

<https://doi.org/10.1038/s41746-026-02410-1>

# Flight rules for clinical AI: lessons from aviation for human-AI collaboration in medicine

Check for updates

Ariel Yuhan Ong<sup>1,2,3</sup> ✉, David A. Merle<sup>1,2,4</sup>, Andreas Pollreisz<sup>5</sup>, Siegfried K. Wagner<sup>1,2</sup>, Mertcan Sevgi<sup>1,2</sup>, Pearse A. Keane<sup>1,2</sup>, Roman Huemer<sup>6</sup>, Julian Oehling<sup>6</sup>, Markus Jäger<sup>6</sup> & Josef Huemer<sup>1,2,7</sup> ✉

The parallels between medicine and aviation are well-recognised. The aviation industry's early experience with automation improved safety and efficiency, but simultaneously introduced new vulnerabilities and occasionally created misplaced trust in complex systems. Aviation has developed a robust safety framework in response to these costly lessons. In this Perspective, which draws from the experiences of clinicians and aviation experts, we argue that it is now time for the medical community to consider how we can learn from these lessons as artificial intelligence (AI) becomes increasingly integrated into clinical care. We propose that this requires a shift in perspective from AI as “autopilot” to collaboration with a “digital copilot”, as well as considerations of practicalities such as scenario-based training, clinician benchmarking, and minimum unaided practice, with the ultimate aim of optimising human-AI collaboration to improve patient care.

The parallels between medicine and the aviation industry have long been recognised. Both involve high-stakes decision-making (often under uncertainty), and both emphasise a strong culture of safety and multidisciplinary team working to safeguard human life<sup>1</sup>. Given these similarities, medicine has borrowed much from the aviation industry over the past few decades: the origins of surgical checklists, safety time-outs, the “just culture” of incident reporting and analysis, and human factors simulation training can all be traced back to flight safety (Fig. 1)<sup>1,2</sup>, which continues to inspire other initiatives such as sustainable quality and safety improvement practices in healthcare<sup>3</sup>.

While significant quality gains have been achieved in healthcare over the same period of time, variability in practice and an appreciable morbidity and mortality from preventable medical errors persist<sup>4,5</sup>, even as airline passenger safety continues to improve at pace, with mortality rates halving every decade<sup>6</sup>. Part of this discrepancy might lie with contextual differences, such as the biological variability inherent across patients (as opposed to standardised aircraft), and patients being autonomous agents (as opposed to passive passengers on a flight). In addition, the field of aviation has well-resourced and central governance structures that support safety culture and accountability in aviation<sup>7</sup>, in contrast to healthcare, where analogous structures may be more fragmented or less well funded. However, despite

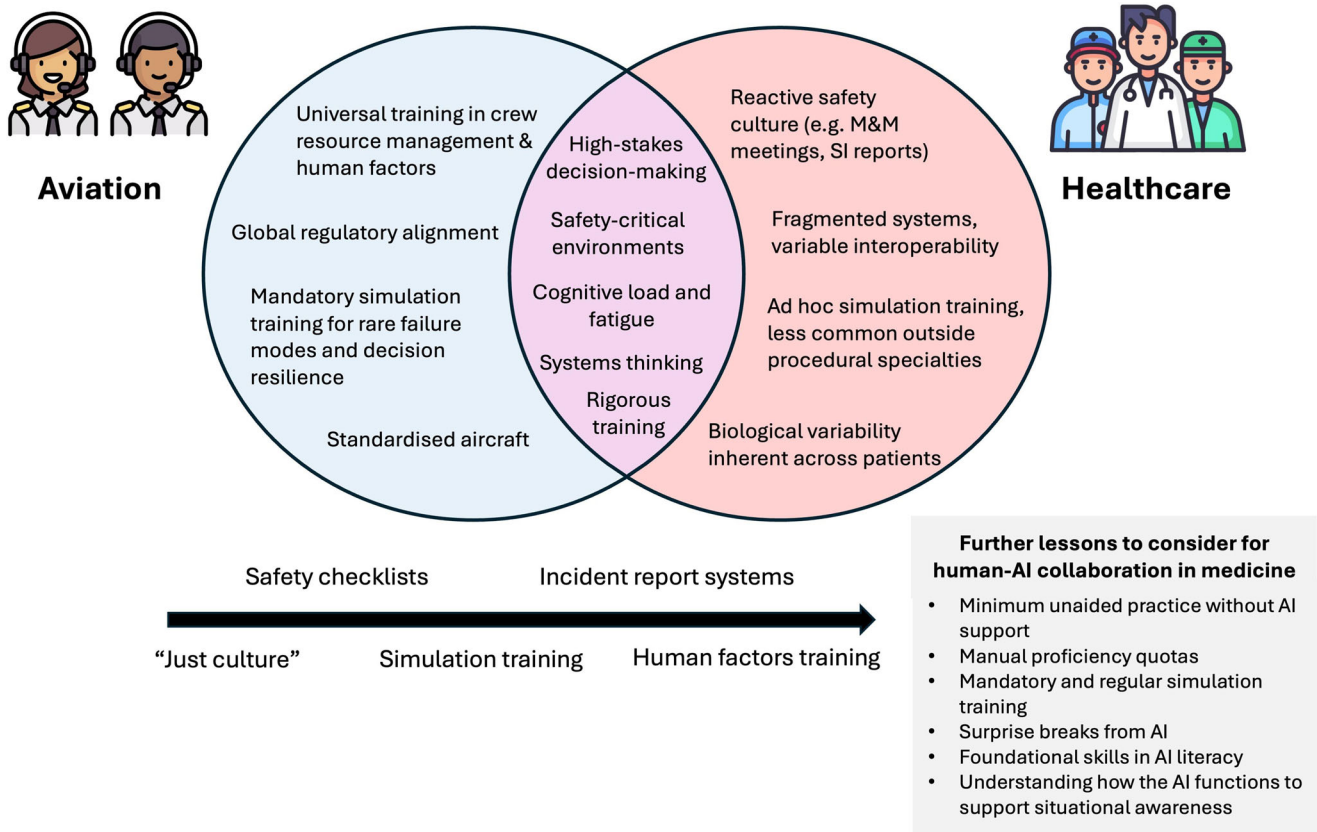
some inherent dissimilarities, aviation's track record demonstrates that meaningful progress is possible when safety is prioritised in a systematic and regulated manner.

These parallels have only deepened as artificial intelligence (AI) becomes increasingly integrated into clinical care. While the aviation industry's early experience with automation improved safety and efficiency, it simultaneously introduced new vulnerabilities such as loss of manual skills, reduction of situational awareness, and misplaced trust in complex systems. Aviation has developed a robust safety framework around automation in response to these costly lessons. In this Perspective, we argue that with AI poised to profoundly reshape medical workflows, we must now consider how we can learn from these lessons to avoid making the same mistakes. Here, it is important to note that healthcare is not a single homogenous system, but comprises multiple different environments, settings, and clinical contexts<sup>8</sup>. Accordingly, the parallels drawn with aviation are not intended as a blueprint, but as a source of transferable principles that can be applied across different key areas where relevant.

## Early lessons from automation in aviation

Automation was introduced in aviation to improve safety and consistency. However, the “automation paradox” soon became apparent. This

<sup>1</sup>NIHR Moorfields Biomedical Research Centre, London, UK. <sup>2</sup>Institute of Ophthalmology, University College London, London, UK. <sup>3</sup>Oxford Eye Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, UK. <sup>4</sup>Center of Ophthalmology, University Hospital Tübingen, Tübingen, Germany. <sup>5</sup>Department of Ophthalmology and Optometry, Medical University of Vienna, Vienna, Austria. <sup>6</sup>Flight Safety Department, Lufthansa German Airlines, Frankfurt, Germany. <sup>7</sup>Department of Ophthalmology and Optometry, Kepler University Hospital, Linz, Austria. ✉e-mail: [ariel.ong@nhs.net](mailto:ariel.ong@nhs.net); [josef.huemer1@nhs.net](mailto:josef.huemer1@nhs.net)



**Fig. 1 | Conceptual relationship between the aviation and healthcare industries.** The Venn diagram compares the similarities and differences. The directional arrow illustrates the historical adoption of key safety concepts such as safety checklists, simulation training, and “just culture” from aviation in healthcare thus far. The grey

box highlights further lessons that healthcare can consider adopting from aviation, in order to aid human-AI collaboration. M&M morbidity and mortality, SI serious incident, AI artificial intelligence.

phenomenon describes how increasing automation erodes human skills and awareness, which begets the need for more automation, resulting in a vicious cycle that further increases the dependency on these systems<sup>9</sup>. Such behaviour led a senior pilot to coin the term “children of the magenta line”—a metaphor to describe younger pilots who became dependent on the use of autopilot and computer-generated ‘magenta lines’ displayed on their instruments, and who lacked the skills and confidence to fly the plane manually<sup>9</sup>.

Although accidents and the ensuing incident investigations catalysed a shift in aviation safety philosophy starting in the late 1970s, it took decades of incident investigations and persistent advocacy for policymakers and aircraft manufacturers to acknowledge that flight safety required investment in human capabilities and not just technology alone. The resulting reforms (spanning the late 1970s and 80s and becoming embedded in practice by the late 1990s) were grounded in the recognition that automation changes rather than eliminates human responsibility<sup>10</sup>. However, subsequent accidents, including the example of Asiana Airlines Flight 214 in San Francisco in 2013, demonstrates that this challenge persists even within relatively mature safety systems. In that incident, the pilots relied on the auto-throttle system to regulate engine thrust and airspeed, but failed to realise that they had misconfigured the automation. They fully relied on the automation to monitor the aircraft’s speed, which consequently became too slow to sustain flight, destroying the aircraft, killing four passengers and injuring many more.

Automation improves flight performance and reduces cognitive load, which is beneficial in the vast majority of situations where systems behave as expected. However, it can be catastrophic in rare instances of errors or unanticipated outcomes, especially as higher levels of automation comes at the cost of diminished situational awareness such as reduced attention to in-flight instruments<sup>11</sup>. De-skilled pilots, who are overly used to operating with

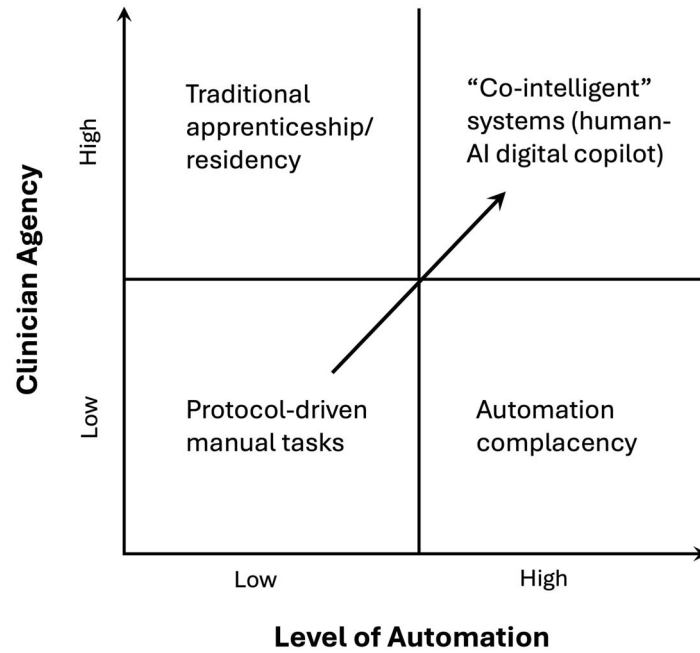
automated systems, are often slower to detect anomalies, less prepared to take manual control, or more reliant on automated cues that may themselves be erroneous. Requiring pilots to actively measure and record key metrics such as altitude, fuel status, and time-to-destination and compare them to the pre-calculated flight plan at regular intervals are simple examples of actively seeking engagement to maintain an overview of the flight situation.

In medicine, there is the same risk that AI algorithms can enhance performance and efficiency or reduce workload under routine conditions but undermine clinicians’ vigilance, particularly when errors, system failures, or unexpected situations occur. Parallel examples are starting to emerge in clinical care—a recent study found that endoscopists who regularly performed AI-assisted colonoscopy performed worse at detecting adenomas—a 6.0 percentage point absolute reduction<sup>12</sup>, which has clinically significant implications for interval colorectal cancer risk<sup>13</sup>—after their AI assistant was removed from the equation, suggesting a potential dependency effect. Clinicians have also expressed other concerns about deskilling with AI exposure in terms of identifying clinical signs from physical examination<sup>14</sup>, even as inadequate clinical skills has been identified as a key cause of medical errors<sup>15</sup>. Instruction in the art of history-taking and examination has been central to the practice of medicine since the days of William Osler<sup>16</sup>, but increasing dependence on AI may accelerate the erosion of these skills as well as clinical reasoning—the iterative process of hypothesis generation, observation, and evidence synthesis at the heart of medicine<sup>17,18</sup>.

### From autopilot to digital copilot

Just as the introduction of automation changed the roles and responsibilities of pilots, the increasing integration of AI may also influence how clinicians

**Fig. 2 | Matrix illustrating how clinician agency may interact with levels of automation in medical artificial intelligence, accompanied by an explanation of each quadrant.** The bottom quadrants of the matrix represent the risk states of disempowered manual practice (low automation/ low agency) and automation complacency (high automation/ low agency). The upper quadrants represent more optimal configurations, from the traditional medical apprenticeship model (low automation/ high agency) to co-intelligent systems (high automation/ high agency), where clinicians maintain oversight. The arrow represents the trajectory towards co-intelligence and achieving automation efficiency without eroding human expertise.



Quadrant	Explanation	Examples
Low agency / Low automation	Clinicians follow rigid protocol-driven pathways with limited autonomy; professional judgment is replaced by adherence to pre-defined protocols	Protocolised triage decisions or administrative work, or protocol-driven routine low-risk hypertension or diabetes reviews
High agency / Low automation	Clinicians perform all diagnostic and clinical decision-making tasks without additional technological support, developing deep experiential and procedural expertise	Surgical apprenticeship models where skill transfer depends on observation and repetition, or ECG or X-ray interpretation taught through pattern recognition and feedback without AI support
Low agency / High automation	Highly automated workflows with minimal clinician oversight, or where clinicians defer to AI outputs	Automated AI-generated radiology reports unsupported by human verification
High agency / High automation	Clinicians work collaboratively with AI and maintain oversight within a broader ecosystem	Clinical decision-support systems that highlight risks but leave final judgment to the clinician

practice medicine. Despite the growing tension between automation versus augmentation in medical AI<sup>19</sup>, in practice, very few algorithms truly function independently, as any AI health technology has to operate within the multidisciplinary and multilayered healthcare ecosystem. Varying degrees of human-AI teaming are required for different tasks and contexts, meaning that in reality, “autonomous” and “clinical decision support” algorithms are not dichotomous but are situated on a continuum.

Recognising this reframes AI not as an autopilot that replaces human input, but as a “digital copilot” that supports it. We argue that this conceptual shift provides a more useful model for medicine. Instead of passive oversight or a perfunctory “human-in-the-loop” design, clinicians and AI act as partners or collaborators, either as single human-AI units or in larger teams. Clinicians remain as the “pilot-in-command” and accountable for overall judgement, while the digital copilot can contribute consistency and precision. This is the basis of co-intelligence: a synergistic relationship that combines human contextual reasoning with algorithmic speed and pattern recognition<sup>20</sup>, with clinicians and AI systems operating as interdependent partners within a human-AI team<sup>21</sup>.

This framing also aligns with patient preferences—a large multinational survey found that patients consistently favoured a collaborative diagnostic approach led by clinicians when AI is involved in their care, and

expressed reluctance for AI to reduce physician-patient interaction or replace human physicians altogether<sup>22</sup>. At the same time, the AI safety and ethics literature also cautions that accountability in AI-mediated care is inherently socio-technical and cannot rest with clinicians alone. Instead, clinical judgement should remain as a core component of dynamic approaches to safety assurance that extend beyond design-time safeguards to include developers, systems safety engineers, and organisations across real-world use<sup>23</sup>.

This evolving relationship between humans and AI can be understood through the interplay between the level of automation and clinician agency, which draws from prior human factors research<sup>24,25</sup> (Fig. 2). In low-automation/high-agency settings, clinical work resembles the traditional apprenticeship model. As automation increases, efficiency improves but clinicians may drift toward automation complacency and lose sight of situational awareness<sup>24</sup>. Conversely, rigid protocol-driven workflows can also diminish agency<sup>26,27</sup>. Both extremes limit clinicians’ ability to exercise their clinical judgement. We propose that the optimal configuration lies in high-automation/high-agency systems, where humans and AI function as co-intelligent partners. The challenge, then, is not to resist AI but to ensure that its adoption does not erode the very skills it seeks to support.

## The way forwards: translating aviation's lessons into healthcare

Here, aviation offers another useful precedent—rather than striving to perfect automation alone, flight safety prioritises addressing human fallibility within increasingly automated systems<sup>28</sup>. This same principle is also applicable to medical AI (Fig. 1). We propose five key considerations to navigate these changes, inspired by lessons drawn from flight safety.

### Addressing the risk of skill erosion across generations by benchmarking clinicians and monitoring performance without AI

One of aviation's most costly lessons is that skill decay secondary to automation is predictable and reversible, but only if this is actively addressed. To mitigate the risk of losing manual flight proficiency, pilots are actively encouraged to maintain their manual flying skills on routine flights under appropriate circumstances, for example when the workload is low or there is good weather. Similarly, as AI becomes more routine, minimum unaided practice without AI support should be considered for practicing clinicians. Ironically, while there is significant interest in benchmarking AI algorithms, the same level of enthusiasm has not applied to human clinicians, particularly in non-surgical specialties. The national diabetic retinopathy screening programme in the UK is one of the few examples where human graders are benchmarked on their diagnostic performance on a regular basis to improve quality and consistency<sup>29</sup>, but this is not done in routine clinical practice once clinicians have completed their specialty board exams (or equivalent). It is impossible to assess de-skilling without measurements. Institutions integrating AI (either “autonomous” or decision support systems) could consider monitoring human-AI concordance rates to assess for overreliance and complacency, as well as incorporating manual proficiency quotas, with performance on these no-AI cases providing a baseline benchmark for detecting drift for both the AI and the human user.

### Reconsidering how medical training is designed and structured to facilitate the development of fundamental cognitive and procedural skills before AI is introduced

For younger clinicians trained in an era where AI is increasingly ubiquitous and automating routine clinical tasks, the more fundamental concern is not “deskilling” but “never skilling” or “mis-skilling”<sup>30</sup>. A growing body of evidence suggests that learners develop shallower knowledge with AI-based tools as opposed to self-directed learning<sup>31</sup>. Addressing this risk requires intentional design in medical education, including careful rethinking of how clinical training is structured in an AI-rich environment. Early exposure should prioritise independent reasoning and experiential learning before automation is introduced, meaning that access to AI tools in clinical care should perhaps be restricted until basic clinical competency has been developed, thus allowing AI to function as a scaffold rather than a substitute for skill development. Alongside this, longitudinal monitoring of cognitive and procedural competence is needed to detect early signs of “never skilling” or “mis-skilling”, particularly in specialties where AI use is becoming increasingly routine. Trainees must remain active participants in diagnostic and clinical decision-making processes rather than passive recipients of AI outputs in order to avoid producing clinicians who are competent at using AI rather than understanding patients.

Ultimately, all of these issues are not unique to AI; similar debates have emerged with earlier waves of automation in other safety-critical industries<sup>25</sup>, as well as with previous generations of medical technologies<sup>32</sup>. The enduring tension lies in how to realise the benefits of these tools while helping doctors maintain agency, without eroding the human expertise on which these systems depend, as clinical judgement should remain the ultimate safety redundancy in AI-enabled practice. The rise of generative AI—particularly large language models (LLMs)—presents an additional challenge due to their occasional tendency to generate hallucinations, or plausible-sounding outputs that may not be fully grounded in factual evidence<sup>33</sup>. Confidently incorrect outputs may be misinterpreted as authoritative guidance, and justification chains by the model may not reflect valid reasoning. This stands in stark contrast to

previous generations of safety-critical automation in aviation, which were designed with explicit failure modes and standardised alerts, which could signal when performance limitations are reached or exceeded. For example, autoland systems specify maximum allowable crosswind components and environmental conditions; pilots are prohibited from using such systems outside these certified parameters<sup>34</sup>. Whether and how generative AI can be engineered to exhibit comparable constraint, to communicate uncertainty, and to operate within clearly defined boundaries remains an open question that we need to consider as deployment in healthcare settings becomes more widespread.

### Defining foundational skills in AI literacy and the minimum digital and technical competencies that clinicians using AI should possess

To navigate these changes safely, medical curricula should also consider foundational skills in AI literacy, which some organisations are now beginning to articulate, albeit without significant consistency or traction as yet<sup>35,36</sup>. It is not enough to be able to operate an AI tool. Rather, clinicians should also be trained to use the technology appropriately and optimally<sup>32</sup>. To this end, understanding general concepts about AI (such as LLMs being token predictors rather than an all-knowing oracle) is an essential starting point, as is an appreciation of AI limitations, failure modes, and susceptibility to bias, which has been acknowledged as a substantial concern given the risk of amplifying existing health inequalities if left unrecognised<sup>37,38</sup>.

In order to achieve this, there is a clear need for consensus around the minimum digital and technical competencies necessary for clinicians operating in AI-assisted clinical settings<sup>39</sup>. In addition, competency is not static but requires regular maintenance, especially given the rapid pace of AI advancement and the associated risk of obsolescence. As such, careful deliberation about how much and what to teach in medical school, post-graduate training, and in continuing professional development is also essential.

### Incorporating regular simulation-based training for human-AI teams to monitor dependence and instill situational awareness, and train clinicians when to trust and question their AI support

Robust training and simulation for human-AI teaming should also be considered. The aviation industry mandates line-oriented flight training<sup>40</sup>, where it is compulsory for pilots to spend a set number of hours a year in high-fidelity simulators to develop core competencies such as situational awareness. For example, the European Union mandates 48 h every 3 years, although the specific requirements depend on jurisdiction<sup>41</sup>. Evidence-based training for skill assessment and development rather than task completion has historically been the educational gold standard. This is a familiar concept in some fields of medicine such as ophthalmology, where “vitreous loss safety drills” have been introduced to help familiarise surgeons with the management of uncommon but potentially serious intraoperative complications, in order to equip them with the ability to handle them when they do arise<sup>42</sup>; however, this is often conducted on an ad hoc basis rather in a systematic or regulated manner. In addition, despite the increasing recognition that maintaining team-level situational awareness and shared mental models to develop high-functioning teams can improve patient management in other specialties such as anaesthesia<sup>43</sup>, this concept is uncommon in many other areas of medicine.

We propose going beyond this to consider scenario-based simulation training that recreate contexts where AI fails in order to monitor human dependence on AI systems, and to train clinicians when to trust, when to question, and when to override their “digital co-pilot”. In settings where AI use has become routine, introducing unannounced “surprise breaks” from AI can help assess readiness to operate safely without automation<sup>44</sup>. This should also incorporate structured debriefs where clinicians reflect on AI supported decisions in order to reinforce accountability and situational awareness, akin to the virtual waypoints required of pilots.

### Cultivating operational understanding of AI function in clinical practice to support safe engagement and disengagement

Finally, there is a need for clinicians to develop an operational understanding of how the AI systems they use function in clinical practice. For the average clinician, the goal is not to understand the inner workings of a neural network, but to be able to appreciate the rationale for its decision in operational terms (e.g. highlighting the factors that contributed to a risk score, or pixels in an image that contributed to its decision), in order to support safe engagement with AI outputs or oversight without overwhelming the user with extraneous detail. This ties in with the “explainability conundrum”, which captures a central tension between transparency and clinical utility. While there is significant interest in methods to develop explainable AI (XAI)<sup>45</sup>, overly detailed or poorly calibrated explanations can paradoxically worsen outcomes by leading to unwarranted trust and misinterpretation of model behaviour<sup>45,46</sup>. Explainability should perhaps be judged by its contribution to operational understanding, rather than by the depth or technical fidelity of the explanations alone.

This emphasis on operational understanding aligns with one of the “golden rules” for pilots, which is the need to understand the automated system’s function at all times. When that understanding is lost, the procedure is clear: to reduce the level of automation step by step, or to disengage it entirely, and to regain manual control until situational awareness is restored. A parallel can be drawn with closed-loop systems in anaesthesia, where clinicians must understand the operating limits and know when to override or switch to manual mode if patient parameters drift outside expected bounds<sup>47</sup>.

### Charting a new course towards human-AI collaboration in medicine

Aviation’s experience with automation shows that safety is not achieved by replacing humans in complex systems but by defining their role within them. Ultimately, learning from aviation should go beyond adopting simulation training or checklists alone—human-machine (or human-AI) interaction in medicine should be treated as a discipline in its own right, with defined competencies, measurable outcomes, and explicit recovery mechanisms when the AI fails. Accordingly, safety should be viewed as an emergent property of the human-AI team rather than the technology alone, and regulation will need to evolve accordingly. As such, beyond simply certifying AI as a medical device, as AI becomes more prevalent in healthcare, we should also consider how we regulate the human-AI dyad, and how we maintain competence, accountability, and situational awareness in this context. As we begin to deploy AI in healthcare, considering how we investment in governance and infrastructure will also be essential to avoid reproducing aviation’s early failures and the decades of work that have led to its contemporary safety systems. The future of medical AI will not be defined by how much we automate, but by how well we learn to collaborate in creating a synergistic partnership that augments rather than replaces human judgement, without ever losing sight of the ultimate aim of improving patient care.

### Data availability

No datasets were generated or analysed during the current study.

Received: 20 November 2025; Accepted: 25 January 2026;

Published online: 31 January 2026

### References

- Kapur, N., Parand, A., Soukup, T., Reader, T. & Sevdalis, N. Aviation and healthcare: a comparative review with implications for patient safety. *JRSM Open* **7**, 2054270415616548 (2015).
- Clay-Williams, R. & Colligan, L. Back to basics: checklists in aviation and healthcare. *BMJ Qual. Saf.* **24**, 428–431 (2015).
- Webster, C. S., Henderson, R. & Merry, A. F. Sustainable quality and safety improvement in healthcare: further lessons from the aviation industry. *Br. J. Anaesth.* **125**, 425–429 (2020).
- Hogan, H. et al. Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study. *BMJ Qual. Saf.* **21**, 737–745 (2012).
- Makary, M. A. & Daniel, M. Medical error—the third leading cause of death in the US. *BMJ* **353**, i2139 (2016).
- Barnett, A. & Reig Torra, J. Airline safety: still getting better?. *J. Air Transp. Manag.* **119**, 102641 (2024).
- ICAO. *State of Global Aviation Safety: ICAO Safety Report 2025 Edition* (ICAO, 2025).
- Vincent, C. & Amalberti, R. *Safer Healthcare* (Springer International Publishing, 2016).
- Mingle, K. Children of the Magenta (Automation Paradox, pt. 1). 99% Invisible. <https://99percentinvisible.org/episode/children-of-the-magenta-automation-paradox-pt-1/> (2015).
- Durepaire, X. Manual Flying Policy for Operation. [Video]. Flight Ops Regional Conferences <https://www.airbus-win.com/wp-content/uploads/2020/02/manual-flying-policy-st-en-1.mp4> (2019).
- Causse, M., Mercier, M., Lefrançois, O. & Matton, N. Impact of automation level on airline pilots’ flying performance and visual scanning strategies: a full flight simulator study. *Appl. Ergon.* **125**, 104456 (2025).
- Budzyń, K. et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. *Lancet Gastroenterol. Hepatol.* **10**, 896–903 (2025).
- Corley, D. A. et al. Adenoma detection rate and risk of colorectal cancer and death. *N. Engl. J. Med.* **370**, 1298–1306 (2014).
- Natali, C., Marconi, L., Dias Duran, L. D. & Cabitza, F. AI-induced deskilling in medicine: a mixed-method review and research agenda for healthcare and beyond. *Artif. Intell. Rev.* **58**, 356 (2025).
- Verghese, A., Charlton, B., Kassirer, J. P., Ramsey, M. & Ioannidis, J. P. A. Inadequacies of physical examination as a cause of medical errors and adverse events: a collection of vignettes. *Am. J. Med.* **128**, 1322–1324.e3 (2015).
- Gomes, M., da, M. & Haynes, R. B. William Osler (1849–1919) at the roots of evidence-based medicine. *Can. J. Gen. Intern. Med.* **14**, 23–27 (2019).
- Honavar, S. G. Artificial intelligence in ophthalmology—machines think!. *Indian J. Ophthalmol.* **70**, 1075–1079 (2022).
- Corrao, S. & Argano, C. Rethinking clinical decision-making to improve clinical reasoning. *Front. Med* **9**, 900543 (2022).
- Grüning, M., Trenz, M. & Santos, D. P. dos. Disentangling the tension between automation and augmentation with artificial intelligence in healthcare. *AMCIS 2023 Proceedings* [https://aisel.aisnet.org/amcis2023/sig\\_health/sig\\_health/8](https://aisel.aisnet.org/amcis2023/sig_health/sig_health/8) (2023).
- Mollick, E. *Co-Intelligence: Living and Working with AI* (Penguin Random House, 2024).
- Berretta, S. Defining human-AI teaming the human-centered way: a scoping review and network analysis. *Front. Artif. Intell.* **6**, 1250725 (2023).
- Busch, F. et al. Multinational attitudes toward AI in health care and diagnostics among hospital patients. *JAMA Netw. Open* **8**, e2514452 (2025).
- Habli, I., Lawton, T. & Porter, Z. Artificial intelligence in health care: accountability and safety. *Bull. World Health Organ.* **98**, 251–256 (2020).
- Parasuraman, R. & Manzey, D. H. Complacency and bias in human use of automation: an attentional integration. *Hum. Factors* **52**, 381–410 (2010).
- Bainbridge, L. Ironies of automation. *Automatica* **19**, 775–779 (1983).
- Lau, E. W., Bonnemeier, H. & Baldauf, B. Misuse of guidelines could disadvantage and harm patients. *J. Evid. Based Med.* **17**, 705–707 (2024).
- Al-Najjar, Z. Doctors’ clinical judgment risks being sacrificed on the altar of protocol driven care. *BMJ* q2723 <https://doi.org/10.1136/bmj.q2723> (2024).
- Muecklich, N., Sikora, I., Paraskevas, A. & Padhra, A. Safety and reliability in aviation—a systematic scoping review of normal accident

- theory, high-reliability theory, and resilience engineering in aviation. *Saf. Sci.* **162**, 106097 (2023).
29. Lacey, L. Making the grade—PHE screening. <https://phescreeing.blog.gov.uk/2015/09/21/making-the-grade/> (2015).
  30. Abdunour, R.-E. E., Gin, B. & Boscardin, C. K. Educational strategies for clinical supervision of artificial intelligence use. *N. Engl. J. Med.* **393**, 786–797 (2025).
  31. Melumad, S. & Yun, J. H. Experimental evidence of the effects of large language models versus web search on depth of learning. *PNAS Nexus* **4**, pgaf316 (2025).
  32. Lu, J. Will medical technology deskill doctors?. *IES* **9**, 130 (2016).
  33. Huang, L. et al. A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions. *ACM Trans. Inf. Syst.* **43**, 1–55 (2025).
  34. US Department of Transportation Federal Aviation Administration. *Airplane Flying Handbook*. [https://www.faa.gov/regulations\\_policies/handbooks\\_manuals/aviation/airplane\\_handbook](https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/airplane_handbook) (2021).
  35. Rincón, E. H. H. et al. Mapping the use of artificial intelligence in medical education: a scoping review. *BMC Med. Educ.* **25**, 526 (2025).
  36. HDRUK & Medical Schools Council. Data science in the medical curriculum: equipping medical students for the digital age. <https://www.medschools.ac.uk/wp-content/uploads/2025/05/data-science-in-the-medical-curriculum-1.pdf> (2025).
  37. Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).
  38. Hasanzadeh, F. et al. Bias recognition and mitigation strategies in artificial intelligence healthcare applications. *npj Digital Med.* **8**, 154 (2025).
  39. Schuitemaker, L., Drog, J., Benders, M. & Jongma, K. Physicians' required competencies in AI-assisted clinical settings: a systematic review. *Br. Med. Bull.* **153**, Idae025 (2025).
  40. Hamman, W. R. Chapter 8—Line oriented flight training (LOFT): the intersection of Technical and human factor crew resource management (CRM) team skills. in *Crew Resource Management* 2nd edn (eds Kanki, B. G., Helmreich, R. L. & Anca, J.) 233–263 (Academic Press, 2010).
  41. European Union Aviation Safety Agency. Oversight guidance for the transition to EBT Implementation. <https://www.easa.europa.eu/en/downloads/131390/en> (2025).
  42. Lockington, D., Belin, M. & McGhee, C. N. J. The need for all cataract surgeons to run a regular vitreous loss fire drill. *Eye* **31**, 1120–1121 (2017).
  43. Weller, J. M., Mahajan, R., Fahey-Williams, K. & Webster, C. S. Teamwork matters: team situation awareness to build high-performing healthcare teams, a narrative review. *Br. J. Anaesth.* **132**, 771–778 (2024).
  44. Hunter, R., Moulange, R., Bernardi, J. & Stein, M. Monitoring human dependence on AI systems with reliance drills. arXiv <https://arxiv.org/abs/2409.14055> (2024).
  45. Wong, C. Y. T., Antaki, F., Woodward-Court, P., Ong, A. Y. & Keane, P. A. The role of saliency maps in enhancing ophthalmologists' trust in artificial intelligence models. *Asia Pac. J. Ophthalmol.* 100087 <https://doi.org/10.1016/j.apjo.2024.100087> (2024).
  46. van Royen, F. S. et al. In humble defence of unexplainable black box prediction models in healthcare. *J. Clin. Epidemiol.* 112013 <https://doi.org/10.1016/j.jclinepi.2025.112013> (2025).
  47. Coeckelenbergh, S. et al. Closed-loop anesthesia: foundations and applications in contemporary perioperative medicine. *J. Clin. Monit. Comput.* **38**, 487–504 (2024).

## Acknowledgements

A.Y.O. is supported by a National Institute for Health Research (NIHR)—Moorfields Eye Charity (MEC) Doctoral Fellowship (NIHR303691). D.A.M. is supported by an EURETINA Retinal Medicine Clinical Research Grant. P.A.K. is supported by a UK Research & Innovation Future Leaders Fellowship (MR/T019050/1) and Moorfields Eye Charity with The Rubin Foundation Charitable Trust (GR001753). The views expressed in this publication are those of the authors and not necessarily those of the abovementioned funding bodies, nor the authors' employers.

## Author contributions

A.Y.O., D.A.M., and J.H. conceptualised the manuscript. A.Y.O. wrote the original draft. All authors (A.Y.O., D.A.M., A.P., S.K.W., M.S., P.A.K., R.H., J.O., M.J., J.H.) reviewed and critically revised subsequent drafts, and have read and approved the submitted manuscript.

## Competing interests

P.A.K. is a co-founder of Cascader Ltd. and has acted as a consultant for Retina Consultants of America, Topcon, Roche, Boehringer-Ingelheim, and Bitfount and is an equity owner in Big Picture Medical. He has received speaker fees from Zeiss, Novartis, Gyroscope, Boehringer-Ingelheim, Apellis, Roche, AbbVie, Topcon, and Hakim Group. He has received travel support from Bayer, Topcon, and Roche, and has attended advisory boards for Topcon, Bayer, Boehringer-Ingelheim, RetinAI, and Novartis. J.H. has received travel support from Bayer and Roche, and speaker fees from Zeiss, Roche, and Bayer. The remaining authors do not have any conflicts of interest to declare.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41746-026-02410-1>.

**Correspondence** and requests for materials should be addressed to Ariel Yuhan Ong or Josef Huemer.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026