Clinicians Risk Becoming "Liability Sinks" for Artificial Intelligence

T Lawton,[¶§] P Morgan,[§] Z Porter,[§] A Cunningham,[¶] N Hughes,[§] I Iacovides,[§] Y Jia,[§] V Sharma,[¶] I Habli[§]

¶ Improvement Academy, Bradford Institute for Health Research, Bradford Royal Infirmary, Duckworth Lane, Bradford, BD9 6RJ

§ Assuring Autonomy International Programme, University of York, Heslington, York, YO10 5DD

The Problem

Artificial Intelligence (AI) is often touted as healthcare's saviour, but its potential will only be realised if developers and providers consider the whole clinical context and AI's place within it. One of many aspects of that clinical context is the question of liability.

In the current, standard model of AI-supported decision-making in healthcare, electronic data is fed into an algorithm, typically a machine-learnt model, which combines it all to arrive at a recommendation which is output to a human clinician. The clinician then acts as a final check on the system's recommendation, and can either accept it as-is, or replace it with a decision they make themselves (see Figure 1 below). We are aware of it already being assumed by AI radiology companies, who label their systems as "assistance" and clarify that responsibility lies fully with the user, largely to reassure about safety concerns. Given recent guidance from the National Health Service in England, which clarifies that the final decision must be taken by a healthcare professional,[1] this model looks set to become the norm across the UK healthcare system.

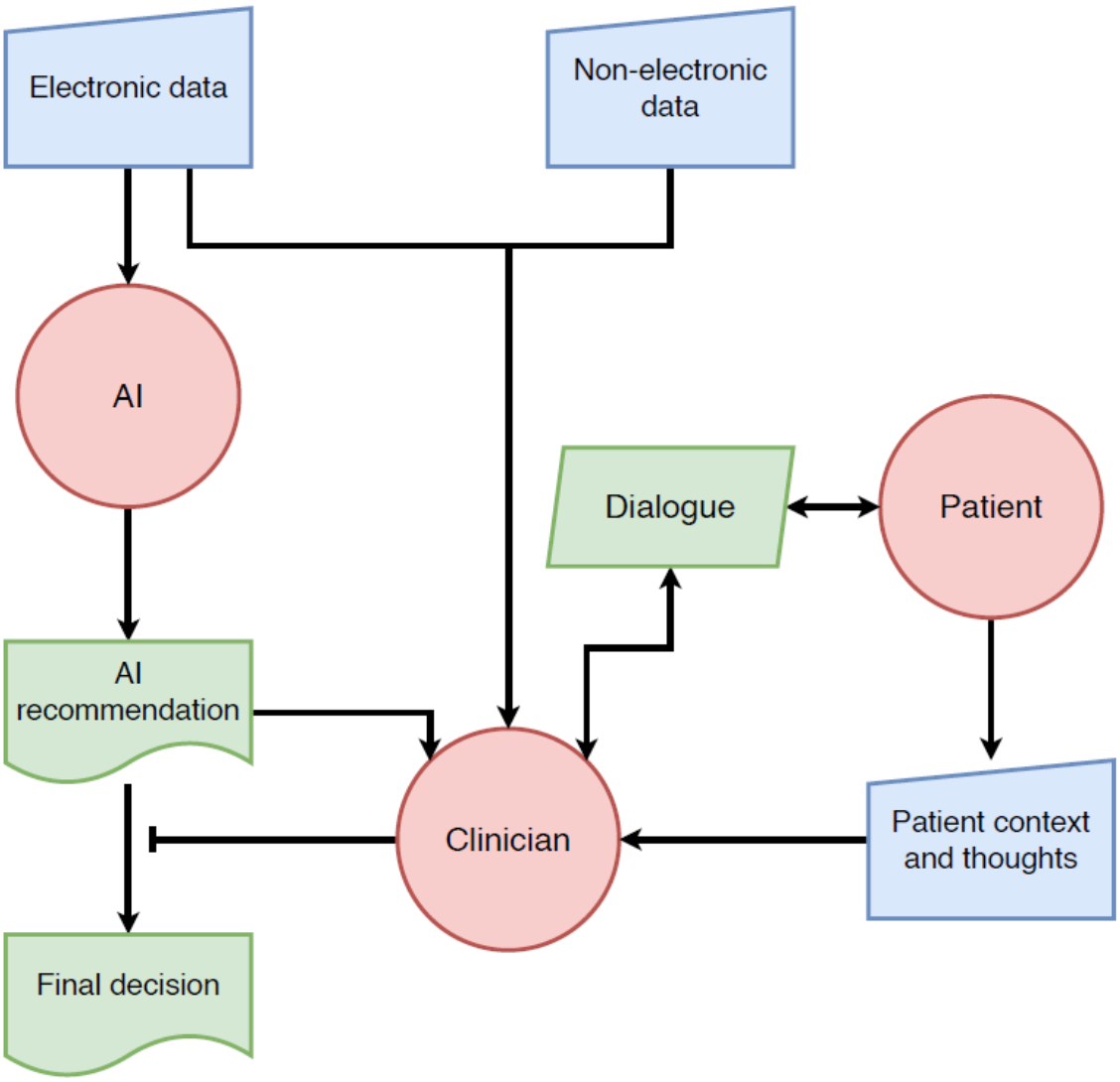

Figure 1 – Current prevalent AI model

But the standard model may not be the best model for AI-supported decision-making in clinical practice. One particular problem is the negative impact on the clinician, as a human facing numerous cognitive and practical challenges when monitoring automation,[2] who is faced with a binary choice of accepting the AI recommendation or ignoring it and reverting to a traditional (non-AI) approach. They risk no longer doing what they are best at, including exercising sensitivity to patient preferences and context, but in effect acting as a sense-check on, or conduit for, the machine. At the same time, the guidance states that the clinician may be held legally accountable for a decision made using the support of AI. Analogous to the way a "heat sink" takes up unwanted heat from a system, the human clinician risks being used here as a "liability sink", where they absorb liability for the consequences of the AI's recommendation whilst being disenfranchised from its decision-making process.

There are similarities here to the field of driver assistance and self-driving systems for cars, where despite the AI being in direct control of the vehicle, in some jurisdictions it seems the human in the driving seat is already being used as a liability sink. For example, a driver activating self-driving mode

typically has to accept that they will take over manual control immediately when required. But US investigation of some Tesla collisions has found that Autopilot aborted control on average less than one second prior to the first impact.[3] This does not give the driver enough time to resume control safely - and yet in practice, for jurisdictions that adopt fault based systems of liability for motor vehicle accidents such the UK, it is likely that they would be liable for the accident. As the most obvious "driver" close to where AI is used in a clinical setting, the clinician could easily end up being held similarly liable for harmful outcomes from AI-based decision-support systems, and carrying this stress and worry, but having limited practical control over their development and deployment, or understanding of how the AI recommendations are reached.[4]

Possible Solutions

The attribution of liability in a whole socio-technical system becomes complex when AI is involved. As well as the humans directly present at the event, there were humans involved in the design and commissioning of the AI system, humans who signed off on its safety, and humans overseeing its running or working in tandem with it. Complexity is further increased with AI because human oversight may be more influenced by automation bias - where humans attribute greater than warranted intelligence to the machine - and because the AI's decision-making cannot be clearly understood by those operating it. Given that automation bias and AI inscrutability are problems across many settings where AI is used, it is no surprise that efforts are already being made to solve them.[5,6]

Whilst we are some way off it being possible, or even appropriate, to hold an AI system itself liable,[7] any of the humans involved in an AI's design, building, provisioning, and operation might be held liable to a degree. Smith and Fotheringham argue that using clinicians as the sole focus for liability is not "fair, just and reasonable".[8] Without a clear understanding of how an AI came to a decision, a clinician is faced with either treating it as a knowledgeable colleague,[9–11] or coming to their own judgement and largely ignoring the AI - or even turning it off. Even if they resolve to make their own decision and then check it against the AI's recommendations, this only avoids the problem when there is agreement. If the AI disagrees, the clinician faces the same dilemma.

Unfortunately, the clinician and their employer via vicarious liability for the clinician's negligence, remain the most attractive defendants to sue.[12] 'Vicarious liability' is when an employer is held liable for the negligence or wrongdoing of an employee. In a medical negligence context, negligence still traditionally focuses on the individual – although moving to a model focused on the system as a whole would be more useful, both for patient safety and for the impact on individual clinicians.[13] Meanwhile, AI systems are currently treated as products, so the software development company (SDC) would only liable to the patient through product liability. In the future, it may be that the AI system is treated as part of the clinical team – and not as a product – so that its 'conduct' could be attributed to those who 'employ' the AI system, which may for instance be the SDC, or clinician's trust.[14] But that is not the current legal context. It is also unclear what 'standard of care' would apply to an AI that is treated as part of the clinical team: that of the reasonable AI system, or that of the reasonable clinician?[15] In a case where the system was being held to the higher standard, the SDC might argue that this is unreasonable. But this implies that their system is simply not good enough -

that its recommendations are inferior to the decisions of a clinician - and few organisations would be willing to deploy an AI system on that basis.

Given the SDC's involvement, Smith and Fotheringham argue that there should be risk pooling between clinicians and SDCs for the harm - with actuarially-based risk pooling insurance schemes to provide cover for AI-related damage.[8] However, these are at present merely proposals. Currently, a clinician (using an AI system) who is held liable in negligence to the patient may seek contribution from the SDC via the Civil Liability (Contribution) Act 1978, although, as with the patient's claim against the SDC there are significant difficulties in doing so, since as noted above establishing that the SDC is itself liable for the damage suffered is problematic. The SDC may also have sought to contractually exclude any right of clinicians to seek such contribution. Thus, in practical terms with systems of this type the clinician remains liable for acting on the recommendations or decisions of an AI they do not and cannot fully understand. Facing the stress and worry of the consequences of using it, many clinicians may refuse to accept the risk, and simply turn off the machine.

Alternative models

Whilst pooling risk might prevent the clinician becoming a liability sink, it may be suboptimal in other ways for the clinician, the patient, and the system as a whole. Figure 1 shows that the entire input of the patient and clinician into the decision is restricted to either accepting the AI's recommendation, or - for this case - ignoring the AI entirely (effectively switching it off and returning to standard practice). This is at odds with the goal of patient-centred decision-making,[16] as the AI cannot easily incorporate patient context and ideas, concerns, and expectations itself - this context is only added by the clinician choosing to accept or replace the AI's output. It may also be frustrating for the clinician by eroding their ability to do what they do best: integrating clinical science and patient context in a dialogue to come to a shared decision.

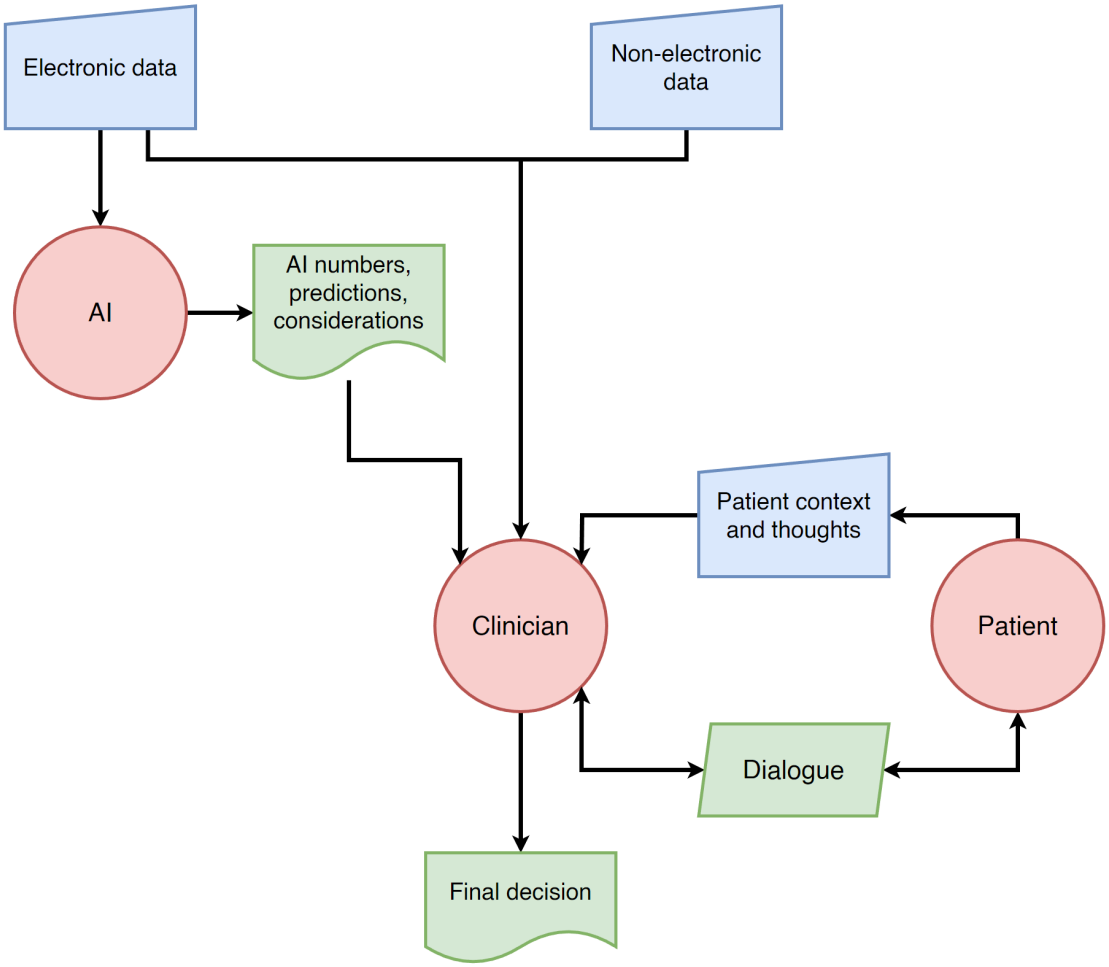

Figure 2 - AI model with alternative outputs to inform patient/clinician dialogue

Fortunately, the currently prevalent model is not the only possible approach. Rather than restructuring systems in a clinical setting around an AI designed to work this way, it may be preferable to explore alternative models which give greater focus to the patient and clinician.[17] In some of these models the output from the AI may not even take the form of a decision or recommendation, but instead show predictions of the effect of different decisions (e.g. treatment options), or highlight the data that is most relevant to the AI model in its decision making. In this way, the explanation of an explainable-AI system may be more useful than the decision or recommendation itself.[18,19] Figure 2 above shows a model where these alternative outputs from the AI system inform a dialogue between the clinician and patient, out of which a decision emerges. In Figure 3 below, a more advanced AI system communicates directly with the patient and a three-way dialogue proceeds before a decision emerges. Other models could be conceived along these lines, bringing the patient and clinician back into the decision-making focus.

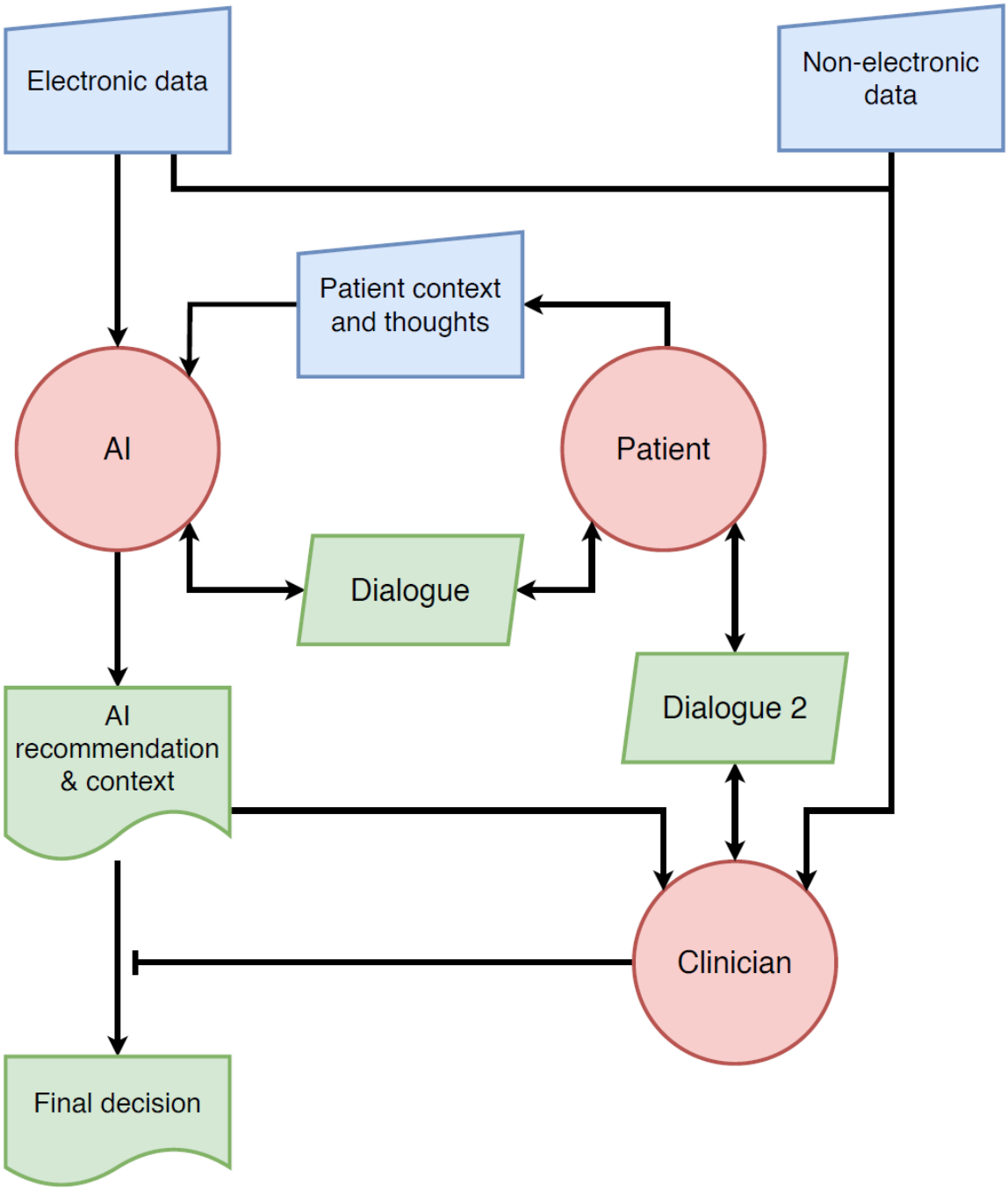

Figure 3 - Advanced AI model capable of sustaining dialogue with the patient

It is notable that with both of the models in Figure 2 and Figure 3, the clinician retains the final decision as recommended by NHS England. Are they still acting as a liability sink for the AI? We would argue that their role here is much more traditional and that they are integrating a variety of data and opinions, in a manner of working that has become familiar with the advent of the multidisciplinary team.[20] Clinicians should feel much more comfortable in accepting liability for a decision where they have genuine understanding and agency, and the socio-technical system as a whole will be much more acceptable to both clinicians and patients as it retains compatibility with patient-centred care.

The question remaining in this setup, however, is the assignment of liability where the advice or information provided by the AI is defective. By returning the clinician to a more traditional role with these models, it becomes more appropriate to treat the AI as a standard medical device. This could be dealt with via product liability, suitably adjusted to take into account the problems within such regimes as applied to AI systems, such as proof of causation, and the failure of the existing Consumer Protection Act 1987 (implementing the European Union's Product Liability Directive ('PLD')[21]) to cover unembodied software. The need for such adjustments has been recognised by the European Union, which has published reform proposals for the PLD. If we do not want clinicians to become liability sinks, similar reforms may need to be considered in the United Kingdom.

In summary, AI systems being developed using current models risk using clinicians as "liability sinks", absorbing liability which could otherwise be shared across all those involved in the design, institution, running, and use of the system. Alternative models can return the patient to the centre of decision-making, and also allow the clinician to do what they are best at, rather than simply acting as a final check on a machine.

Summary

- The benefits of AI in healthcare will only be realised if we consider the whole clinical context and the AI's role in it.
- The current, standard model of AI-supported decision-making in healthcare risks reducing the clinician's role to a mere 'sense check' on the AI, whilst at the same time leaving them to be held legally accountable for decisions made using AI.
- This model means that clinicians risk becoming "liability sinks", unfairly absorbing liability for the consequences of an AI's recommendation without having sufficient understanding or practical control over how those recommendations were reached.
- It also means that clinicians are less able to do what they are best at, specifically exercising sensitivity to patient preferences in a shared clinician-patient decision-making process.
- There are alternatives to this model that can have a more positive impact on clinicians and patients alike.

References

1. NHS England. Information Governance Guidance: Artificial Intelligence [Internet]. NHS England - Transformation Directorate; 2022 [cited 2022 Nov 3]. Available from: https://transform.england.nhs.uk/information-governance/guidance/artificial-intelligence/

2. Bainbridge L. Ironies of automation. In: Johannsen G, Rijnsdorp JE, editors. Analysis, Design and Evaluation of Man–Machine Systems [Internet]. Pergamon; 1983 [cited 2023 Feb 22]. p. 129–35. Available from: https://www.sciencedirect.com/science/article/pii/B9780080293486500269

3. Engineering Analysis 22-002 [Internet]. National Highway Traffic Safety Administation, Office of Defects Investigation; 2022 [cited 2022 Nov 3]. Available from: https://static.nhtsa.gov/odi/inv/2022/INOA-EA22002-3184.PDF

4. Habli I, Lawton T, Porter Z. Artificial intelligence in health care: accountability and safety. Bulletin of the World Health Organization. 2020 Feb;98(4):251–6.

5. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. JAMA Internal Medicine. 2018 Nov 1;178(11):1544–7.

6. McDermid JA, Jia Y, Porter Z, Habli I. Artificial intelligence explainability: the technical and ethical dimensions. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2021 Aug 16;379(2207):20200363.

7. Chesterman S. Artificial intelligence and the limits of legal personality. ICLQ. 2020;69(4):819–44.

8. Smith H, Fotheringham K. Artificial intelligence in clinical decision-making: Rethinking liability. Medical Law International. 2020 Jun 1;20(2):131–54.

9. Wilsher v Essex Area Health Authority [1987] QB 730 (CA). 1987.

10. Junior v McNicol. Times Law Reports, March 26 1959. 1959.

11. Armitage M, editor. Chapter 10: Persons Professing Some Special Skill. In: Charlesworth & Percy on Negligence. 15th ed. London: Sweet & Maxwell; p. 10–147. (Common Law Library).

12. Burton S, Habli I, Lawton T, McDermid J, Morgan P, Porter Z. Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective. Artificial Intelligence. 2020 Feb 1;279:103201.

13. Heywood R. Systemic Negligence and NHS Hospitals: An Underutilised Argument. King's Law Journal. 2021 Sep 2;32(3):437–65.

14. Morgan, Phillip. Chapter 6: Tort Law and Artificial Intelligence – Vicarious Liability. In: Lim E, Morgan P, editors. The Cambridge Handbook of Private Law and Artificial Intelligence. Cambridge University Press;

15. Abbott R. The Reasonable Robot: Artificial Intelligence and the Law [Internet]. Cambridge: Cambridge University Press; 2020 [cited 2023 Feb 22]. Available from: https://www.cambridge.org/core/books/reasonable-robot/092E62F0087270F1ADD9F62160F23B5A

16. Bjerring JC, Busch J. Artificial Intelligence and Patient-Centered Decision-Making. Philos Technol. 2021 Jun 1;34(2):349–71.

17. Birch J, Creel KA, Jha AK, Plutynski A. Clinical decisions using AI must consider patient values. Nat Med [Internet]. 2022 Jan 31 [cited 2022 Feb 1]; Available from: https://www.nature.com/articles/s41591-021-01624-y

18. Jia Y, Mcdermid JA, Lawton T, Habli I. The Role of Explainability in Assuring Safety of Machine Learning in Healthcare. IEEE Transactions on Emerging Topics in Computing. 2022;1–1.

19. Mittelstadt B, Russell C, Wachter S. Explaining Explanations in AI. In: Proceedings of the Conference on Fairness, Accountability, and Transparency [Internet]. New York, NY, USA: Association for Computing Machinery; 2019 [cited 2023 Feb 22]. p. 279–88. (FAT* '19). Available from: https://doi.org/10.1145/3287560.3287574

20. Epstein NE. Multidisciplinary in-hospital teams improve patient outcomes: A review. Surgical Neurology International. 2014;5(Suppl 7):S295.

21. Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products [Internet]. OJ L Jul 25, 1985. Available from: http://data.europa.eu/eli/dir/1985/374/oj/eng

Acknowledgements

This work was supported by The MPS Foundation Grant Programme. The MPS Foundation was established to undertake research, analysis, education and training to enable healthcare professionals to provide better care for their patients and improve their own wellbeing. To achieve this, it supports and funds research across the world that will make a difference and can be applied in the workplace. The work was also supported by the Engineering and Physical Sciences Research Council (EP/W011239/1).

Conflicts of interest

TL has received an honorarium for a lecture on this topic from Al Sultan United Medical Co and is head of clinical artificial intelligence at Bradford Teaching Hospitals NHS Foundation Trust, and a potential liability sink

All other authors report no conflicts of interest

Authors' contributions

TL, ZP, IH - conceptualisation, funding acquisition, writing - original draft & review & editing, analysis, visualisation

PM - writing - original draft, analysis, visualisation, writing - review & editing

AC, NH, JI, YJ, VS - analysis, visualisation, writing - review & editing