

Eight human factors and ergonomics principles for healthcare artificial intelligence

Mark Sujan  ^{1,2} Rachel Pool,³ Paul Salmon⁴

To cite: Sujan M, Pool R, Salmon P. Eight human factors and ergonomics principles for healthcare artificial intelligence. *BMJ Health Care Inform* 2022;29:e100516. doi:10.1136/bmjhci-2021-100516

Received 17 November 2021
Accepted 26 January 2022

INTRODUCTION

The COVID-19 pandemic dramatically accelerated the digital transformation of many health systems in order to protect patients and healthcare workers by minimising the need for physical contact.¹ A key part of healthcare digital transformation is the development and adoption of artificial intelligence (AI) technologies, which are regarded a priority in national health policies.^{2,3} Since 2015, there has been an exponential growth in the number of regulatory approvals for medical devices that use machine learning,⁴ with British standards currently under development in conjunction with international standards. In addition, there are an even larger number of healthcare AI technologies that do not require such approvals, because they fall outside of the narrow definition of medical devices.

The scope of healthcare AI appears seemingly boundless, with promising results being reported across a range of domains, including imaging and diagnostics,⁵ prehospital triage,⁶ care management⁷ and mental health.⁸ However, caution is required when interpreting the claims made in such studies. For example, the evidence base for the effectiveness of deep learning algorithms remains weak and is at high risk of bias, because there are few independent prospective evaluations.⁹ This is particularly problematic, because the performance, usability and safety of these technologies can only be reliably assessed in real-world settings, where teams of healthcare workers and AI technologies co-operate and collaborate to provide a meaningful service.¹⁰ To date, however, there have been few human factors and ergonomics (HFE) studies of healthcare AI.¹¹ There is a need for AI designs and prospective evaluation studies that consider the performance of the overall sociotechnical system, with evidence requirements proportionate to the level of risk.¹²

Reporting guidelines have been developed both for small-scale early clinical intervention trials (DECIDE-AI)¹³ as well as for large-scale clinical trials evaluating AI (SPIRIT-AI)¹⁴ to enhance the quality and transparency of the evidence.

In order to support developers, regulators and users of healthcare AI, the Chartered Institute of Ergonomics and Human Factors (CIEHF) developed a white paper that sets out an HFE vision and principles for the design and use of healthcare AI.¹⁵ Development of the white paper was an international effort bringing together over 30 contributors from different disciplines and was supported by a number of partner organisations including British Standards Institution, the Australian Alliance for AI in Healthcare, the South American Ergonomics Network (RELAESA), US-based Society for Healthcare Innovation, the UK charity Patient Safety Learning, Assuring Autonomy International Programme hosted by the University of York, Human Factors Everywhere and the Irish Human Factors & Ergonomics Society.

HFE PRINCIPLES

HFE as a discipline is concerned with the study of human work and work systems. It is a design-oriented science and field of practice that seeks to improve system performance and human well-being by understanding and optimising the interactions between people and the other elements of the work system, for example, technologies, tasks, other people, the physical work environment, the organisational structures and the external professional, political and societal environment.¹⁶

Current implementations of healthcare AI typically adopt a technology-centric focus, expecting healthcare systems (including staff and patients) to adapt to the technology. In this technology-centric focus, the function,



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Human Factors Everywhere, Woking, UK

²Chartered Institute of Ergonomics and Human Factors, Birmingham, UK

³NHS England, Redditch, UK

⁴Centre for Human Factors and Sociotechnical Systems, University of the Sunshine Coast, Maroochydore DC, Queensland, Australia

Correspondence to

Dr Mark Sujan;
mark.sujan@humanfactorsever
ywhere.com

**Table 1** Eight human factors and ergonomics principles for healthcare AI

Situation awareness	Design options need to consider how AI can support, rather than erode, people's situation awareness.
Workload	The impact of AI on workload needs to be assessed because AI can both reduce as well as increase workload in certain situations.
Automation bias	Strategies need to be considered to guard against people relying uncritically on the AI, for example, the use of explanation and training.
Explanation and trust	AI applications should explain their behaviour and allow users to query it in order to reduce automation bias and to support trust.
Human–AI teaming	AI applications should be capable of good teamworking behaviours to support shared mental models and situation awareness.
Training	People require opportunities to practise and retain their skill sets when AI is introduced, and they need to have a baseline understanding of how the AI works. Attention needs to be given to the design of effective training that is accessible and flexible. Staff should be provided with protected time to undertake training during their work hours.
Relationships between people	The impact on relationships needs to be considered, for example, whether staff will be working away from the patient as more and more AI is introduced.
Ethical issues	AI in healthcare raises ethical challenges including fairness and bias in AI models, protection of privacy, respect for autonomy, realisation of benefits and minimisation of harm.

AI, artificial intelligence.

performance and accuracy of AI are optimised, but these aspects are considered in isolation. This perspective raises various critical considerations that are often overlooked in the design and implementation of advanced technologies, sometimes with catastrophic consequences. From an HFE point of view, the design of healthcare AI needs to transition from the technology-centric focus towards a systems perspective. Applying a systems focus, AI should be designed and integrated into clinical processes and healthcare systems meaningfully and safely, with a view to optimising overall system performance and people's well-being. Understanding how a sociotechnical system works comes from taking time to look at the elements of the system and how they interact with each other. HFE provides several frameworks and methods to achieve this, including Systems Engineering Initiative for Patient Safety¹⁷ and Cognitive Work Analysis.¹⁸ These frameworks usually involve the use of observation or ethnography for data collection in order to provide a rich contextual description of how work is carried out ('work-as-done'¹⁹) and of people's needs.

The CIEHF white paper identifies eight core HFE principles, see **table 1**. Some of these are very familiar from the wider literature on automation and date back to the 1970s and 1980s but retain their importance in the novel context of healthcare AI. For example, the potentially adverse impact of highly automated systems on user situation awareness and workload, along with the potential for over-reliance and automation bias, became apparent decades ago in a series of transportation accidents and incidents.^{20 21} These 'ironies of automation'²² can arise when technology is designed and implemented without due consideration of the impact on human roles or the interaction between people and the technology, which can result in inadequate demands on the human, such as lengthy periods of passive monitoring, the need to

respond to abnormal situations under time pressure and difficulties in understanding what the technology is doing and why. Alarm fatigue, that is, the delayed response or reduced response frequency to alarms, is another phenomenon associated with automated systems that has been identified from major industrial accidents, such as the 1994 explosion and fires at the Texaco Milford Haven refinery. In intensive care, it has been suggested that a healthcare professional can be exposed to over 1000 alarms per shift, contributing to alarm fatigue, disruption of care processes and noise pollution, with potentially adverse effects on patient safety.²³ Developers of AI need to be mindful of these phenomena and not create technologies that add additional burden to healthcare professionals.

However, the use of more advanced and increasingly autonomous AI technologies also presents novel challenges that require further study and research. AI technologies can augment what people do in ways that were not possible when machines simply replaced physical work, but in order to do this effectively the AI needs to be able to communicate and explain to people its decision-making. This can be very challenging when using machine learning algorithms that produce complex and inscrutable models. Many approaches to explainable AI simply focus on providing detailed accounts of how an algorithm operates, but for explanations to be useful they need to be able to accommodate and be responsive to the needs of different users across a range of situations, for example, a patient might benefit from a different type of explanation compared with a healthcare professional. In this sense, rather than providing a description of a specific decision, explanation might be better regarded as a social process and a dialogue that allows the user to explore AI decision-making by interacting with the AI and by interrogating AI decisions.²⁴



It is also important to build trust among staff to report any safety concerns with the AI. Many safety incidents are not currently reported and recorded in incident reporting systems.²⁵ While an AI system can potentially log every piece of data and every one of its actions to provide an auditable history, healthcare professionals require assurance and reassurance of how these data would be used during a safety investigation. If clinicians are held accountable for incidents involving AI unless they can prove otherwise, then this might reduce their willingness to trust and accept AI systems.

Many applications of healthcare AI will be used within teams of healthcare workers and other professionals, as well as patients. The computational capabilities of AI technologies mean that AI applications will have a much more active and dynamic role within teams than previous IT systems and automation, in effect potentially becoming more like a new team member than just a new tool. Effective human–AI teaming will become increasingly critical when designing and implementing AI to ensure that AI capabilities and human expertise, intuition and creativity are fully exploited.²⁶

Part of effective human–AI teaming is handover from the AI to the healthcare professional when it becomes necessary.¹⁰ To achieve this, the AI needs to recognise the need for handover and then execute the handover effectively. Handover between healthcare professionals is a recognised safety-critical task that remains surprisingly challenging and error prone in practice.²⁷ The use of structured communication protocols (eg, age–time–mechanism–injuries–signs–treatments) could improve the quality of handover even if challenges remain in their practical application.²⁸ Consideration should be given to the development of comparable approaches for the structured handover between AI and healthcare professionals.

While the intention of designers is to use AI to improve efficiency of workflows by taking over tasks from healthcare professionals, there is a danger that staff might get pulled into other activities instead or that the healthcare professional spends more time interacting with the AI. Lessons should be learnt from the introduction of other digital technologies, such as electronic health records, where it has been suggested that, for example, in emergency care physicians spend more time on data entry than on patient contact.²⁹ The impact of integrating AI into an already computer-focused patient encounter needs to be carefully considered.

The use of healthcare AI also raises significant ethical issues. Technical challenges, including the potential for bias in data, have been highlighted, and have been incorporated into international guidelines and reporting standards.³⁰ However, it is also important to address wider issues around fairness and impact on different stakeholder groups.³¹ At European level, the High-Level Expert Group on AI published ‘Ethics Guidelines for Trustworthy AI’.³² The guidelines are based on a fundamental rights impact assessment and operationalise ethical principles through seven key requirements: human agency and oversight;

technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being and accountability. HFE approaches can support addressing these ethical requirements through understanding stakeholders and their diverse needs and expectations.

BUILDING HFE CAPACITY

The systems perspective on healthcare AI set out in the CIEHF white paper is going to be instrumental in realising national AI strategies and delivering the benefits for patients and health systems. The digital transformation needs to be underpinned by HFE capacity within the health sector. Until very recently, there was no formal career structure for healthcare professionals with an interest in HFE. In the UK, this is changing with the recent introduction of both academic and learning-at-work routes towards accredited status of technical specialist or TechCIEHF (healthcare).³³ Enhancing the professionalisation of HFE knowledge among those with responsibility for quality improvement, patient safety and digital transformation can support healthcare organisations in making better informed AI adoption and implementation decisions.

There is also a need for funding bodies and regulators to require evidence that suitable HFE expertise is included in the design and evaluation of healthcare AI. Funding specifications frequently reflect only the technology-centric perspective of AI rather than reinforcing a systems approach. While inclusion of qualitative research to support scaling of healthcare AI from the lab to clinical environments is useful, it cannot replace the benefits of early inclusion of HFE expertise already during the design stage of AI technologies. Human behaviour is highly context dependent and adaptive as people navigate complexity and uncertainty and this needs to inform the design of AI to ensure that the use of AI in health and care systems is meaningful and safe. Regulators are trying to catch up on the technical AI expertise required, but the effective regulation of these technologies should also be supported through the recruitment of suitably qualified HFE professionals to establish appropriate interdisciplinary expertise in the advancement of AI technologies in healthcare.

Twitter Mark Sujan @MarkSujan

Contributors All authors contributed equally to the idea and drafting of the manuscript and reviewed and approved the final version.

Funding This work was supported in part (MS) by the Assuring Autonomy International Programme, a partnership between Lloyd’s Register Foundation and the University of York.

Competing interests All authors are coauthors of the Chartered Institute of Ergonomics and Human Factors white paper referred to in the manuscript. MS is a member of the Governing Council of the Chartered Institute of Ergonomics and Human Factors.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which



permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Mark Sujan <http://orcid.org/0000-0001-6895-946X>

REFERENCES

- 1 Peek N, Sujan M, Scott P. Digital health and care in pandemic times: impact of COVID-19. *BMJ Health Care Inform* 2020;27:e100166.
- 2 Joshi I, Morley J. *Artificial intelligence: how to get it right. putting policy into practice for safe data-driven innovation in health and care*. London: NHSX, 2019.
- 3 UK AI Council. *AI roadmap*. London, 2021.
- 4 Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digit Health* 2021;3:e195–203.
- 5 McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577:89–94.
- 6 Blomberg SN, Folke F, Ersbøll AK, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. *Resuscitation* 2019;138:322–9.
- 7 Komorowski M, Celi LA, Badawi O, et al. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med* 2018;24:1716–20.
- 8 Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated Conversational agent (Woebot): a randomized controlled trial. *JMIR Ment Health* 2017;4:e19.
- 9 Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting Standards, and claims of deep learning studies. *BMJ* 2020;368:m689.
- 10 Sujan M, Furniss D, Grundy K, et al. Human factors challenges for the safe use of artificial intelligence in patient care. *BMJ Health Care Inform* 2019;26:e100081.
- 11 Scott I, Carter S, Coiera E. Clinician checklist for assessing suitability of machine learning applications in healthcare. *BMJ Health Care Inform* 2021;28:e100251.
- 12 Greaves F, Joshi I, Campbell M, et al. What is an appropriate level of evidence for a digital health intervention? *The Lancet* 2018;392:2665–7.
- 13 Vasey B, Clifton DA, Collins GS, et al. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. *Nat Med* 2021;27:186–7.
- 14 Rivera SC, Liu X, Chan A-W, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *BMJ* 2020;370:m3210.
- 15 Sujan M, Baber C, Salmon P, et al. *Human factors and ergonomics in healthcare AI*. Wootton Waven: Chartered Institute of Ergonomics and Human Factors, 2021.
- 16 Carayon P, Schoofs Hundt A, Karsh B-T, et al. Work system design for patient safety: the SEIPS model. *Qual Saf Health Care* 2006;15:i50–8.
- 17 Carayon P, Wooldridge A, Hoonakker P, et al. SEIPS 3.0: Human-centered design of the patient journey for patient safety. *Appl Ergon* 2020;84:103033.
- 18 Vicente KJ. *Cognitive work analysis: toward safe, productive, and healthy computer-based work*. CRC Press, 1999.
- 19 Hollnagel E. Why is work-as-imagined different from work-as-done? In: Wears R, Hollnagel E, Braithwaite J, eds. *The resilience of everyday clinical work*. Farnham: Ashgate, 2015.
- 20 Parasuraman R, Riley V. Humans and automation: use, misuse, disuse, abuse. *Hum Factors* 1997;39:230–53.
- 21 Sarter NB, Woods DD, Billings CE. Automation surprises. In: Salvendy G, ed. *Handbook of Human Factors & Ergonomics*. Wiley, 1997: 1926–43.
- 22 Bainbridge L. Ironies of automation. *Automatica* 1983;19:775–9.
- 23 Ruskin KJ, Hueske-Kraus D. Alarm fatigue: impacts on patient safety. *Curr Opin Anaesthesiol* 2015;28:685–90.
- 24 Weld DS, Bansal G. The challenge of crafting intelligible intelligence. *Commun ACM* 2019;62:70–9.
- 25 Levinson DR. OEI-06-09-00091 Hospital incident reporting systems do not capture most patient harm. Washington, DC: Department of Health and Human Services, 2012.
- 26 Saenz MJ, Revilla E, Simón C. Designing AI systems with Human-Machine teams. *MIT Sloan Management Review* 2020;61:1–5.
- 27 Sujan MA, Chessum P, Rudd M, et al. Managing competing organizational priorities in clinical handover across organizational boundaries. *J Health Serv Res Policy* 2015;20:17–25.
- 28 Sujan MA, Chessum P, Rudd M, et al. Emergency care handover (ECHO study) across care boundaries: the need for joint decision making and consideration of psychosocial history. *Emerg Med J* 2015;32:112–8.
- 29 Hill RG, Sears LM, Melanson SW. 4000 clicks: a productivity analysis of electronic medical records in a community hospital ED. *Am J Emerg Med* 2013;31:1591–4.
- 30 Challen R, Denny J, Pitt M, et al. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf* 2019;28:231–7.
- 31 Wawira Gichoya J, McCoy LG, Celi LA, et al. Equity in essence: a call for operationalising fairness in machine learning for healthcare. *BMJ Health Care Inform* 2021;28:e100289.
- 32 High-Level Expert Group On Artificial Intelligence. *Ethics guidelines for trustworthy AI*. Brussels: European Commission, 2019.
- 33 Sujan M, Pickup L, Bowie P, et al. The contribution of human factors and ergonomics to the design and delivery of safe future healthcare. *Future Healthcare Journal* 2021;8:e574–9.