AI & SOCIETY (2021) 36:535–545
https://doi.org/10.1007/s00146-020-01019-6

OPEN FORUM

Clinical AI: opacity, accountability, responsibility and liability

Helen Smith[1]

Received: 29 November 2019 / Accepted: 8 July 2020 / Published online: 25 July 2020

Abstract
The aim of this literature review was to compose a narrative review supported by a systematic approach to critically identify and examine concerns about accountability and the allocation of responsibility and legal liability as applied to the clinician and the technologist as applied the use of opaque AI-powered systems in clinical decision making. This review questions (a) if it is permissible for a clinician to use an opaque AI system (AIS) in clinical decision making and (b) if a patient was harmed as a result of using a clinician using an AIS's suggestion, how would responsibility and legal liability be allocated? Literature was systematically searched, retrieved, and reviewed from nine databases, which also included items from three clinical professional regulators, as well as relevant grey literature from governmental and non-governmental organisations. This literature was subjected to inclusion/exclusion criteria; those items found relevant to this review underwent data extraction. This review found that there are multiple concerns about opacity, accountability, responsibility and liability when considering the stakeholders of technologists and clinicians in the creation and use of AIS in clinical decision making. Accountability is challenged when the AIS used is opaque, and allocation of responsibility is somewhat unclear. Legal analysis would help stakeholders to understand their obligations and prepare should an undesirable scenario of patient harm eventuate when AIS were used.

Keywords Artificial intelligence · Clinical decision making · Law · Ethics · Responsibility · Accountability

1 Introduction

Computer systems use algorithms; a set of rules which dictate their behaviour (Weizenbaum 1976). The term artificial intelligence is not widely defined (House of Lords: Select Committee on Artificial Intelligence 2018) but it can be used to identify the element of a computer system which takes information, processes it and dispenses an output.

The process by which an artificially intelligent system (AIS) determines its output can be obscured; i.e., it can be difficult to achieve meaningful scrutiny of the reasoning for an AIS's output when modern computing methods are used (Knight 2017). This makes the process by which an AIS makes its outputs comparable to a black box; the process is 'opaque' (Fenech et al. 2018). Opacity is a relative concept rather than absolute; a process used by an AIS may be so complex that it is effectively obscured to a non-technically trained clinical user, whilst remaining simple to understand to a technologist who is proficient in that area of computer science. A clinician may be additionally skilled in the design and use of AIS in the clinical environment, but this is currently not a required professional standard.

There has been the development of AISs which are designed to support clinical users which might thus directly influence clinical decision making; for example, IBM's Watson for Oncology. Here, the AIS would accept information about the clinician's desired patient, would process that information and then make a recommendation for the patient's care (Ross and Swetlitz 2017). The future scenario of using AIS to aid clinicians could offer a forthcoming step-change in clinical decision-making activities and, if adopted, will potentially create new dynamics between stakeholders. Clinicians ('clinicians' include, but are not limited to, doctors, nurses and other professions allied to health) are confronted with incorporating emerging technology in the clinical environment (Hancock 2018), thus have a vested interest in its safe roll-out.

Historically, healthcare has largely been focussed on the relationship of the patient and the clinician (with variable

✉ Helen Smith
helen.smith@bristol.ac.uk

[1] Centre for Ethics in Medicine, University of Bristol, Bristol, UK

involvement of peripheral personnel attached to the clinical area). The clinician's recommendation is discussed with the patient and, through the shared decision making and the use of informed consent, a plan of care is devised which aims to balance the patient's preferences and needs (NICE 2012). Decisions on clinical recommendations are made by clinicians based on their evaluation of information which they had gathered about the patient; they are solely responsible for their own clinical decision making. Computer systems may have previously assisted in the administration of healthcare, but their outputs have not directly and purposefully advised human clinical decision-making at the point of care. The introduction of AIS has the potential to alter relationships in the clinical environment (Fenech et al. 2018, p25). If AIS such as IBM's Watson for Oncology are adopted, the technologist who designed and deployed the system in question would be joining the clinician in the decision making space for the first time.

The clinician and the technologist are inextricably linked when AIS are used in clinical decision making; without the clinician the technologist's AIS cannot reach the patient, without the technologist there is no AIS to offer the clinician to use in their clinical decision making. It is the relationship between these two stakeholders which this review is chiefly concerned with.

Using an emerging technology raises concerns on how it can be a new source of error (Fenech et al. 2018). Mistakes in the high-risk area of healthcare might lead to significant consequences for the patient affected (Harwich and Laycock 2018); this is important to consider as patients encounter clinicians at times in their lives when they are potentially at their most vulnerable (NMC 2018). Conscientious clinicians are aware of how clinical errors can increase the potential for an increase in morbidity and mortality (Makary and Daniel 2016); thus, making it reasonable that clinicians should oversee all applications of AIS use. But if clinicians are to be presented with an opaque AIS to aid their decision making, they might not understand how the AIS has made its recommendation for patient care.

Concern for AISs being a new source of error and the nature of opacity creates novel problems when applied to the clinical environment. Mukherjee (2017) outlines the scenario of a clinician having no idea how an opaque AIS's answer is created when they asked it a question and, as identified by Char et al. (2018), those who create the AISs for use are unlikely to be at the patient's bedside with the clinician.

This raises two pertinent points. First, that the clinician cannot fully account (i.e., be accountable) for the AIS output that they are being offered for the patient. Second, that the technologist is a novel actor in the clinical environment who is not physically present at the point of clinical decision making. Under these conditions it is reasonable to question, (a) if it is permissible for a clinician to use an opaque AIS in clinical decision making if there is an additional risk of error and (b) if a patient was harmed as a result of a clinician using an AIS's suggestion, how would responsibility and legal liability be allocated?

When considering these questions, it is worth noting that accountability, when identified as a person's explanation and justification for their intentions and beliefs about their behaviour (Dignum 2019; Oshana 2004), differs from responsibility. A person's account of their actions is linked to responsible behaviour, which is characterised by the "common norms which govern conduct" (Oshana 2004, pp.257). If one cannot rationally account for their behaviour in accordance with the accepted norms, any claim that their actions are responsible could be open to challenge; thus one of the ways that responsible actors demonstrate their responsibility is by being accountable.

2 Aims

The aim of this literature review was to explore concerns about the use of opaque AIS in clinical decision making. The issues of accountability, and the allocation of responsibility and legal liability as applied to the clinician and the technologist were examined. This review employed a narrative review supported by a systematic approach.

3 Methods

Employing a systematically inspired strategy to select and review the literature aids data capture (Khan et al. 2003). The expectation was to find non-homogenous materials in the literature searches, thus careful selection of the type of review process was needed which would accommodate this. No single theoretical framework proved ideal; Braun and Clarke (2006) identify this as an issue in the selection of research methodology and recommend that "the theoretical framework and methods match what the researcher wants to know". Thus, I have adopted Strech and Sofaer's (2012) four-step model of systematic reviews and adapted it to incorporate the concept of Braun and Clarke's (2006) use of themes in steps three and four for the purposes of this review as outlined below.

3.1 Step 1: Formulate the review question and eligibility criteria

The literature review aimed to answer the questions "Is it permissible for a clinician to use an opaque AI system in clinical decision making?" and "What concerns are there about opacity, accountability, responsibility and legal liability when considering the stakeholders of technologists and clinicians in the creation and use of AI systems in clinical decision making?".

To aid the selection of items to include in this review, an inclusion/exclusion criteria specific to the literature review's aims was used to determine the eligibility of materials to be considered for review. This helped the author to identify items with relevant arguments and argument themes whilst checking for flaws, credibility, contribution, relevance and coherence in each item selected for inclusion to this literature review. Each item selected from search results must be formally reviewed to ensure quality.

The applied inclusion/exclusion criteria were as follows: That the items will have content pertaining to ethical and legal issues in applications of AIS in clinical decision making as it relates to opacity, accountability, responsibility and liability in healthcare. Inclusion was limited to those items generated in the past 10 years to aid relevance, but more weighting was given to the value of the item's contribution, rather than its age. Items from all areas of clinical practice where AIS can be applied were considered (i.e., inclusive of all fields of medicine, surgery, paediatrics, adult, mental health etc.). Literature had to be presented in English. Items which discuss legal theory were limited specifically to the context of the law of England and Wales (else this review would have become unwieldy with international comparative examples). The literature search found diverse materials in a multitude of formats such as journal articles, books, opinion pieces, reports, editorials, items of discussion and analysis; each item was judged on its strengths. It is impossible to exhaustively identify every way that items were considered weak, but as a guide, those with incoherent, invalid, weak or fatally flawed arguments were excluded.

3.2 Step 2: Identify all of the literature that meets the eligibility criteria

The following search string was applied to all databases:

"Artificial intelligence" AND (liability OR responsibility OR accountability OR transparency OR opacity) AND (ethic* OR law) AND (healthcare OR clinical OR medical).

The author performed searches using this search string in nine relevant databases; these searches were performed in February 2018. The databases chosen were from a spread of disciplines, not just limited to healthcare, law, and ethics but also to computing and general scientific sources. Hand searches were additionally performed on the websites of organisations who collectively regulate clinical professionals: The General Medical Council, the Nursing and Midwifery Council, and the Health and Care Professions Council. Relevant grey literature originating from governmental and non-governmental organisations which had been found outside of the searches and come to the attention of the author during the period of composing the review were included for consideration alongside the formal search results.

This approach generated 185 non-duplicate citations. The author screened each title with the inclusion/exclusion criteria to decide the relevance of each item. Items which passed title screening proceeded to abstract screening and were again subjected to the inclusion/exclusion criteria. In total, 36 items passed title and abstract screening; these 36 items were then subjected to full-text screening by being read fully and the inclusion/exclusion criteria applied again. Ten items were excluded after full-text screening, which left 26 articles. Information relevant to this review was identified in each item, and each item's contribution was assessed for quality and relevance before the item was included in this review.

Using a PRISMA diagram, (Fig. 1) shows the show the databases used, the number of items identified by each database search and how each item outside of the inclusion criteria was excluded from the final collection of literature for synthesis and analysis.

3.3 Step 3: Extract and synthesise data by the allocation of pertinent points to theme headings

The EndNote Online reference management system was used to capture the citations from the searches from each database. From the EndNote hosted catalogue, results were screened, and irrelevant items removed as per the inclusion/exclusion criteria. An independent second review of 10% of the search results was performed by an academic who was external to this review to ensure the robustness and reliability of the selection process (as exemplified by Kyte et al. 2013). This process yielded 26 items of literature is included in this literature review.

Data extraction was performed on these 26 items using Strech and Sofaer's (2012) method of extraction and coding of data; this technique's strength lies in its promotion of the identification of ethical analysis and argument within an item's content. Relevant arguments and argument themes were identified whilst checking for flaws, credibility, contribution, relevance and coherence in each item selected for inclusion to this literature review.

The concept of Braun and Clarke's (2006) 'themes' were adopted so that any additional themes found in the literature during the data extraction process could be flexibly considered for addition in the review's findings. Using themes allowed the additional areas of 'why accountability was important' as well as 'why opacity interferes with accountability' to be recognised and explored alongside the initial core topics of opacity, accountability, responsibility and liability.

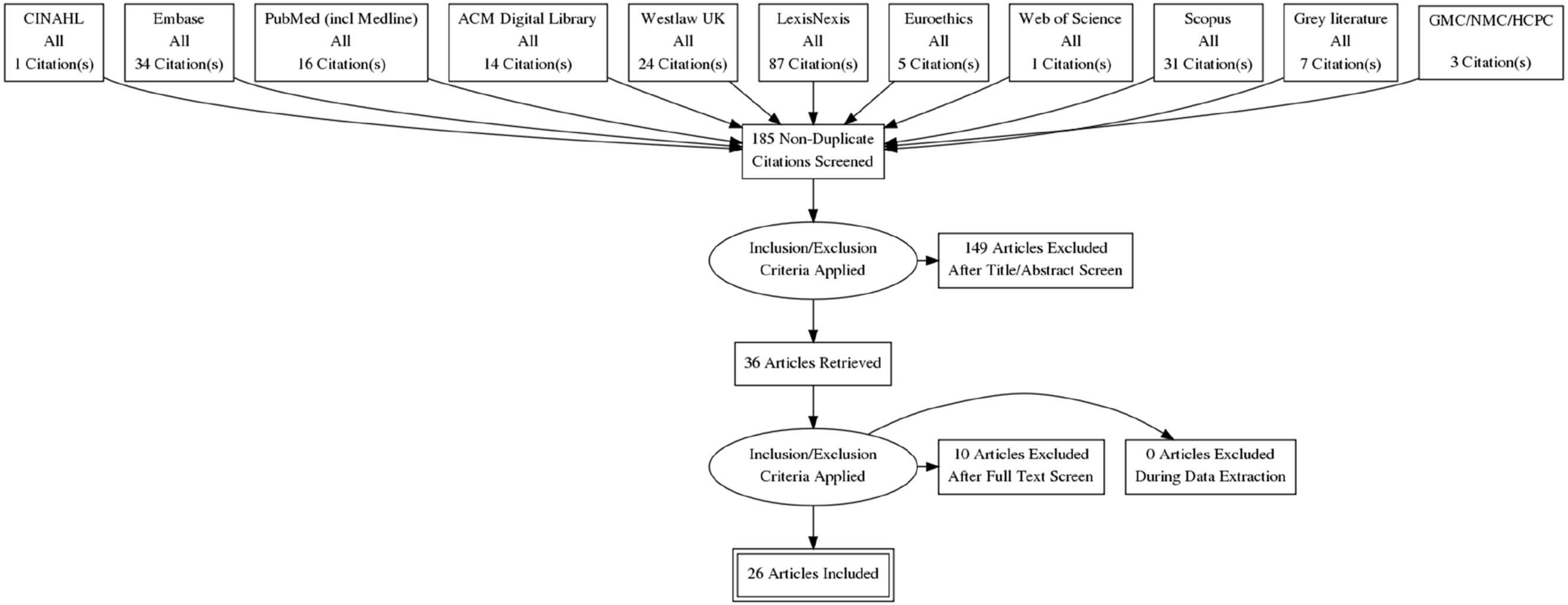

Figure 1 Box 1 PRISMA diagram

3.4 Step 4: Derive and present results organised by themes

The following findings resulted from the above careful searches and selection process. The findings have been structured as per the themes identified in the research question. Corralling of themes enabled stratification of information, thus aiding the analysis and critique of the literature when identifying concerns of AI use in clinical decision making.

4 Findings

As per step 4, the 26 items selected for this review were examined for content related to concerns related to the research question's key themes of AIS opacity, accountability, responsibility, and legal liability regarding the clinical use of AIS in decision making.

The following is the key findings and the high-level literature synthesis generated from that identified in the review.

Regarding accountability, clinicians have a regulatorily enforced professional requirement to be able to account for their actions, whereas technologists do not; instead, ethical codes of practice are employed in this sector. This comparison synthesises the question asking if technologists should also be regulated if their AIS is to be deployed in the clinical environment and directly affect patients.

Regarding opacity, clinicians will be challenged on issues of safety and accountability when using AIS's which do not explain their outputs. If a clinician cannot account for the output of the AIS they are using, they cannot fully account for their actions if they choose to use that output. This lack of accountability raises the potential safety issue of using unverified or unvalidated AISs in the clinical environment. Opacity is not a problem limited only to clinicians; it can also affect technologists. To recognise this, scenarios which encompassed how opacity can affect each stakeholder are detailed.

Regarding responsibility, there is a lack of formal clarification regarding who is responsible for the outcomes of AIS use. There is an agreement that one should take responsibility for their actions when choosing to use an AIS; this included evaluating the AIS's outputs before using them in the clinical context. Technologists were found to be responsible for the accuracy of their systems, but the literature pushed back against the idea of technologists holding any responsibility for the effect that their AIS would have in the clinical environment; this was justified via their devices assisting the clinician rather than replacing the clinician. The potential for responsibility to be shared in the future was mentioned and a possible retrospective approach identified to determine the allocation of responsibility to shareholders as per an analysis of each given incident. The literature agrees that an AIS should not be responsible for itself but may carry out tasks if appropriately supervised.

Regarding liability, the literature has not predicted the outcomes of negligence and liability in this area due to lacking a body of case law in this area.

This review's findings are expounded upon in the following discussion, where each theme is addressed in turn and progressively links one theme to the next.

4.1 Why is accountability important?

Professionalism is the vehicle which formalises the notion of trust within organisational structures which gathers those with similar skill sets together. By cohorting these skilled persons, standardisation of desirable behaviours can be achieved which serve to promote trust within that professional group (NMC 2018). Codes of conduct are created by the statutory bodies who oversee their respective healthcare professional groups. In the UK, the General Medical Council (GMC), the Nursing and Midwifery Council (NMC) and the Health and Care Professional Council (HCPC) cover a significant number of practicing clinical professionals; these shall be the three bodies I call upon to exemplify codes of conduct and professional issues.

Accountability from clinicians is required by the GMC (2013), NMC (2018), and HCPC (2016) codes of conduct. GMC (2013) and HCPC (2016) codes of conduct specifically require that the clinician must be able to justify once own decisions, the NMC (2018) stipulates that a Registered Nurse should be able to fully explain all aspects of a patient's care. The existence and enforcement of these codes result in the clinician's requirement to provide good care with an emphasis on safety. Breach of these codes of conduct would lead to the clinician being exposed to sanctions from their professional regulator, for example, the clinician is prevented from practicing.

Interestingly, despite it being a requirement in professional clinical practice, the literature searches failed to yield a unified definition of accountability and, therefore, I have developed my own definition for the purposes of this literature review which encompasses the spirit of that aimed for by the governing bodies; accountability is when an individual is obliged to explain (account) to those who are entitled to ask (e.g., regulators, a patient) for their decision-making process which guided their actions or omissions.

Hengstler et al (2016, p.106) identify trust as "the willingness to be vulnerable to the actions of another person". Given that the patient is already vulnerable due to the nature of their ailment and that a clinician may have to do harm to create the conditions whereby the patient may heal (e.g., a surgical incision whilst under general anaesthesia), trust is, logically, both a relevant and necessary quality which the patient will need if they are to be comfortable to approach a clinician for help and for them to tolerate the treatment pathway under that clinician's care. Armstrong (2018) describes how even when a clinician may be uncertain about their decision making, the act of communicating and expressing that uncertainty can lead to increased trust from their patient rather than the loss of their confidence. It is reasonable to deduce that if a clinician communicates their uncertainty to their patients, they are acting in an accountable manner; thus, accountability and patient trust are linked.

The historical background of professional cultural carefulness in the clinical professions does not appear to be shared in the field of computer science (Whitby 2015). This was exemplified by Lanfear (evidence to House of Lords: Select Committee on Artificial Intelligence 2018, p.122) who was unable to describe how his artificial intelligence company, Nvidia, was ensuring compliance of their own corporate ethical principles: "as a technologist it is not my core thinking". Whitby (2015), p.227 identifies a lack of compulsory professional standards or formal qualifications for technologists, and that the information technology (IT) industry is "barely regulated".

Whitby (2015, p.227) states that whilst medicine is highly regulated "the IT industry is barely regulated at all." Ethical codes of practice do exist for technologists; for example, there is a Code of Ethics and Professional Conduct published by the world's largest computing society, the Association for Computing Machinery (ACM 2018). The ACM code recommends that decision-making "is accountable to and transparent to all stakeholders" and stipulates qualities that technologists should possess, such as avoiding harm and acting honestly. In practice, the ACM does not have the power to enforce rules upon individuals beyond low impact punitive measures, such as termination of membership of the ACM. Termination would only demonstrate disproval from the ACM body, and it would not prevent a technologist from continuing their practice (ethical or not), but it is conceivable that ACM membership termination potentially might affect their access to activities such as future collaboration or opportunities such as funding.

There is an enforced requirement for the clinicians to be personally professionally accountable, via their professional codes of conduct, but no similarly enforced requirement for technologists to be personally professionally accountable. Technologists do not have a direct relationship with patients, but they are designing AISs which aim to contribute to clinical decision making with the clinician at the patient's bedside. This raises the question of if there ought to be a requirement for technologists to create and deploy AIS to be used in clinical decision making to be regulated in a similar fashion to the clinicians? Or, conversely, is regulation of technologists necessary if they are not directly interacting with the patients?

Having seen at how accountability is an enforced requirement for clinicians and not for technologists, the next section shall explore how the use of an opaque AIS interferes with accountability.

4.2 How does opacity interfere with accountability?

The inner workings of computerised systems are not always made visible. Opacity is when the process by which an output from an AIS is made is either too complex to be understood by one, many, or all stakeholders or that the decision-making process has been withheld completely from the stakeholder. Opacity is not the sole term used to describe this problem, for example, when an AIS's decision-making process is obscured it can be described as a "black box" (Mukherjee 2017), or as not being transparent (Hengstler et al. 2016). For the purposes of this review, these terms may be used interchangeably but the meaning remains consistent.

Mukherjee's (2017) commentary identifies that AIS are being developed in such a way that the process by which an AIS's outputs are calculated can be opaque; some of these systems are being designed for use in healthcare contexts with the goal being to help clinicians to improve patient outcomes. The problem here is that a clinician will ask an opaque AIS a question and they may have no idea how the answer outputted to them was created (Mukherjee 2017). This is additionally complicated by the opinion that using an AIS outputs which are delivered without verification risks the use of unpredictable or unwanted outputs (House of Commons: Science and Technology Committee 2016). Hengstler et al (2016) identified that trust is key to ensuring perceived risk reduction and that trust will be reinforced if the trustor is given algorithms which are transparent. Thus, it is reasonable to say that trust will be hard to win from the clinician if they are faced with an opaque AIS to use.

As a solution, verification and validation of AISs are recommended by the Association for the Advancement of Artificial Intelligence (House of Commons: Science and Technology Committee 2016, p.16):"it is critical that one should be able to prove, test, measure and validate the reliability, performance, safety and ethical compliance—both logically and statistically/probabilistically—of such robotics and artificial intelligence systems before they are deployed."

Verification and validation might assist the clinician to reasonably account for their actions if they chose to use an AIS; as identified earlier, the clinical codes of professional conduct do not permit practice which is not accountable (GMC 2013; NMC 2018; HCPC 2016). There is no mention in the literature reviewed of how a clinical user would know if the verification and validation of an AIS was appropriate, or how sufficient levels of safety from an AIS could be determined. To find a solution, it is reasonable to assume that this epistemic question will require future collaboration between clinicians and technologists.

It has been argued though that it is not solely AIS which can be opaque; that clinicians are also opaque. When interrogated, clinicians are not always able to explain exactly how they may come to a decision for an individual patient because their clinical judgement would be drawing from their experience as well as accepted rules which guide clinical care (Miles 2007). But this does not seem to be considered problematic in the literature reviewed. Thrum (interviewed by Sukel 2017b) exemplifies that if a clinician advises their patient that they have a melanoma, the patient does not interrogate the clinician's decision; instead they accept the biopsy and the subsequent treatment suggested. Thrum described how patients have traditionally accepted the opacity of medical decision-making and that diagnostic procedures and treatments are usually embraced without interrogating the practitioner's method of determination. One could say that it seems that it's acceptable for people to be opaque, but not the AIS that they are using, but the patient can take advantage of their clinician being professionally bound to be accountable for their practice (GMC 2013; NMC 2018; HCPC 2016); something which neither an AIS nor its creator currently is not bound to.

4.3 Examples of opaque AI scenarios

This review identified three main scenarios in the literature reviewed which illustrated the potential clinical use of opaque AISs and identified opacity as a source for concern regarding accountability of clinical decision making:

1) The AIS is understandable to one or more stakeholders but not all. Thus the AIS is not opaque to the technologist who builds it but is opaque to the end user: the clinician (Hartman 1986). The clinician might argue that they cannot use an opaque AIS as they would not be able to account for the determination of its outputs, thus be working against their code of professional conduct (GMC 2013; NMC 2018; HCPC, 2016). Given that technologists are not regulated and arguing that the public would not tolerate clinicians without qualifications to practice, Whitby (2015), p.227 finds it remarkable, "if not downright alarming", that clinicians would base their decision making on AIS created by "gifted amateurs".
2) Scenario 2 is as per scenario 1, but here the clinician does not hold specialist knowledge of the area which the AIS is advising them on (Ross and Swetlitz 2017). The use of IBM Watson for Oncology in UB Songdo Hospital, Mongolia, was investigated by Ross and Swetlitz (2017). They reported that this AIS is being used to advise generalist doctors who have either little or no training in cancer care. They describe that Watson works by looking at a patient's medical record, choosing what it calculates as the patient's options from a list of treatments, scoring those treatments as a percentage based upon how appropriate they are for the patient, and then presenting these options as recommendations for the clinician to consider. The options are presented to the clinician as a list ranked ordered by a score from highest to lowest. The AIS is opaque to the clinician as it is unable to explain why it gives treatments their scores (Hogan and Swetlitz video embedded in Ross and Swetlitz 2017). Suggestions from Watson are reportedly followed at UB Songdo Hospital almost at a rate of 100% despite the programme not explaining how its output was generated. Ross and Swetlitz (2017) demonstrate why this is concerning by describing the experience of an oncologist using the same Watson AIS in a South Korean hospital."Sometimes, he will ask Watson for advice on a patient whose cancer has not spread to the lymph nodes, and Watson will recommend a type of chemotherapy drug called taxane. But, he said, that therapy is normally used only if cancer has spread to the lymph nodes. And, to support the recommendation, Watson will show a study demonstrating the effectiveness of the taxane for patients whose cancer did spread to their lymph nodes.Kang is left confused as to why Watson recommended a drug that he does not normally use for patients like the one in front of him. And Watson can not tell him why." (Ross and Swetlitz (2017). Watson may arguably be safe in the hands of someone such as Kang who knows the subtle differences in the appropriate use of each of the treatments that the AIS recommends, but when the same technology is deployed in areas where that experience is lacking, the patient is at risk of receiving inappropriate treatments due to a lack of clinical safeguarding. In Mongolia the specialised clinical knowledge base was not ever present. A clinician may look to the technologist to provide reassurance that the AIS can be trusted, but in this scenario that reassurance is lacking. It would have been reassuring to know that Watson had been exposed to critical review by third parties outside IBM, but Ross and Swetlitz (2017) assert that this did not happen. It also appears that the company has also distanced itself from Watson's own outputs when applied in clinical practice; an IBM executive has been quoted by Hengstler et al (2016), p.115 saying that "Watson does not make decisions on what a doctor should do. It makes recommendations based on hypothesis and evidence based [sic.]".
3) The AIS is opaque to both the clinician and the technologist; its processes cannot be understood, resulting in outputs which may lack context (Mukherjee 2017). This risks a AIS's outputs being misunderstood, for example,

the AIS being used in a context which does not match its intended use resulting in its outputs being misapplied (Doroszewski 1988). Here it is arguable that accountability is unachievable by anyone prior to clinical use.

From the review's findings so far, it may be said that opacity may interfere with one's ability to account for using an AIS in clinical decision making. That there are multiple scenarios where using an opaque AIS in clinical decision making could raise issues of safety. That there is merit in an opaque AIS being subjected to a process of validation prior to use, but that such validation needs to be understood by the clinical user as being appropriate and sufficient. Given that the clinical professional bodies require their members to be accountable and to ensure patient safety, in the absence of an appropriate process of validation, the answer to this review's first question of "is it permissible for a clinician to use an opaque AI system in clinical decision making?" is currently 'no'.

4.4 Responsibility

As with accountability, no unified definition of responsibility was yielded by the literature searches. Therefore, for the purposes of this review, responsibility is assigned to one or more agents who hold the duty or obligation to respond or act correctly for an act or omission; the agent/s also are ascribed the blame or praise for the outcomes of their acts/omissions. Allocation of the responsibility for the consequences of AIS use may one day become needed if there are unintended consequences of AIS use; the following outlines those concerns. Understanding of the allocation of responsibility is illustrated by Whitby (2015); he is concerned that lack of clarification here regarding who holds responsibility for actions involving AIS use could result in a detriment to patient welfare (e.g., stakeholders blaming the AIS or each other rather than proactively ensuring that the AIS is functioning and being applied correctly).

Can AISs be responsible for themselves? Luxton (2014) is concerned that systems do not share the human suffering of moral consequences. Van Wynsberghe (2014) agrees, if a AIS cannot be punished, it cannot assume responsibility for roles incorporating the care of humans. Whitby (2015) warns that managers of AIS users should be explicit that clinicians cannot blame the AIS to avoid responsibility. Somewhat contrary to this, Van Wynsberghe (2014) holds that AISs can be delegated small roles where no harm to the patient can be caused; but may only carry out these roles when supervised by clinicians who hold responsibility for the patient. Here the clinician is the one who ensures that the AIS works as intended when deployed.

The literature seemed to agree that clinicians should take responsibility for opaque AIS's that they chose to use in clinical decision-making. Delvaux (2017) asserts that an AIS should assist the clinician; that the planning and final decision for the execution of a treatment must be made by a clinician. Pouloudi and Magoulas (2000) warn that an AIS's user is responsible for evaluating its outputs before using them. Whitby's (2015) insists that clinicians should maintain responsibility for outcomes when they use AISs and that clinicians ought not be allowed to escape that responsibility by blaming the AIS should negative outcomes arise. Kohane (interviewed in Sukel 2017a) explains that if there is a human clinician in the decision-making loop, the responsibility remains with them; the human would undertake to ensure that that which is advised by the AIS is safe and appropriate for the patient, the responsibility for the patient outcome of using that AIS output lies with them too. When discussing AISs which make diagnoses, Kohane (interviewed in Sukel 2017a) also states that if there is a decision-making disagreement between an AIS and the clinician using it, human third parties could "break the tie".

The literature was less clear regarding allocating responsibility to technologists. The ACM code states that "public good is always the primary consideration" and that its members should minimise the negative effects of their work such as threats to health and safety (ACM, 2018). But, beyond the ACM Code, the literature is divided, and it all seems to depend on what it is that one is asking the technologist to be responsible for.

Delvaux's report (2017, point 56), Doroszewski's essay (1988) and Vallverdu and Casacuberta's discussion (2015) place responsibility for an AIS's accuracy at the door of the person who trained that system. Doroszewski (1988) stresses the importance of this responsibility upon the technologist as the consequences of misrepresenting information in an AIS to be used in healthcare can be dire. Whitby (2015) underlines and specifies that technologists must share responsibility for consequences with clinicians when inappropriate advice is given by an AIS and used by the clinician. Doroszewski (1988) demonstrates that allocation of responsibility to an individual may not be easy though as, in the case of multiple authors making additions to an AIS, it might not be obvious who will take responsibility for the accuracy of the AIS which is created.

Some technologists are pushing back against this responsibility and refer to how their AISs are designed to defend against being assigned responsibility for the use of their creations outputs. Fenech et al.'s interviews (2018) identified the opinion that technologists should not hold responsibility for a system when it was designed to assist clinical decision-making rather than replacing it (e.g., the DeepMind system); that in this case, the responsibility remained with the clinician using it. This opinion was echoed by an IBM spokesperson interviewed in Hengstler et al. (2016), p.115), that Watson makes recommendations for a clinician and does not make the ultimate decision for patient treatment. Inthorn et al.'s discussion (2015) holds that doctors should retain the authority of decision-making as justifying and explaining the treatment to patients is their role, not the technologist's. The only exception to this rule is when a AIS is designed to work without clinical supervision; here, the technologist should be held responsible for AIS outcomes (Fenech et al, 2018; Kellmeyer et al. 2016).

The question of the allocation of responsibility in the event that harm is caused should a clinician use an AIS does not appear to have been fully resolved in the literature reviewed. There seems to be agreement that clinicians should be responsible for their actions when they use AIS, but there is no united agreement that technologists should be allocated responsibility too. Whitby (2015) stated that responsibility could be shared between clinicians and technologists should medical accidents or incidents that take place, but no authors suggest how this responsibility should be allocated between clinicians and technologists. This lack of clarification would concern clinicians and technologists as they would not be able to plan for the consequences of their contributions to healthcare when using or deploying AIS.

Instead of making definitive statements about who should be responsible for what, Whitby (2015) suggests that in the event of a negative outcome from using an AIS's outputs, interdisciplinary investigations should include all stakeholders and that blame should not be allocated, rather than the aim should be to prevent future harms. This balanced approach appears fairer to me as it recognises the complexity of the contributions made by each stakeholder in the process leading up to the AIS being used.

4.5 Liability

The Government Office for Science (2016) and Clarke (in House of Lords: Select Committee on Artificial Intelligence 2018) confirm that there is no body of case law yet to guide negligence and liability in this area. Little (in House of Lords: Select Committee on Artificial Intelligence 2018) advises that if civil and criminal liabilities and responsibilities are not considered before individual cases are brought, the resolutions resulting from existing legal frameworks may not be desirable. Yeung points out that if the courts have to find a solution for responsibility and liability then someone would have been harmed already. The Law Society (House of Commons: Science and Technology Committee 2016) explains that the downfall of relying on common law is that legal principles are developed after an untoward event, which is both expensive and stressful to stakeholders. Additionally, it can be said that reliance on common law makes forward planning difficult to carry out as well as planning and acquisition of appropriate insurance problematic.

Due to the lack of clarity, especially for that of dynamic AIS, the Law Commission was asked to investigate if current legislation is adequate to address liability and to make recommendations on this area. (Select Committee on Artificial Intelligence 2018). There has been no word of the Law Commission starting this requested work, and no one has speculated what the content of this review could be. There are several articles discussing issues of liability when using AISs in clinical decision making in the USA, but nothing recently focussed in English law where this review is concerned. Bainbridge (1991) discussed how the areas of negligence and contract could be applied in English law when AISs are used in the clinical context, but this work is of limited value nearly 30 years later as negligence law has moved on. This review's second question asked, "what concerns are there about opacity, accountability, responsibility and liability when considering the stakeholders of technologists and clinicians in the creation and use of AI systems in clinical decision making?".

This literature review has suggested that there are multiple multifaceted concerns which I shall now outline in brief. That it is not possible to account for an opaque AIS's outputs; thus, if one cannot account for the outputs, one cannot give a reasonable account for choosing to use those outputs. That if technologists provide opaque AISs to aid clinical decision making, they may find that clinicians choose not to use them as it would affect their ability to be accountable practitioners. That the formulation whereby responsibility is allocated is not concrete; that there seems to be a consensus that clinicians should hold responsibility for choosing to use an opaque AIS, but that there is no such accord for technologists joining them in holding that responsibility even though some authors feel that this response could be shared. That there is no case law or legislation in the law of England and Wales which is specific to negligence and liability cases in the use of AISs in clinical decision making; this lack of clarity has prevented stakeholders from confident future planning in the undesirable scenario of patient harm. That waiting for the courts to find a solution to the allocation of responsibility and liability would require that someone came to harm first. It is reasonable to say that there is a current opportunity to proactively address these issues before harm takes place, rather than allowing harm to take place and retrospectively allocating ethical and legal responsibility. One wonders: if this opportunity is taken, avoidable harm could be prevented.

4.6 Limitations

The author recognises that there are limitations to this review, the following notes those which they were able to identify.

This review found a lack of consistency in the language used when considering opacity as well as an enormous variety of subgroups of AIS systems in use. These two factors challenged the author to appropriately and inclusively recognise the multitudes of terms and programming language in existence which populate the literature discussing this review's concerns.

Regarding the subgroups of algorithm types in AI; this review intentionally did not identify a particular group (such as machine learning) lest the discussion become sidetracked by specifically which AIS's are being used rather than consideration of how the AIS are being used. Currently, machine learning is well-represented in the current debate, but this has not always been the case and another subgroup may prove to be more popular in the future (in House of Lords: Select Committee on Artificial Intelligence 2018).

Regarding the lack of consistent terminology made the literature searches challenging; for example, AI opacity could also be described as the AIS being a black box, or that there was a lack of AIS transparency. Increasing the number of search terms to attempt to capture the variety of terms did not improve the number of search results returned, nor the relevance of those search results. Ultimately 'opacity' was adopted as the primary descriptor and employed as the search term as it yielded the highest volume of relevant results in the literature searches.

The author suspects that there has likely been much relevant material from a variety of worthy sources which has been lost to this review due to the changing nature of how information (especially regarding technology) has been communicated in recent years. Relevant and worthy ideas, concepts and opinions are no longer routinely published in the traditional way, i.e., via peer-reviewed journals, thus are not admitted to academic database searches which are the main pathway for discovering material for a systematic review such as this. For this reason, media items from outside of the realms of the traditional academic sources were selected when they were determined as pertinent to this review. For example, Ross and Swetlitz's useful and demonstrative report of the use of IBM Watson would have been lost to this review had media been excluded.

Since the searches were performed in 2018, there could well have been more materials published which would be worthy of inclusion which has not been captured.

5 Conclusion

This literature review suggests that there are multiple concerns about opacity, accountability, responsibility and liability when considering the stakeholders of technologists and clinicians in the creation and use of AIS in clinical decision making. Accountability is challenged when the AIS in use is opaque, and allocation of responsibility is somewhat unclear. Legal analysis would help stakeholders to understand their obligations and prepare should an undesirable scenario of patient harm eventuate when AISs are used.

Acknowledgements The author thanks the editor and reviewers for their insightful and helpful comments which greatly improved this article. The author thanks Dr Jonathan Ives, Dr Giles Birchley, and Professor Andrew Charlesworth, for offering their patience, expertise and comments in helping the development of this review's earlier incarnation as a Ph.D. thesis chapter. The author thanks Miss Alice Toms for her independent second review of the search results, Dr Verity Jones for proofreading and support, and Mr Kit Fotheringham for his comments in final drafting. Thanks also to Mr Christopher Martin for proofreading, coffee making, and being an all-round jolly good husband.

Author contributions Single author, all own work.

Funding The author received no funding for the production of this work.

Compliance with ethical standards

Conflict of interest The author declares no conflicts of interest.

References

Armstrong K (2018) If you cannot beat it, join it: uncertainty and trust in medicine. Ann Int Med 168(11):818–819

Association for computing machinery (2018) acm code of ethics and professional conduct. https://www.acm.org/code-of-ethics. Accessed Feb 2018

Bainbridge DI (1991) Computer-aided diagnosis and negligence. Med Sci Law 31:127–136

Braun V, Clarke V (2006) Using thematic analysis in psychology. Q Res Psychol 3(2):77–101

Char DS, Magnus D, Shah NH (2018) Implementing machine learning in health care — addressing ethical challenges. New Eng J Med 378(11):981–983

Delvaux M, Affair (2017) C.o.L. Report 27 Jan 2017; with recommendations to the Commission on Civil Law Rules on Robotics [2015/2103(INL)]. (online: European Parliament)

Dignum V (2019) Responsible artificial intelligence: how to develop and use AI in a responsible way. Springer, Switzerland

Doroszewski J (1988) Ethical and methodological aspects of medical computer data bases and knowledge bases. Theoret Med 9(2):117–128

Fenech M, Strukelj N, Buston O (2018) Ethical, social and political challenges of artificial intelligence in health: future Advocacy report for the Wellcome Trust. https://wellcome.ac.uk/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf. Accessed Feb 2018

General Medical Council (2013) Good medical practice: general medical council. https://www.gmc-uk.org/-/media/documents/good-medical-practice---english-1215_pdf-51527435.pdf. Accessed Feb 2018

Government Office for Science (2016) Challenges and uses of AI set out by UK government: "Artificial intelligence: opportunities and implications for the future of decision making". https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf. Accessed Feb 2018

Hancock M. (2018) Matt Hancock writes in the Health Service Journal about his admiration for the NHS and its staff: Gov.UK. https://www.gov.uk/government/speeches/matt-hancock-writes-in-the-health-service-journal-about-his-admiration-for-the-nhs-and-its-staff. Accessed July 2018

Hartman DE (1986) On the use of clinical psychology software. practical, legal, and ethical concerns. Prof Psychol Res Pract 17(5):462–465

Harwich E, Laycock K. (2018) Thinking on its own: AI in the NHS Reform. https://www.reform.uk/publication/thinking-on-its-own-ai-in-the-nhs/. Accessed Feb 2018

Health Care Professions Council (2016) Standards of conduct, performance and ethics. https://www.hcpc-uk.org/publications/standards/index.asp?id=38. Accessed Feb 2018

Hengstler M, Enkel E, Duelli S (2016) Applied artificial intelligence and trust—the case of autonomous vehicles and medical assistance devices. Technol Forecast Soc Chang 105:105–120

House of Commons: Science and Technology Committee (2016) Robotics and artificial intelligence fifth report of session 2016–2017 House of Commons, UK. https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf. Accessed Feb 2018

House of Lords: Select Committee on Artificial Intelligence (2018) AI in the UK: ready, willing and able? House of Lords, UK. https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf. Accessed Apr 2018

Inthorn J, Tabacchi ME, Seising R (2015) Having the final say: machine support of ethical decisions of doctors. In: Rysewyk S, Pontier M (eds) Machine Medical Ethics (Intelligent Systems, Control and Automation: Science and Engineering). Springer, Switzerland

Kellmeyer P, Cochrane T, Müller O, Mitchell C, Ball T, Biller-Andorno JJ, Fins N (2016) The effects of closed-loop medical devices on the autonomy and accountability of persons and systems. Camb Q Healthc Ethics 25(4):623–631

Khan KS, Kunz R, Kleijnen J, Antes G (2003) Five steps to conducting a systematic review. J R Soc Med 96:118–121

Knight W. (2017) 'The Dark Secret at the Heart of AI', MIT Technol Rev. https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/. Accessed Feb 2018

Kyte DG, Draper H, Ives J, Liles C, Gheorghe A, Calvert M, Timmer A (2013) Patient reported outcomes (PROs) in clinical trials: is 'in-trial' guidance lacking? A systematic review. PLoS ONE 8(4):e60684

Luxton DD (2014) Recommendations for the ethical use and design of artificial intelligent care providers. Artif Intell Med 62(1):1–10

Makary MA, Daniel M (2016) Medical error—the third leading cause of death in the US. BMJ 353:2139

Miles A (2007) Science: a limited source of knowledge and authority in the care of patients. a review and analysis of: 'how doctors think. clinical judgement and the practice of medicine. J Evaluation Clin Pract 13(4):545–563

Mukherjee, S. (2017) A.I. Versus M.D. The New Yorker. https://www.newyorker.com/magazine/2017/04/03/ai-versus-md. Accessed Feb 2018

NICE (2012) Patient experience in adult NHS services: improving the experience of care for people using adult NHS services: clinical guideline [CG138]. https://www.nice.org.uk/guidance/cg138/chapter/1-Guidance#enabling-patients-to-actively-participate-in-their-care. Accessed Feb 2018

Nursing and Midwifery Council (2018) The code for nurses and midwives. Nursing and Midwifery Council, London. https://www.nmc.org.uk/standards/code/read-the-code-online/. Accessed February 2018

Oshana M (2004) Moral accountability. Philos Top 32(1–2):255–274

Pouloudi A, Magoulas GD (2000) Neural expert systems in medical image interpretation: development, use, and ethical issues. J Intel Syst 10(5–6):451–472

Ross C. and Swetlitz I. (2017) IBM pitched Watson as a revolution in cancer care. It is nowhere close. STAT News. https://www.statnews.com/2017/09/05/watson-ibm-cancer/. Accessed Feb 2018

Strech D, Sofaer N (2012) How to write a systematic review of reasons. J Med Ethics 38:121–126

Sukel K (2017a) Artificial Intelligence ushers in the era of superhuman doctors, New Scientist. https://www.newscientist.com/article/mg23531340-800-artificial-intelligence-ushers-in-the-era-of-superhuman-doctors/. Accessed Feb 2018

Sukel K (2017b) With a little help from AI friends. New Scientist 235(3134):36–39

Editorial (2018) Opening the black box of machine learning. Lancet Respir Med 6(11):801

Vallverdú J, Casacuberta D (2015) Ethical and technical aspects of emotions to create empathy in medical machines. In: Rysewyk S, Pontier M (eds) Machine Medical Ethics (Intelligent Systems, Control and Automation: Science and Engineering). Springer, Switzerland

van Wynsberghe A (2014) To delegate or not to delegate: care robots, moral agency and moral responsibility. Anniver AISB Convention 20:14

Weizenbaum J (1976) Computer power and human reason: from judgement to calculation. Penguin Books Ltd, Harmondsworth, England

Whitby B (2015) Automating medicine the ethical way. In: Pontier M (ed) Rysewyk Machine Medical Ethics (Intelligent Systems, Control and Automation: Science and Engineering). Springer, Switzerland

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