliberation. Once artificial intelligence becomes institutionalised, it may be difficult to reverse: a proactive role for government, regulators and professional groups will help ensure introduction in robust research contexts, and the development of a sound evidence base regarding real-world effectiveness. Detailed public discussion is required to consider what kind of AI is acceptable rather than simply accepting what is offered, thus optimising outcomes for health systems, professionals, society and those receiving care.

In this article we consider the ethical, legal and social implications of the use of artificial intelligence (AI) in breast cancer care. This has been a significant area of medical AI development, particularly with respect to breast cancer detection and screening. The *Digital Mammography DREAM Challenge*, for example, aimed to generate algorithms that could reduce false positives without undermining cancer detection [1]; the best algorithm to date has reported 80.3–80.4% accuracy, and the next phase aims to develop an algorithm that can 'fully match the accuracy of an expert radiologist' [2]. In another example, Google Deepmind Health, NHS Trusts, Cancer Research UK and universities are developing machine learning technologies for mammogram reading [3]. Eventually, AI may seem as unremarkable as digital Picture Archiving and Communication Systems (PACS) do now. But at this critical moment, before widespread implementation, we suggest proceeding carefully to allow measured assessment of risks, benefits and potential harms.

In recent years there has been a stream of high-level

* Corresponding author.
E-mail addresses: stacy_carter@uow.edu.au (S.M. Carter), wendy.rogers@mq.edu.au (W. Rogers), win@uow.edu.au (K.T. Win), Helen.Frazer@svha.org.au (H. Frazer), bernadette.richards@adelaide.edu.au (B. Richards), nehmat.houssami@sydney.edu.au (N. Houssami).

governmental, intergovernmental, professional and industry statements on the Ethical, Legal and Social Implications (ELSI) of AI (at least 42 such statements as of June 2019 [4]), expressing both excitement about the potential benefits of AI and concern about potential harms and risks. The possible adverse consequences of AI are concerning enough in the context of social media platforms, or retail services. In the case of breast cancer detection, prognostication or treatment planning, potential harms compel serious examination. There is limited scholarship on the ELSI of the use of AI in breast cancer care, so the discussion that follows will draw heavily on, and apply, the general literature on the ELSI of AI.

This paper proceeds in three parts. In part one, we consider what AI is, and examples of its development and evaluation for potential use in breast cancer care. In part two, we outline the ELSI of AI. In the final section, we will anticipate future directions for AI in breast cancer care and draw some conclusions.

1. What is AI and how is it being used in breast cancer care?

There is much conceptual and terminological imprecision in public discussions about AI. Below is a general characterisation of the history and nature of AI; we acknowledge that this is contestable. We focus on the issues most relevant to our argument.

1.1. What is AI

Popular AI debate often focuses on Artificial General Intelligence (AGI) or strong AI: the kind that may, one day, be 'able to accomplish any cognitive task at least as well as humans' [5]. However most current advances involve narrow forms of AI, which can now complete a wide range of specific tasks, such as playing a board game, translating between languages, listening and responding to human instructions, or identifying specific patterns in visual data (e.g. recognising faces from CCTV images, or suspicious areas on a mammograms). These algorithms are built using a range of approaches and techniques (e.g. machine learning, deep learning, neural networks). There has been an important shift in AI development—including in healthcare—between the 1970s and now, and especially since 2006 [6]. As depicted in Fig. 1, this can be characterised as a shift from 'old AI' to 'new AI'.

Early clinical decision support systems utilised expert systems techniques, where humans were required to provide rules to the system. These 'old' clinical AIs—for example those using decision tree techniques—relied on humans to select which attributes should be included, so were largely transparent. It was easy to tell what a human-defined algorithm was doing.

In contrast, 'new AI' is characterised by the use of novel machine learning techniques (especially deep learning) that enable an algorithm to independently classify and cluster data. Rather than being explicitly programmed to pay attention to specific attributes or variables, these algorithms have the capacity to develop, by exposure to data, the ability to recognise patterns in that data. A goal is set (e.g. distinguish 'cancer' vs 'no cancer' on mammographic images) and the algorithm is exposed to large volumes of training data which may be more or less heterogeneous: for example, image data only, or image data and clinical outcomes data. Over time, the algorithm, independent of explicit human instruction, 'learns' to identify and extract relevant attributes from the data to achieve the goal. Unsupervised learning (for example self-organising map approaches) allow the algorithm to group data based on similarity, and to independently reduce the dimensionality of the data used to inform a decision. This *dimension reduction* process simplifies the data by, for example, selecting a subset of the data to process, or building a new set of features from the original data. These machine processes are not always reversible. Recently, developers have begun using ensemble methods, which combine several different AI techniques.

The performance of these new forms of AI is far better than that of 'old AI'. This has been largely responsible for recent rapid advances in healthcare applications. However, results obtained from these newer forms of AI, especially neural networks, can be difficult to interpret, not least because of their increasing complexity and their capacity for unsupervised or non-restricted learning. These algorithms are referred to as 'black boxes' because the detail of their operations is not always understandable to human observers.

1.2. How is AI being developed and used for breast cancer care?

Breast cancer has been of interest since the days of 'old AI', applied to problems including differential diagnosis [7], grading [8], and prognosis and management [9]. Breast cancer is also an important use case in 'new AI', especially for screening and diagnosis [10,11]. Image reading relies heavily on the visual acuity, experience, training and attention of human radiologists; the best human image reading is imperfect and it is not always clear how it can be improved [12,13], creating some appetite for AI solutions. 'New AI' is also being applied to breast cancer risk calculation, prognostication and clinical decision-support [14–16], management planning [17], and translation of genomics into precision medicine [18,19]. There is strong competitive private sector development of AI-based products, particularly for image-processing including mammography reading and tissue pathology diagnostics [10,20]. There are now breast cancer detection AI systems available for clinical application; marketing for these products promises greater efficiency, accuracy and revenue. Evidence to date, however, does not always warrant the enthusiasm of the marketing. A recent scoping review, for example, found that the accuracy of AI models for breast screening varied widely, from 69.2% to 97.8% (median AUC 88.2%) [10]. In addition, most studies were small and retrospective, with algorithms developed from datasets containing a disproportionately large number of mammograms with cancer (more than a quarter, when only 0.5–0.8% of women who present for mammography will have cancer detected), potential data bias, limited external validation, and limited data on
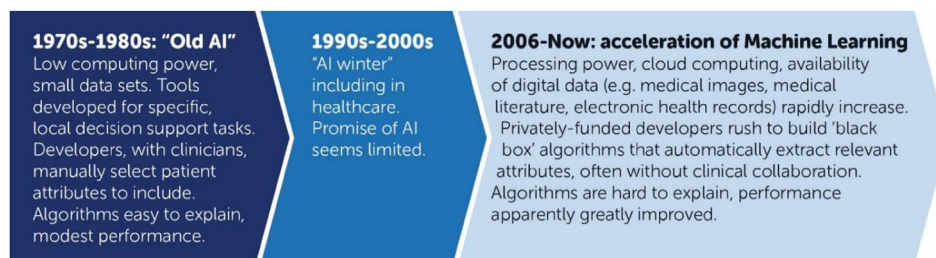

1970s-1980s: "Old AI"	1990s-2000s	2006-Now: acceleration of Machine Learning
Low computing power, small data sets. Tools developed for specific, local decision support tasks. Developers, with clinicians, manually select patient attributes to include. Algorithms easy to explain, modest performance.	"AI winter" including in healthcare. Promise of AI seems limited.	Processing power, cloud computing, availability of digital data (e.g. medical images, medical literature, electronic health records) rapidly increase. Privately-funded developers rush to build 'black box' algorithms that automatically extract relevant attributes, often without clinical collaboration. Algorithms are hard to explain, performance apparently greatly improved.

Fig. 1. The change from 'old AI' to 'new AI' (dates are approximate).

which to compare AI versus radiologists' performance [10].

There is also a burgeoning market in AI-based breast cancer detection outside of mainstream healthcare institutions. Start-ups developing AIs for breast screening have launched at least annually since 2008, using angel and seed capital investment, and in both high- and middle-income countries, especially India. We searched for these start-ups in August 2019.[1] Within the startups we identified, the most interesting pattern observed was a relationship between the technology offered and market targeted. Some startup companies are developing AI to improve mammography workflows and marketing these to health systems; they are sometimes collaborating with or being bought out by larger technology providers. A second group of companies is offering AI-based services using novel and unproven test technologies; they are more likely to be marketing direct to consumers, and this marketing often features women below the usual age recommended for mammography. The commercial nature of the healthcare AI market is important in considering ethical, legal and social implications, to which we now turn.

2. What are the ethical, legal and social implications?

In considering any technology, the potential risks and harms need to be evaluated against the potential benefits. In what follows we will consider issues including the values reflected in AI systems, their explainability, the need for a focus on clinical outcomes, the potential for bias, the problem of transferability, concerns regarding data, legal, moral and professional responsibility, the effect on patient experience, and possible explanations for the push to implementation. Fig. 2 summarises drivers, risks, solutions and desired outcomes.

2.1. Whose values should be reflected in algorithms, and can these be explained?

It is commonly argued that AI is 'value neutral': neither good nor bad in itself. In our view this is a potentially problematic way of thinking. Certainly, AI has the capacity to produce both good and bad outcomes. But every algorithm will encode values, either explicitly, or more commonly in the era of 'new AI', implicitly [21]. To give an example from breast screening: like any breast screening program, a deep learning algorithm may prioritise minimising false negatives over minimising false positives or perform differently depending on the characteristics of the breast tissue being imaged, or for women from different sociodemographic groups. Pre-AI, the performance of screening programs was a complex function of several factors, including the technology itself (e.g. what digital mammograms are capable of detecting) and collective human judgement (e.g. where programs set cut-offs for recall). Post new-AI, these same factors will be in play, but additional factors will be introduced, especially arising from the data on which AIs are trained, the way the algorithm works with that data, and the conscious or unconscious biases introduced by human coders.

The 'black box' problem in deep learning introduces a critical issue: *explainability* or *interpretability* [22—25]. If an algorithm is explainable, it is possible for a human to know how the algorithm is doing what it is doing, including what values it is encoding [26]. At present, less explainable algorithms seem to be more accurate, and

it is not clear whether accuracy and explainability must inevitably be traded off, or whether it is possible to have both. Of course, human clinicians' explanations for their own decisions are not perfect, but they are legally, morally and professionally responsible for these decisions, are generally able to provide some explanation, and can be required to do so. In contrast, future healthcare AIs may recommend individualised diagnostic, prognostic and management decisions that cannot be explained. This contrasts with older decision-support tools such as population-based risk calculators, which encode transparent formula and make general recommendations at a population level rather than specific recommendations for individual patients. Explainable medical AI is an active area for current research [23,27]. If AI is integrated into clinical practice, systems will need quality assurance processes—established well in advance—that can ensure understanding, review and maintenance of consistency in results. Before implementation there will also need to be clear expectations regarding explainability. We will consider the ethical implications of explainability in a later section on responsibility.

So far, we have argued that: 1) AI systems will inevitably encode values; and 2) these values may be difficult to discern. We now consider two central ways in which AIs are likely to encode values. The first relates to outcomes for patients; the second to the potential for systematic but undetected bias.

2.2. Effects on outcomes for patients and health systems

Evidence-based healthcare should require that new interventions or technologies improve outcomes by reducing morbidity and mortality, or deliver similar health outcomes more efficiently or cheaply, or both. Despite this, screening and test evaluation processes in the last half century have tended to implement new tests based on their characteristics and performance, rather than on evidence of improved outcomes [28]. At present, reporting of deep learning-based systems seems to be following the same path, focusing strongly on comparative accuracy (AI vs expert clinician) [11], with few clinical trials to measure effects on outcomes [22,25]. Noticing and acting on this now could help mitigate a repeat of problems that emerged in screening programs in the late 20th century, including belated recognition of the extent of overdiagnosis [29]. Combining imaging, pathological, genomic and clinical data may one day allow learning systems to discriminate between clinically significant and overdiagnosed breast cancers, effectively maximising the benefit of screening and minimising overdiagnosis. At present, however, AIs are being pointed towards the goal of more accurate detection, much like early mammographic systems [10]. Only deliberate design choices will alter this direction to ensure AI delivers more benefits than harms.

Creating an algorithm to address overdiagnosis will, however, be limited by the availability of training and validation data. Current human clinical capabilities do not, on the whole, allow identification of overdiagnosed breast cancers: population level studies can estimate the proportion of cases that are overdiagnosed, but clinicians generally cannot determine which individuals are overdiagnosed [30]. Autopsy studies provide one body of evidence of overdiagnosis, demonstrating a reservoir of asymptomatic disease in people who die of other causes. It may be possible to design studies that compare data from autopsy studies (including radiological, clinical and pathological and biomarker data to determine cancer subtype) with data from women who present clinically with breast cancer, and thus build an AI to distinguish between potentially fatal and non-fatal breast cancers. There are likely limitations to such a method, however, including that breast cancers found in autopsy studies may differ in important ways from cancers

---

[1] We searched Google on August 2nd, 2019, combining breast + screening + startup + AI; when we exhausted new webpages we pearled from the ones we had found. Note that we searched specifically for startups, not for the products being developed within existing mammography machine vendors, who are also working on AI platforms and solutions.
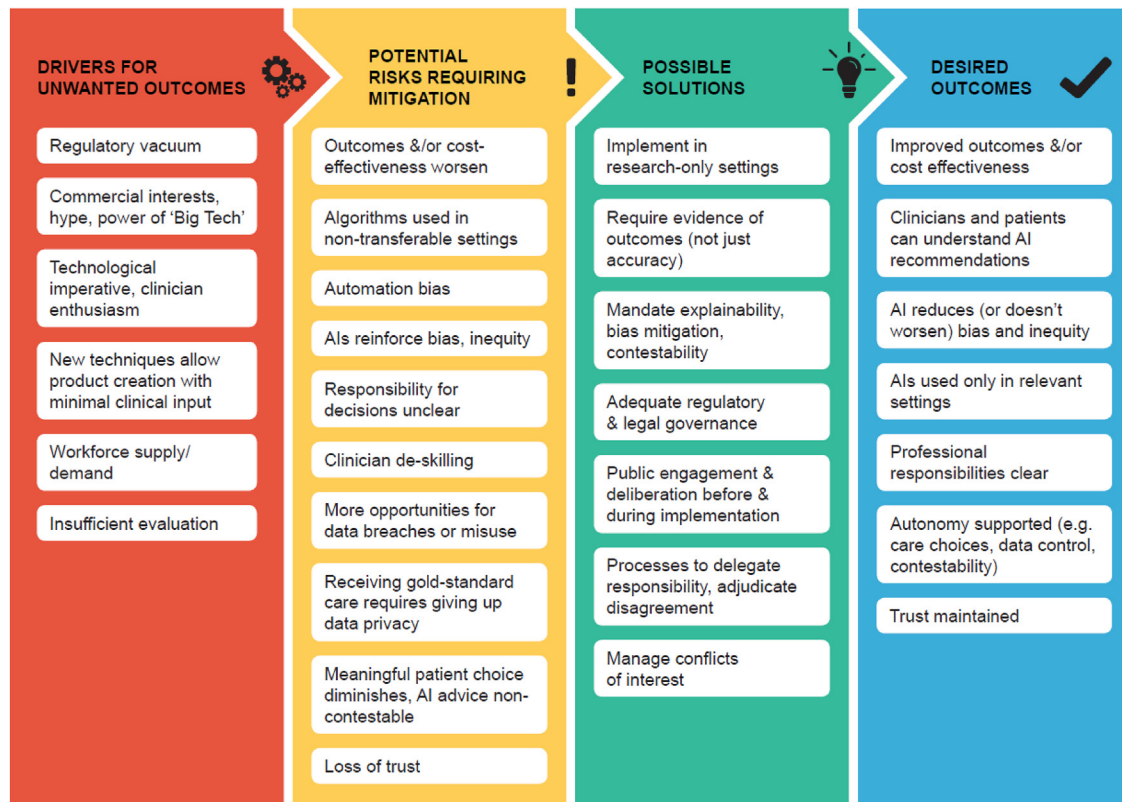
Fig. 2. Drivers, potential risks, possible solutions and desired outcomes.

overdiagnosed in screening participants. Watchful waiting trials of sufficiently low-risk cancer subtypes offer another possible strategy for building overdiagnosis-detecting AI. Heterogenous medical records from participants in such trials could conceivably be used as data to build deep learning algorithms. These algorithms might be able to identify patterns or characteristics that are imperceptible to human observers, and therefore not currently included in epidemiological studies. This could assist in building a more general AI to distinguish overdiagnosed cancers from cancers that will progress clinically. However, such an algorithm will not be produced using current research protocols, which overwhelmingly focus on increasing sensitivity and specificity according to current diagnostic criteria.

2.3. Bias and transferability in AI

A significant and well-observed problem for machine learning systems is bias in outcomes, arising from bias in training data or design choices [24,25,31,32]. Fundamentally, machine learning systems are 'made of' data. By exposure to massive datasets they develop the ability to identify patterns in those datasets, and to reproduce desired outcomes; these abilities are shaped not just by their coding, but also by the data they are fed. There is now extensive evidence from fields including higher education, finance, communications, policing and criminal sentencing that feeding biased data into machine learning systems produces systematically biased outputs from those systems; in addition, human choices can skew AI systems to work in discriminatory or exploitative ways [33]. It is already well-recognised that both healthcare and evidence-based medicine are biased against disadvantaged groups, not least because these groups are under-represented in the evidence base [34]. AI will inevitably reinforce this bias unless explicit

human choices are made to counter it. Ethical AI development programs are strongly focused on reducing bias. IBM Watson (a significant actor in healthcare AI) has launched several products designed to detect bias in data and algorithms [35]; a team from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) and Massachusetts General Hospital (MGH) have reported a deep learning mammogram algorithm specifically designed to perform equally well for American women of African and Caucasian descent [36].

In addition to problems of bias, transferability is a significant issue for AI; high-performing algorithms commonly fail when transferred to different settings [37]. An algorithm trained and tested in one environment will not automatically be applicable in another environment. If an algorithm trained in one setting or with one population group was to be used in another setting or group, this would require deliberate training on data from the new cohort and environment. Even with such training transferability is not inevitable, so AI systems require careful development, testing and evaluation in each new context before use in patient care. A high degree of transparency about data sources and continued strong distinction between efficacy and effectiveness will be necessary to protect against these potential weaknesses.

2.4. Data ownership, confidentiality and consent

Because AI systems require large quantities of good-quality data for training and validation, ownership and consent for use of that data, and its protection, are critical issues. This is complicated by the rapid movement of large technology companies such as IBM and Google into healthcare, as well as the proliferation of start-ups developing breast cancer-related products [25]. Some governments have made healthcare data available to developers: for example,

the Italian government released to IBM Watson the anonymised health records of all 61 million Italians, including genomic data, without individual consent from patients, and granted exclusive use rights to the company [25]. Such releases raise significant concerns regarding what economic or public goods should be required in exchange for provision of these very valuable data sets to private providers [38]. Expansion of data use also increases opportunities for data leakage and re-identification, as demonstrated by a range of ethical hacking exercises such as the Medicare Benefits Schedule Re-identification Event in Australia in 2016 [39]. The University of Chicago Medical Centre provided data to allow Google to build a predictive AI-powered electronic health record (EHR) environment; at time of writing both are being sued for alleged misuse of patient EHR data [40]. The case rests in part on Google's ability to re-identify EHR data by linking it with the vast individuated and geolocated datasets they already hold. Traditional models of opt-in or opt-out consent are extremely difficult to implement at this scale and in such a dynamic environment [41]. If algorithms enter individual clinical care and clinicians become dependent on these systems for decision-making, patients who do not share their health data may not be able to receive gold-standard treatment, creating a tension between consent and quality of care [42,43].

Addressing these data-related issues will be critical if the promise of healthcare AI is to be delivered in a justifiable and legitimate way. Some AI developers are experimenting with creative forms of data protection. For example, to enable the *DREAM Challenge* mentioned earlier, challenge participants were not given access to data but instead sent their code to the DREAM server using a Docker container, a self-contained package of software that can be transferred and run between computers. When run on the DREAM server, the container had access to the DREAM data, permitting all models to be trained and evaluated on the same imaging data. The competitors received their results but not any of the imaging or clinical data. Grappling with data protection and use in such a granular way will be a critical foundational step for any health system wanting to implement deep learning applications.

2.5. Legal risk and responsibility

Because AI is being introduced into healthcare exponentially, we currently have, as one author noted, 'no clear regulator, no clear trial process and no clear accountability trail' [44]. Scherer refers to this as a regulatory vacuum: virtually no courts have developed standards specifically addressing who should be held legally responsible if an AI causes harm, and the normally voluble world of legal scholarship has been notably silent [45]. One exception is in the area of data protection, especially protection of health information. General and specific regimes are emerging, including the *General Data Protection Regulation 2016/79* (GDPR) [46] in the European Union and *The Health Insurance Portability and Accountability Act (HIPAA)* [47] in the USA; in Australia personal health information is given general protection under the Australian Privacy Principles found in the *Privacy Act 1988* (Cth) [48]. Further work between stakeholders will be needed to continue to address privacy and data protection issues.

There is a clear 'regulatory gap' [49] or 'regulatory vacuum' [45] with respect to the clinical application of AI, and it is not clear how best to fill this gap. Existing frameworks (such as tort law, product liability law or privacy law) could be adapted, applying law that intervenes once harm has been done (looking backwards, '*ex poste*'). Alternatively, a new, 'purpose built' regulatory framework could be created, looking forward and setting up a set of guidelines and regulation to be followed ('*ex ante*'). It is beyond scope to consider this in detail: the crucial point is that meaningful debate is

needed, as well as care to ensure that legal developments reflect AI capabilities and the need to protect vulnerable patients. This will require a realistic identification and characterisation of the legal issues and a pro-active approach to creation of an appropriate framework. It will not be possible to 'retro-fit' the law once the technology has been developed; regulators and developers must co-operate to ensure that adequate protections are put in place, but also that appropriate innovation in AI technologies is supported rather than restricted by regulation.

2.6. Medical moral and professional responsibility

The debate about medical AI often features concern that human clinicians (particularly diagnosticians such as radiologists) will become redundant. Although some are strong advocates of this 'replacement' view of AI [50], others propose bringing together the strengths of both humans and machines, arguing that AIs may address workforce shortages (e.g. in public breast screening where radiologists are under-supplied [51]), or even that machines will free clinicians' time to allow them to better care for and communicate with patients [37]. Whichever of these is most accurate, future decision making is likely to involve both clinicians and AI systems in some way, requiring management of machine-human disagreement, and delegation of responsibility for decisions and errors [25,32].

AI, particularly non-explainable AI, potentially disrupts traditional conceptions of professional medical responsibility. If, for example, AI becomes able to perform reliable and cost-effective screening mammograms, a woman's mammogram might conceivably be read only by an AI-based system, and negative results could be automatically notified to her (e.g. through a secure online portal). This would be efficient for all concerned. However, it would also create cohorts of people undergoing medical testing independent of any human readers, undermining the traditional model in which people receiving interventions are patients of an identified medical practitioner who takes responsibility for that aspect of their care. It is not clear whether these women would be patients, and if so, whose. If doctors are directly involved in patient care, but decisions depend on non-explainable AI recommendations, doctors will face challenges regarding their moral and legal responsibility. They may be expected to take responsibility for decisions that they cannot control or explain; if they decline, it is not clear to whom responsibility should be delegated. It may be difficult to locate the decision-points in algorithm development that should trigger attributions of responsibility (and unlike healthcare professionals, developers are not required to put the interest of patients first). Shifting attributions of responsibility may affect patients' trust in clinicians and healthcare institutions and change medical roles. Clinicians may have more time to talk with patients, for example, but if decisions are algorithmically driven there may be limits to what clinicians are able to explain [31]. These observations are speculative but require careful consideration and research before moving to implementation.

In addition to raising questions about responsibility, AI has implications for human capacities. Routinisation of machine learning could change human abilities in a variety of ways. First, clinicians are likely to lose skills they do not regularly use [26], if they read fewer mammograms those skills will deteriorate. Conversely, workflow design using AI to triage out normal mammograms may mean that human readers see proportionally *more* breast cancer and become better at identifying cancer, but less familiar with variants of normal images. Second, automation bias means that humans tend to accept machine decisions, even when they are wrong [37,52,53]. Clinicians' diagnostic accuracy, for example, has been shown to decrease when they view inaccurately

machine-labelled imaging data [22]. Commentators have suggested over-automation should be resisted [24], or clinicians should be trained to avoid automation bias [52]; however cognitive biases are common in *current* healthcare [54], so this may not be possible.

2.7. Patient knowledge, experience, choice and trust

Although public engagement is seen as key to the legitimacy of and trust in AI-enabled healthcare [25,55], little is known about the values and attitudes of the public or clinicians towards AI. Recent US research found that general support for AI was higher amongst those who were wealthy, male, educated or experienced with technology; support for governance or regulation of AI was lower among younger people and those with computer science or engineering degrees. This work also showed low understanding of some future risks of AI uptake [56]. We are currently undertaking similar research in the Australian context. It could be argued that patients currently understand little about how health technologies (for example, mammographic screening systems) work, so perhaps also do not need to understand AI. However, given the implications of AI for medical roles and responsibilities, and the increasing public conversation regarding the dangers of AI in areas such as surveillance and warfare, it seems likely that sustaining trust in healthcare systems will require at least some public accountability about the use of AI in those systems.

Implementing AI may change the choices that are offered to patients in more or less predictable ways. For example, the range of treatment options may be shaped and narrowed by taking account not only of research evidence, but also personal health and non-health data, leaving the patient with a narrower range of options than currently exists. Choices will be further narrowed if funding (either through health insurance or government rebates) is limited to those options that are AI-approved. AI recommendations may replace current models of EBM and clinical guidelines, but risk replicating the ethical problems with these current approaches [57].

2.8. Explaining the pressure to implement healthcare AI

There does seem to be promise in AI for breast cancer care, particularly if it is able to improve the performance of screening (e.g. reducing overdiagnosis, personalising screening). However, there are also real risks. There is currently great momentum towards implementation, despite the risks: the social science literature helps explain why. In short, social, cultural and economic systems in healthcare create strong drivers for rapid uptake. Healthcare AI is subject to a *technological imperative,* where operating at the edge of technical capability tends to be seen as superior practice, technologically sophisticated interventions are preferred, and the fact that something can be done creates pressure for it to be done [58]. And healthcare AI—which is overwhelmingly proprietary and market-driven—shows all the hallmarks of *biocapital* [59,60], a highly speculative market in biotechnological innovation, in which productivity relies not on the delivery of concrete value so much as on selling a vision of the future. In this case, the vision being sold is of AI-powered technologies transforming the experience of healthcare provision and consumption, vastly increasing cancer control, and delivering massive investment opportunities [61]. Sociological evidence suggests these promises may not be realised; they are not truths, but powerful imaginings evoked to inspire consumers, practitioners and investors to commit in spite of the profound uncertainties that currently exist around AI [62] (It is worth noting also that these promises also contrast strongly with the measured, prospective and systematic culture of evidence based practice.). A long history of activism, fundraising and private

interests in breast cancer care has also created a large population of passionate consumers and advocates keen to fund and promote any novel development, creating a fertile seedbed for the big promises of tech companies [63]. Identifying these social conditions and seeing how they particularly apply in breast cancer care should give us pause in relation to hype regarding the promise of AI, and inspire healthy critique of the ambitious promises being made [11].

2.9. The future of AI in breast cancer care

The high level of activity in breast cancer care AI development creates both opportunities and responsibilities. Vanguard technologies can set standards for good practice but can also become infamous examples of failure and unintended consequences. What might it take for breast cancer AI to become an exemplar of excellent practice? Our central conclusion is this: *Health system administrators, clinicians and developers must acknowledge that AI-enabled breast cancer care is an ethical, legal and social challenge, not just a technical challenge.* Taking this seriously will require careful engagement with a range of health system stakeholders, including clinicians and patients, to develop AI collaboratively from outset to implementation and evaluation. We note that there is already a tradition of Value Sensitive Design in data and information science that could inform such engagement [64,65]. We make specific recommendations below.

Use of non-explainable AI should arguably be prohibited in healthcare, where medicolegal and ethical requirements to inform are already high. Building useful AI requires access to vast quantities of high-quality data; this data sharing creates significant opportunities for data breaches, harm, and failure to deliver public goods in return. These risks are exacerbated by a private market. Research suggests that publics are willing to support the use of medical data for technological development but only with careful attention to confidentiality, control, governance and assured use for public interest [66]: developers and data custodians should take heed.

It is critical that health systems and clinicians require AI providers to demonstrate explicitly what values are encoded in the development choices they have made, including the goals they have set for algorithms. Developers and health systems must learn from past mistakes in screening and test evaluation so as to avoid implementing algorithms based on accuracy data alone: data about outcomes should be required, even though this will slow implementation. Because transferability and effectiveness (as opposed to efficacy) are known problems for AI, no algorithm should be introduced directly into clinical practice. Instead AI interventions should be proven step-wise in research-only settings, following a model such as the IDEAL framework for evidence-based introduction of complex interventions [67]. Because bias is a known problem with AI, evaluation in breast cancer care AIs should pay close attention to the potential for systematically different effects for different cohorts of women. The first opportunity for AIs to take on a task currently performed by humans is likely to be reading large volumes of images, such as digital mammograms or digitised histopathology slides. This provides an opportunity to augment rather than replace human function: for example, initially adding AI to existing human readers rather than replacing readers and revealing AI decisions only after human decisions have been made. This would allow implementation design to minimise automation bias and deskilling and could act as a relatively low-risk 'experiment' in tracking the potential harms, benefits and 'unknowns'.

Significant conflicts of interest—for both clinicians and proprietary developers— have the potential to skew AI use in several ways. On the clinical side, AI may be manipulated or regulated—for example, by professional bodies—to mandate human medical

involvement to protect incomes and status, even though it may eventually be more effective and cost efficient for both patients and health services for the AI to act alone. On the market side, current financial interests in AI are hyping potential benefits, and risk tying radiology infrastructure to particular proprietary algorithms, creating future intractable conflicts of interest. This is nothing new: commercial interests have always been a part of the breast cancer care landscape, providing pharmaceuticals, diagnostic technologies and medical devices; such involvement has not always worked in patients' favour. The commercialisation of diagnostic or treatment decisions (as opposed to the provision of data to inform those decisions) is a different order of commercial interest, one to which health systems should not easily or unreflectively acquiesce.

A cursory reader of the current breast cancer AI literature might reasonably conclude that the introduction of AI is imminent and inevitable, and that it will radically improve experience and outcomes for clinicians, patients and health systems. While this may prove to be true, our aim here has been to introduce caution by examining the ELSI of healthcare AI, and to ask what might make the introduction of AI, on-balance, a good rather than a bad thing in breast cancer care. Once AI becomes institutionalised in systems it may be difficult to reverse its use and consequences: due diligence is required before rather than after implementation. A strong and proactive role for government, regulators and professional groups will help ensure that AI is introduced in robust research contexts and that a sound evidence base is developed regarding real-world effectiveness. Rather than simply accepting what is offered, detailed public discussion about the acceptability of different options is required. Such measures will help to optimise outcomes for health systems, professionals, society, and women receiving care.

Declarations of interest

Prof. Carter reports grants from National Health and Medical Research Council (NHMRC) Centre of Research Excellence 1104136 and a grant from the University of Wollongong Global Challenges Scheme, outside the submitted work.

Prof. Rogers reports funding from the CIFAR in association with UK Research and Innovation and French National Centre for Scientific Research (CNRS) outside the submitted work. She is also a co-chair of the program on Safety, Quality and Ethics of the Australian Alliance on AI and Health Care.

Prof. Houssami reports grants from a National Breast Cancer Foundation (NBCF Australia) Breast Cancer Research Leadership Fellowship, outside the submitted work.

A/Prof. Frazer reports affiliations with BreastScreen Victoria Radiology Quality Group, State Quality Committee, and Research Committee, BreastScreen Australia Tomosynthesis Technical Standards Working Group, Cancer Australia Expert Working Group for Management of Early Breast Cancer, Royal Australian and New Zealand College of Radiologists Breast Imaging Advisory Committee, and Convenor of Breast Interest Group (BIG) scientific meetings, Medicare Benefits Scheme Breast Imaging Review Committee, Sydney University Faculty of Health Science Adjunct Associate Professor, Honorary Clinical Director of BREAST.

Associate Professor Bernadette Richards and Associate Professor Khin Than Win have no interests to declare.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Digital Mammography DREAM Challenge. Sage bionetworks [cited 2019 5th Feb]. Available from: https://www.synapse.org/#!Synapse:syn4224222/wiki/401745. [Accessed 26 August 2019].
[2] IBM and sage bionetworks announce winners of first phase of DREAM digital mammography challenge [updated June 2], https://sagebionetworks.org/in-the-news/ibm-and-sage-bionetworks-announce-winners-of-first-phase-of-dream-digital-mammography-challenge/2017. [Accessed 26 August 2019].
[3] No author. Cancer research UK imperial Centre, OPTIMAM and the jikei university hospital. DeepMind Technologies Limited; 2019. https://deepmind.com/applied/deepmind-health/working-partners/health-research-tomorrow/cancer-research-imperial-optimam-jikei/. [Accessed 26 August 2019].
[4] Berggruen Institute. Keynote: AI Ethics and governance. 2019 June 20. https://www.berggruen.org/activity/global-artificial-intelligence-technology-2019-conference-day-1-keynote-ai-ethics-and-governance/. [Accessed 26 August 2019].
[5] Tegmark M. Life 3.0. UK: Penguin Random House; 2017.
[6] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006;313(5786):504–7. https://doi.org/10.1126/science.1127647.
[7] Heathfield H, Kirkham N. A cooperative approach to decision support in the differential diagnosis of breast disease. Med Inf 1992;17(1):21–33.
[8] Heathfield HA, Winstanley G, Kirkham N. Computer-assisted breast cancer grading. J Biomed Eng 1988;10(5):379–86.
[9] Leaning MS, Ng KE, Cramp DG. Decision support for patient management in oncology. Med Info 1992;17(1):35–46.
[10] Houssami N, Kirkpatrick-Jones G, Noguchi N, Lee CI. Artificial Intelligence (AI) for the early detection of breast cancer: a scoping review to assess AI's potential in breast screening practice. Expert Rev Med Devices 2019;16(5):351–62. https://doi.org/10.1080/17434440.2019.1610387.
[11] Houssami N, Lee CI, Buist DSM, Tao D. Artificial Intelligence for breast cancer screening: opportunity or hype? Breast 2017;36:31–3. https://doi.org/10.1016/j.breast.2017.09.003.
[12] van der Gijp A, Ravesloot CJ, Jarodzka H, van der Schaaf MF, van der Schaaf IC, van Schaik JPJ, et al. How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. Adv Health Sci Educ Theor Pract 2017;22(3):765–87. https://doi.org/10.1007/s10459-016-9698-1.
[13] Gur D, Bandos AI, Cohen CS, Hakim CM, Hardesty LA, Ganott MA, et al. The "laboratory" effect: comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. Radiology 2008;249(1):47–53. https://doi.org/10.1148/radiol.2491072025.
[14] Jiang X, Wells A, Brufsky A, Neapolitan R. A clinical decision support system learned from data to personalize treatment recommendations towards preventing breast cancer metastasis. PLoS One 2019;14(3):e0213292. https://doi.org/10.1371/journal.pone.0213292.
[15] Nindrea RD, Aryandono T, Lazuardi L, Dwiprahasto I. Diagnostic accuracy of different machine learning algorithms for breast cancer risk calculation: a meta-analysis. Asian Pac J Cancer Prev APJCP 2018;19(7):1747–52.
[16] Somashekhar SP, Sepúlveda MJ, Puglielli S, Norden AD, Shortliffe EH, Rohit Kumar C, et al. Watson for oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board. Ann Oncol 2018;29(2):418–23. https://doi.org/10.1093/annonc/mdx781.
[17] Sherbet GV, Woo WL, Dlay S. Application of artificial intelligence-based technology in cancer management: a commentary on the deployment of artificial neural networks. Anticancer Res 2018;38(12):6607–13. https://doi.org/10.21873/anticanres.13027.
[18] Low SK, Zembutsu H, Nakamura Y. Breast cancer: the translation of big genomic data to cancer precision medicine. Cancer Sci 2018;109(3):497–506. https://doi.org/10.1111/cas.13463.
[19] Xu J, Yang P, Xue S, Sharma B, Sanchez-Martin M, Wang F, et al. Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. Hum Genet 2019;138(2):109–24. https://doi.org/10.1007/s00439-019-01970-5.
[20] Shamai G, Binenbaum Y, Slossberg R, Duek I, Gil Z, Kimmel R. Artificial intelligence algorithms to assess hormonal status from tissue microarrays in patients with breast cancer. JAMA Netw Open 2019;2(7):e197700. https://doi.org/10.1001/jamanetworkopen.
[21] Carter SM. Valuing healthcare improvement: implicit norms, explicit normativity, and human agency. Health Care Anal 2018;26(2):189–205. https://doi.org/10.1007/s10728-017-0350-x.
[22] Cabitza F, Rasonini R, Gensini GF. Unintended consequences of machine learning in medicine? J Am Med Assoc 2017;318(6):517–8. https://doi.org/

10.1001/jama.2017.7797.

[23] Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain?. 2017. arXiv:1712.09923 [cs.AI].

[24] Loder J, Nicholas L. Confronting Dr. Robot: creating a people-powered future for AI in health. London: NESTA Health Lab; 2018.

[25] Fenech M, Strukelj N, Buston O. Ethical, social, and political challenges of artificial intelligence in health. United kingdom. Future advocacy. Welcome Trust; 2018.

[26] Cowls J, Floridi L. Prolegomena to a white paper on an ethical framework for a good AI society. SSRN; 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3198732. [Accessed 26 August 2019].

[27] No author. Call for papers — explainable AI in medical informatics & decision making. https://human-centered.ai/special-issue-explainable-ai-medical-informatics-decision-making/:Springer/NatureBMCMedicalInformaticsandDecisionMaking. [Accessed 26 August 2019].

[28] Schünemann HJ, Mustafa RA, Brozek J, Santesso N, Bossuyt PM, Steingart KR, et al. GRADE guidelines: 22. The GRADE approach for tests and strategies—-from test accuracy to patient-important outcomes and recommendations. J Clin Epidemiol 2019;111:69—82. https://doi.org/10.1016/j.jclinepi.2019.02.003.

[29] Houssami N. Overdiagnosis of breast cancer in population screening: does it make breast screening worthless? Cancer Biol Med 2017;14:1—8. https://doi.org/10.20892/j.issn.2095-3941.

[30] Carter SM, Degeling C, Doust J, Barratt A. A definition and ethical evaluation of overdiagnosis. J Med Ethics 2016;42(11):705—14. https://doi.org/10.1136/medethics-2015-102928.

[31] Coiera E. On algorithms, machines, and medicine. Lancet Oncol 2019;20(2): 166—7. https://doi.org/10.1016/S1470-2045(18)30835-0.

[32] Taddeo M, Floridi L. How AI can be a force for good. Science (New York, NY) 2018;361(6404):751—2.

[33] O'Neil C. Weapons of math destruction. London: Penguin Random House; 2016.

[34] Rogers WA. Evidence based medicine and justice: a framework for looking at the impact of EBM upon vulnerable or disadvantaged groups. J Med Ethics 2004;30(2):141—5.

[35] IBM Research. AI and bias. IBM; 2019. https://www.research.ibm.com/5-in-5/ai-and-bias/.

[36] Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. Radiology 2019;292(1):60—6. https://doi.org/10.1148/radiol.2019182716.

[37] Topol E. Deep Medicine: how artificial intelligence can make healthcare human again. New York: Hachette; 2019.

[38] Leshem E. IBM Watson Health AI gets access to full health data of 61m Italians. 2018. https://medium.com/@qData/ibm-watson-health-ai-gets-access-to-full-health-data-of-61m-italians-73f85d90f9c0:Medium. [Accessed 26 August 2019].

[39] Culnane C, Rubinstein BIP, Teague V. Health data in an open world. arXiv [Internet]. 2017 Dec 5. Available from: https://arxiv.org/abs/1712.05627. [Accessed 26 August 2019].

[40] Cohen IG, Mello MM. Big data, big tech, and protecting patient privacy. J Am Med Assoc 2019. https://doi.org/10.1001/jama.2019.11365.

[41] Racine E, Boehlen W, Sample M. Healthcare uses of artificial intelligence: challenges and opportunities for growth. Healthcare Management Forum; 2019.

[42] Azencott CA. Machine learning and genomics: precision medicine versus patient privacy. Philosl Trans A Math Phys Eng Sci 2018;376(2128). https://doi.org/10.1098/rsta.2017.0350. pii: 20170350.

[43] Char DS, Shah NH, Magnus D. Implementing machine learning in health care-addressing ethical challenges. N Engl J Med 2018;378(11):981—3. https://doi.org/10.1056/NEJMp1714229.

[44] McCartney M. AI in medicine must be rigorously tested. BMJ 2018;361.

[45] Scherer MU. Regulating Artificial Intelligence systems: risks, challenges, competencies, and strategies. Harv J Law Technol 2015;29(2):353—401. https://doi.org/10.2139/ssrn.2609777.

[46] Regulation (EU. The European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. General Data Protection Regulation; 2016/679 of.

[47] United States. The health insurance portability and accountability act (HIPAA). Washington, D.C.: U.S. Pub. L; 1996. p. 104—91. Stat. 1936. Web.

[48] Privacy act. 1988 (cth) (austl.).

[49] Gunkel DJ. Mind the gap: responsible robotics and the problem of responsibility. Ethics and Information Technology; 2017. p. 1—14. https://doi.org/10.1007/s10676-017-9428-2.

[50] Hinton G. Machine learning and the market for Intelligence2016. Available from: https://www.youtube.com/watch?v=2HMPRXstSvQ. [Accessed 26 August 2019].

[51] van den Biggelaar FJ, Nelemans PJ, Flobbe K. Performance of radiographers in mammogram interpretation: a systematic review. Breast 2008;17(1):85—90.

[52] Coiera E. The fate of medicine in the time of AI. Lancet 2018;392(10162): 2331—2. https://doi.org/10.1016/S0140-6736(18)31925-1.

[53] Gretton C. The dangers of AI in health care: risk homeostatsis and automation bias. Data Sci 2017 31/01/2019. Available from: https://towardsdatascience.com/the-dangers-of-ai-in-health-care-risk-homeostasis-and-automation-bias-148477a9080f. [Accessed 26 August 2019].

[54] Scott IA, Soon J, Elshaug AG, Lindner R. Countering cognitive biases in minimising low value care. Med J Aust 2017;206(9):407—11. https://doi.org/10.5694/mja16.00999.

[55] Ensuring data and AI work for people and society. London: The Ada Lovelace Institute; 2018.

[56] Zhang B, Dafoe A. Artificial intelligence: American attitudes and trends. Oxford, UK: Center for the Governance of AI. Future of Humanity Institute. University of Oxford; 2019.

[57] Rogers WA. Are guidelines ethical? Some considerations for general practice. Br J Gen Pract 2002;52(481):663—8.

[58] Burger-Lux MJ, Heaney RP. For better and worse: the technological imperative in health care. Soc Sci Med 1986;22(12):1313—20.

[59] Helmreich S, Labruto N. Species of biocapital, 2008, and speciating biocapital, 2017. In: Meloni M, Cromby J, Fitzgerald D, Lloyd S, editors. The palgrave handbook of biology and society. London: Palgrave Macmillan UK; 2018. p. 851—76.

[60] Rajan KS. Biocapital: the constitution of postgenomic life. Durham: Duke University Press; 2006.

[61] CBInsights. Research briefs: 12 startups fighting cancer with artificial intelligence [updated Sept 15], https://www.cbinsights.com/research/ai-startups-fighting-cancer/2016. [Accessed 26 August 2019].

[62] Beckert J. Imagined futures : fictional expectations and capitalist dynamics. Cambridge, Mass: Harvard University Press; 2016.

[63] King S. Pink ribbons, inc. Minneapolis: University of Minnesota Press; 2006.

[64] Friedman B, Peter H, Kahn J, Borning A. Value sensitive design and information systems. In: Zhang P, Galletta D, editors. Human-computer interaction and management information systems: foundations. Armonk: Routledge; 2006. p. 348—72.

[65] Yetim F. Bringing discourse Ethics to value sensitive design: pathways toward a deliberative future. AIS Trans Hum-Comput Interact 2011;3(2):133—55.

[66] Aitken M, de St, Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med Ethics 2016;17(1):73.

[67] McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, et al. No surgical innovation without evaluation: the IDEAL recommendations. Lancet 2009;374(9695):1105—12. https://doi.org/10.1016/S0140-6736(09)61116-8.