

Invited State-Of-The-Art Review

Risks of automation in medicine - a review article for the obstetrics case

Joris Fournel¹, Aasa Feragen¹ & Martin Tolsgaard²

1) Department of Applied Mathematics and Computer Science, Visual Computing - Technical University of Denmark (DTU), Denmark, 2) Department of Gynaecology, Fertility and Obstetrics&&;, The Juliane Marie Centre, Copenhagen University Hospital, Denmark

Dan Med J 2026;73(4):A10250852. doi: 10.61409/A10250852

ABSTRACT

The current overoptimism surrounding AI threatens to disrupt clinical environments. We discuss the risks related to invasive automation and describe how AI performance exaggeration may lead to clinical errors. We also investigate cognitive damage to clinicians: automation bias and de-skilling, by exploring their relation to well-defined psychological principles. Finally, we propose solutions for automation that actually benefit medicine: generalise model bias assessments and quality control, and ensure that AI integration does not erode clinicians' clinical expertise.

KEY POINTS

- AI overoptimism may disrupt clinical environments.
- The assumption of inherent machine superiority over clinicians is increasingly promoted, with potentially detrimental consequences for clinical practice.
- Researchers are incentivised to exaggerate AI performance. This constitutes a recipe for clinical errors.
- Current AI systems pose a major, documented, yet overlooked threat to clinicians' cognitive capacities.

Will "artificial intelligence" (AI) systems benefit or disrupt clinical practice? Is the automation of clinical decisions going to improve or worsen clinical care? Are AI systems going to make doctors more or less competent?

The current overoptimistic atmosphere is detrimental to answering these questions. Claims of extraordinary AI performance far outpace reports of negative results, which are hardly ever published [1]. On social media, positive sentiments towards AI in medical imaging are expressed

five times more often than negative sentiments [2]. Some described this trend as “hype” [3], whereas Sam Altman, the CEO of OpenAI, characterised the broader AI atmosphere as a “bubble” [4].

Hype and enthusiasm are emotional states that distort our perception of reality, leading us to overlook dangers in hasty decisions. Therefore, this work proposes a sober examination of the potential dangers of automation, which is urgently needed to avoid disrupting clinical environments as fragile as the human lives they care for.

We discuss when introducing automation – a machine performing a task instead of a clinician – may harm clinics: automation is not justified a priori; model errors lead to clinical errors (technical risks); the judgment of the clinicians is impaired by AI usage (cognitive risks).

The main concern uniquely addressed in this paper includes concepts such as de-skilling and, expanding beyond existing work [5], their relation to cognitive psychology and their effect on longer-term complications for clinicians.

Specifically, we focus on AI in obstetrics as a use case that represents a clinical specialty that draws on imaging, surgery and decision-making. Whereas most published literature involving AI in medicine has been rooted in radiology or pathology, AI and automation are increasingly used in clinical specialities such as obstetrics, cardiology and surgery.

Is automation always a good idea?

Before risking introducing a machine, one should have solid reasons to do so. Unjustified automation will always do more harm than good. For example, focusing on the automated task in isolation from its broader environment may cause us to overlook a downstream negative impact. Furthermore, the arguments provided when introducing automation may reinforce a false and insidious narrative about AI or doctors. Therefore, reviewing how researchers generally motivate new automations may serve to alert us to existing trends.

The body of research in obstetrics provides a good basis for analysing these arguments. Automated models have been developed for probe guidance [6], standard plane identification [7], foetal biometric measurement [8], Doppler information extraction [9], anomaly highlighting [10], gestational age prediction [11], birth weight prediction [12] and prognosis including preterm [13], intra-uterine growth restriction [14] and pre-eclampsia prediction [15]. Deep learning has also been applied in foetal cardiotocography and foetoscopic surgery [16, 17].

Here, we review and classify authors’ justifications for developing models found in this literature, ordered from what we consider the most legitimate to the most questionable, both in terms of validity and overshadowed consequences:

1. Screening and risk stratification improved by machines. Obstetric medicines occasionally rely on simple (but powerful) risk stratification parameters. Automatic models can help clinicians discover and extract new risk parameters - for example, by refining baseline cervical-length

thresholds in preterm birth screening [13].

2. Improve clinician training by incorporating automated support systems. For example, Lei et al. reported that a trainee cohort achieved prenatal screening quality requirements in significantly fewer training cycles when assisted by an automated system. Similarly, improved operator performance in perinatal ultrasound screening has been described [18]. This is a particularly desirable application because it strengthens clinical expertise rather than replacing it.

3. Retrieve information from healthcare records. When clinicians cannot scroll through a patient's history to find relevant information, a model can be used to query it. ChatEHR is an AI-based software currently being piloted at Stanford Medicine [19]. Such applications are certainly a net gain, provided they do not introduce tool dependency. However, AI summarisation (as ChatEHR also claims to do) differs from retrieval: it carries a risk of grounding clinical decisions in incorrect and hallucinated "information".

4. Free clinicians from "time-consuming" tasks. Some tasks are *assumed* by authors to be an obstacle to higher-level tasks [20]. This is also described as "reduce the workload" or "reducing the duration of the examination" [9]. For example, Matthew et al. reported an average time saving of 7.62 minutes per scan from adopting an AI-assisted approach for biometric measurements and plane detection [8]. However, some "time-consuming" tasks may be essential for clinicians to develop their skills.

5. Replace the trained clinician. Examples: in a study entitled "*No sonographer, no radiologist*", Arroyo et al. advocate for sonographer-less obstetric care in rural and under-resourced communities [21], while Aguado et al. argue for Doppler image analysis "even for non-trained readers" [9]. Similarly, Ramirez Zegarra et al. recommend automation because ultrasound acquisition "requires years of training and extensive knowledge of fetal anatomy" [22]. To support this argument, a context of "global shortage of imaging experts" is invoked [20].

6. "Machines are intrinsically superior to humans". This justification might seem surprising initially, but it is one of the most frequently suggested. Machines are said to have reproducibility, absolute consistency over time, whereas clinicians' "performance" does not possess those attributes [20]. Human intelligence is presented as "error-prone", subjective and altered by fatigue, whereas machines introduce reproducibility, absolute consistency over time and never get tired [20]. Overall detection rates of fetal malformations are described as "low" due to the "human factor" [22], even when cited studies report prenatal detection "already accounting for 50% or more of critical congenital heart defects detected in many programmes" and as "increasing", with some very high rates in certain areas (87% in France). Intraobserver variability in foetal biometry measurements is described as "high" [22] when the reference study reports 3.0% to 6.0% differences. This narrative is not specific to obstetric medicine, as outperformance claims over humans have become common, as in this title: "ChatGPT with GPT-4 outperforms emergency department physicians in diagnostic accuracy: retrospective analysis" [23]. As Drogot et al. rightly note: "these outperformance claims often lack specificity, contextualisation and empirical

grounding" [24]. Even so, the *idea* that machines surpass clinicians' intelligence and skill is repeated and used as a major implicit argument for automation. This systematic depreciation of human performance may deter students from engaging in clinical training, affect doctors' morale and lead to unrestrained automation.

Technical Risks

Let us now suppose that a properly grounded automation reaches the clinics. The machine can still produce errors that go unnoticed and are used in clinical decision-making. Below, we review the technical and systemic factors that increase the likelihood of such occurrences and explain how some of them can be addressed.

First, this risk is documented. In a study by Matthew et al., AI tools saved a satisfactory set of 13 ultrasound views in 73% of cases, whereas manual scanning achieved a 98% success rate [8]. Another study estimating gestational age reported mean errors ranging from 1.45 to 7.73 days when using poorly segmented images [25]. Such errors can have clear and serious clinical implications, such as missed post-term pregnancies.

In particular, the likelihood of these errors increases in ultrasound imaging due to certain specific factors: noisy images (speckle noise); weak contrast between tissues compared with magnetic resonance imaging (MRI) or computed tomographies; significant variations depending on the patient, foetal position, probe type, probe angle, applied pressure and coupling quality; non-isotropic pixel spacing (which fluctuates depending on direction) and partial fields of view; changing numbers of images and the appearance order of structures per session. Additionally, text and callipers increase the risk of shortcut learning, where machines rely on misleading cues and fail to generalise well beyond the training set [26].

Second, skewed validation increases the likelihood of trusting errors in practice. More often than not, models are evaluated with unrealistic test datasets that foster performance exaggeration and unawareness of model biases [27]. This distortion is aggravated by publication bias, where studies reporting "state of the art" performance are more likely to be published, regardless of reproducibility [1, 28].

A model is biased when its performance reliably declines on certain subgroups: ethnicity, Body Mass Index (BMI), image quality, machine (ultrasound machines are replaced more frequently than other devices [29]), etc. Without transparency into these biases, the user has no insight into which patients can be safely assessed by the model. However, to date, studies that evaluate model bias have been the exception [7, 30]; their absence has been common. Detailed bias analyses should become standard practice. Then, depending on the context, the user can make an informed decision about whether the risk of error is acceptable.

Other **solutions** to avoid models' errors in clinical workflows include automatic quality control methods that associate a metric of quality to the output of a model [25] or to its input [31].

Explainable models that provide explanations along with predictions can facilitate the detection of absurd outputs [31]. Even so, clinicians have been shown to trust erroneous results even when provided with explanations [32].

Cognitive Risks for the Clinician

Even a justified and error-free automation will be harmful if it impairs the critical judgment of its users, the clinicians. This section reviews existing proof of this phenomenon, describes its psychological cause and suggests solutions to avoid it.

Short term: automation bias

Automation bias is the user's tendency to follow an automatic system "decision" even when it is incorrect and in the presence of contradictory information. In a randomised clinical trial, clinicians favoured automated decision-making systems despite contradictory or clinically nonsensical information [33]. Automated support was given to 457 clinicians to diagnose clinical vignettes. The diagnostic accuracy modestly increased from 73% to 76% with support from a good model, but dropped from 73% to 62% with outputs from a biased model. Providing model explanations did not mitigate this harmful effect (73% to 64%). As Khera et al. rightly note, this concerning automation bias occurred in «controlled settings, without the usual pressure on time [34].» In another study, Dratsch highlighted that this loss of competence occurred regardless of level of clinical experience [32]. Their prospective experiment asked 27 radiologists to assess 50 mammograms with AI assistance. The machine suggestion was incorrect for 12 mammograms. The AI correctness significantly impacted the percentage of correct ratings for inexperienced (80% versus 20%), moderately experienced (81% versus 25%) and even very experienced (82% versus 45%) radiologists.

Long-term effect: de-skilling or impact on critical thinking and sense of responsibility

Erosion of clinical expertise

Until now, we have focused extensively on the short-term effects of automation on the clinician's faculties. However, it seems reasonable to also consider the long-term consequences of automation. Everything is dictated by the following principle: any knowledge or skill that is not practised is *lost*.

To predict the long-term impact on clinicians, examining how automation has influenced human capacities in other sectors can be helpful. For example, having digital devices remember information for us has led to digital amnesia [35] and a loss of spatial memory [36].

A notable warning comes from the experience with cockpit automation. During the eighties, the American Congress asked NASA to investigate how automation affected pilots. First, Earl Wiener, analysing crash reports, concluded that some major accidents were caused by automation [37]. Stephen Casner from the Cames Research Center then examined the issue of inattention and skill

retention. By observing pilot-computer interactions in simulators, he showed that pilots' ability to make complex cognitive decisions suffered a palpable reduction from automation; the more automation there was, the more pilots reported "mind wandering" or thinking about inconsequential topics [38].

More recently, some researchers have begun studying AI-induced de-skilling. In a paper entitled *Your Brain on ChatGPT*, Nataliya Kosmyna followed the neural activity of essay writers for four months and observed a significantly lower activity (and connectivity) in the brain of those "assisted" by ChatGPT [39]. Michael Gerlich sent a questionnaire to more than 600 persons and reported a -0.68 (!) correlation coefficient between AI usage and critical thinking score [40].

Clinicians possess no special immunity to this long-term, insidious de-skilling [41]. AI systems only offer an illusion of thinking themselves [42] and foster the illusion of understanding in users [43]. Hence, clinicians would maintain the impression of performing the tasks themselves: de-skilling may occur without their noticing. Clinicians' cognitive perspective is far richer than the models' correlational process [44]. This loss must be avoided.

Erosion of responsibility

Another clinician's virtue that requires practice is moral decision-making, as it presupposes responsibility. This key aspect is rarely discussed [20]. Here, a well-documented psychological principle should be considered: the diffusion of responsibility or "bystander effect" [45]. In the presence of another potential agent, individuals are less likely to assume responsibility and act autonomously, even when doing so would benefit the group as a whole [45]. The Milgram experiment highlighted that such blind compliance is aggravated when the other agent is perceived as an authority or an "expert" [46].

The cause is mainly psychological

Introducing a machine next to the clinician can result in: (1) Turning a solitary worker into a perceived group of two operators; (2) Offering an effortless way to reach an objective; (3) Creating an apparently safe, comfortable environment, where safety nets make errors and inattention inconsequential. But all these conditions have been associated with a performance drop in Psychology:

1. The Ringlemann effect (or social loafing). Performance and motivation have been shown to decline significantly when a *perceived* co-worker is added to a task, compared with working alone [47, 48].

2. The principle of least effort. This principle states that people naturally choose the path of least resistance or "effort" [49]. Humans perceive avoiding effort as gratifying [50] and reduce effort through "cognitive offloading". The principle of least effort results in cognitive miser: "People are limited in their capacity to process information, so they take shortcuts whenever they can" [51]. For example, the mere presence of a smartphone has been shown to reduce the available cognitive capacity [52]. "People look up information that they actually know or could easily learn,

but are unwilling to invest the cognitive cost associated with encoding and retrieval" [53].

3. The Yerkes Dodson law (or necessity of pressure). This law states that human performance declines in the absence of a stimulating and competitive environment [54, 55].

Solutions

What is unacceptable is the creation of an environment that encourages idleness, mind-wandering, false security or irresponsibility. It would undermine sound medical practice. A recent review on automation in foetal ultrasound described the "ideal" sonographer-machine collaboration as one in which the system would "work in real time", and where the sonographer would utilise its outputs and identify when it fails [56]. The clinician would be the "assessor" of machine output. This is precisely the setting that Stephen Casner identified as undesirable in pilots: "What we're doing is using human beings as safety nets or backups to computers, and that's completely backward; it would be much better if the computing system watched us and chimed in when we do something wrong" [38, 57]. Earl Wiener reached the exact same conclusion regarding cockpit automation [37]. And the above statements align with Pascal Blatzer reflecting upon automation bias in medical application [58]. We must respect the laws of human psychology when building automated systems.

More research is needed to identify and eliminate automation bias among clinicians. The ideal configuration in which the machine actually benefits the clinician's cognitive and moral engagement has yet to be identified. Such a result can only flow from interdisciplinary collaboration involving psychology, learning science, human-computer interaction, machine learning and medical scientific communities.

To this day, the Food and Drug Administration (FDA) leaves the door wide open to clinician deskilling, never mentioning deskilling or automation bias in its January 2025 guidance document [59]. Only design validation is mandatory; bias assessment and quality control, advised; and complete passivity of the clinician, permitted: "AI-enabled devices span a *continuum of decision-making roles* from more autonomous systems to supportive tools" [59].

Once identified by researchers, the principles that regulate healthy and harmful human-computer interaction must be translated into regulations, with the assistance of legal advisors, to ensure integrated devices will not cause deskilling.

Conclusion

This review identified risks in automating clinics, using obstetrics as an illustrative example. We hope that our work, offering consistent food for thought, will inspire meaningful reflections among clinicians and researchers. Among other undesirable outcomes, a situation in which de-skilled clinicians become completely dependent on machines owned by a few private companies must be avoided at all costs. In particular, our main contribution is to connect psychological

principles to automation bias and deskilling issues. We also highlighted the shortcomings of current regulations with regard to de-skilling.

However, a number of concerns still remain; we have not discussed the crucial influence of the industrial-academic ecosystem and associated publication bias on AI. We also have not discussed the patient-focused risks of depersonalisation, disruption of relationships between clinicians and patients, or the risk of patient deception when AI diagnostics are framed as "personalised" to overcome patients' resistance to AI.

Correspondence *Joris Fournel*. E-mail: jorfo@dtu.dk

Accepted 26 January 2026

Published 26 February 2026

Conflicts of interest JF reports financial support from or interest in the Novo Nordisk Foundation. AF and MT report financial support from or interest in the Novo Nordisk Foundation and Prenaital Aps. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. These are available together with the article at ugeskriftet.dk/DMJ.

References can be found with the article at ugeskriftet.dk/DMJ

Cite this as Dan Med J 2026;73(4):A10250852

doi 10.61409/A10250852

Open Access under Creative Commons License [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)

REFERENCES

1. Saidi P, Dasarathy G, Berisha V. Unraveling overoptimism and publication bias in ML-driven science. *Patterns*. 2025;6(4):101185. <https://doi.org/10.1016/j.patter.2025.101185>
2. Almanaa M. Trends and public perception of artificial intelligence in medical imaging: a social media analysis. *Cureus*. Published online September 2024. <https://doi.org/10.7759/cureus.70008>
3. Banja J. AI hype and radiology: a plea for realism and accuracy. *Radiol Artif Intell*. 2020;2(4):e190223. <https://doi.org/10.1148/ryai.2020190223>
4. OpenAI's Sam Altman sees AI bubble forming as industry spending surges. *CNBC*. Published online 2025. <https://www.cnbc.com/2025/08/18/openai-sam-altman-warns-ai-market-is-in-a-bubble.html>
5. Natali C, Marconi L, Dias Duran LD, Cabitza F. AI-induced deskilling in medicine: a mixed-method review and research agenda for healthcare and beyond. *Artif Intell Rev*. 2025;58(11). <https://doi.org/10.1007/s10462-025-11352-1>
6. Wong CK, Lin M, Raheli A, et al. An automatic guidance and quality assessment system for Doppler imaging of the umbilical artery. In: *Simplifying medical ultrasound*. Springer Nature Switzerland; 2023:13-22. https://doi.org/10.1007/978-3-031-44521-7_2
7. Taksoee-Vester CA, Mikolaj K, Bashir Z, et al. AI-supported fetal echocardiography with quality assessment. *Sci Rep*. 2024;14(1). <https://doi.org/10.1038/s41598-024-56476-6>

8. Matthew J, Skelton E, Day TG, et al. Exploring a new paradigm for the fetal anomaly ultrasound scan: artificial intelligence in real time. *Prenat Diagn.* 2021;42(1):49-59. <https://doi.org/10.1002/pd.6059>
9. Aguado AM, Jimenez-Perez G, Chowdhury D, et al. AI-enabled workflow for automated classification and analysis of feto-placental Doppler images. *Front Digit Health.* 2024;6. <https://doi.org/10.3389/fdgth.2024.1455767>
10. Enache IA, Iovoaica-R&;mescu C, Ciobanu &;G, et al. Artificial intelligence in obstetric anomaly scan: heart and brain. *Life.* 2024;14(2):166. <https://doi.org/10.3390/life14020166>
11. Stringer JSA, Pokaprakarn T, Prieto JC, et al. Diagnostic accuracy of an integrated AI tool to estimate gestational age from blind ultrasound sweeps. *JAMA.* 2024;332(8):649. <https://doi.org/10.1001/jama.2024.10770>
12. Dülger Ö, Dursun A, Osmano&;lu UÖ. Fetal birth weight estimation with machine learning techniques in 15–40 weeks of pregnancy. *Van Med J.* 2024;31(3):186-191. <https://doi.org/10.5505/vmj.2024.09797>
13. Pegios P, Sejer EPF, Lin M, et al. Leveraging shape and spatial information for spontaneous preterm birth prediction. In: *Simplifying Medical Ultrasound.* Springer Nature Switzerland; 2023:57-67. https://doi.org/10.1007/978-3-031-44521-7_6
14. Zamagni G, Monasta L, Lees C, Stampalija T. EP08.51: Machine learning methods to predict fetal growth restriction: a systematic review. *Ultrasound Obstet Gynecol.* 2024;64(S1):197-197. <https://doi.org/10.1002/uog.28379>
15. Gupta K, Balyan K, Lamba B, et al. Ultrasound placental image texture analysis using artificial intelligence to predict hypertension in pregnancy. *J Matern Fetal Neonatal Med.* 2021;35(25):5587-5594. <https://doi.org/10.1080/14767058.2021.1887847>
16. Spairani E, Steyde G, Subitoni L, et al. A semi-supervised deep learning approach to automate the identification of fetal behavioral states in fetal heart rate tracings. In: *2024 IEEE International Symposium on Medical Measurements and Applications (MeMeA).* IEEE; 2024:1-6. <https://doi.org/10.1109/MeMeA60663.2024.10596918>
17. Sadda P, Imamoglu M, Dombrowski M, et al. Deep-learned placental vessel segmentation for intraoperative video enhancement in fetoscopic surgery. *Int J Comput Assist Radiol Surg.* 2018;14(2):227-235. <https://doi.org/10.1007/s11548-018-1886-4>
18. Tan Y, Peng Y, Guo L, Liu D, Luo Y. Cost-effectiveness analysis of AI-based image quality control for perinatal ultrasound screening. *BMC Med Educ.* 2024;24(1). <https://doi.org/10.1186/s12909-024-06477-w>
19. Armitage H. Clinicians can chat with medical records through new AI software, ChatEHR. Published June 5, 2025. <https://med.stanford.edu/news/all-news/2025/06/chatehr.html>
20. Drukker L, Noble JA, Papageorghiou AT. Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. *Ultrasound Obstet Gynecol.* 2020;56(4):498-505. <https://doi.org/10.1002/uog.22122>
21. Arroyo J, Marini TJ, Saavedra AC, et al. No sonographer, no radiologist: new system for automatic prenatal detection of fetal biometry, fetal presentation, and placental location. *PLoS One.* 2022;17(2):e0262107. <https://doi.org/10.1371/journal.pone.0262107>
22. Ramirez Zegarra R, Ghi T. Use of artificial intelligence and deep learning in fetal ultrasound imaging. *Ultrasound Obstet Gynecol.* 2023;62(2):185-194. <https://doi.org/10.1002/uog.26130>
23. Hoppe JM, Auer MK, Strüven A, et al. ChatGPT with GPT-4 outperforms emergency department physicians in diagnostic accuracy: retrospective analysis. *J Med Internet Res.* 2024;26:e56110.

- <https://doi.org/10.2196/56110>
24. Drogts J, Milota M, van den Brink A, Jongsma K. Ethical guidance for reporting and evaluating claims of AI outperforming human doctors. *npj Digit Med*. 2024;7(1). <https://doi.org/10.1038/s41746-024-01255-w>
 25. Cengiz S, Almakk I, Yaqub M. FUSQA: fetal ultrasound segmentation quality assessment. In: International Conference on Medical Imaging with Deep Learning. 2023. <https://api.semanticscholar.org/CorpusID:257405469>
 26. Lin M, Weng N, Mikolaj K, et al. Shortcut learning in medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. Springer Nature Switzerland; 2024:623-633. https://doi.org/10.1007/978-3-031-72111-3_59
 27. Sarno L, Neola D, Carbone L, et al. Use of artificial intelligence in obstetrics: not quite ready for prime time. *AJOG MFM*. 2023;5(2):100792. <https://doi.org/10.1016/j.ajogmf.2022.100792>
 28. Serra-Garcia M, Gneezy U. Nonreplicable publications are cited more than replicable ones. *Sci Adv*. 2021;7(21). <https://doi.org/10.1126/sciadv.abd1705>
 29. Renewal of radiological equipment. *Insights Imaging*. 2014;5(5):543-546. <https://doi.org/10.1007/s13244-014-0345-1>
 30. Fournel J, Pegios P, Sejer EPF, et al. The cervix in context: bias assessment in preterm birth prediction. In: Fairness of AI in Medical Imaging. Springer Nature Switzerland; 2025:43-52. https://doi.org/10.1007/978-3-032-05870-6_5
 31. Lin M, Ambsdorf J, Sejer EPF, et al. Learning semantic image quality for fetal ultrasound from noisy ranking annotation. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). IEEE; 2024:1-5. <https://doi.org/10.1109/ISBI56570.2024.10635225>
 32. Dratsch T, Chen X, Rezazade Mehrizi M, et al. Automation bias in mammography: the impact of artificial intelligence BI-RADS suggestions on reader performance. *Radiology*. 2023;307(4). <https://doi.org/10.1148/radiol.222176>
 33. Jabbour S, Fouhey D, Shepard S, et al. Measuring the impact of AI in the diagnosis of hospitalized patients: a randomized clinical vignette survey study. *JAMA*. 2023;330(23):2275. <https://doi.org/10.1001/jama.2023.22295>
 34. Khera R, Simon MA, Ross JS. Automation bias and assistive AI: risk of harm from AI-driven clinical decision support. *JAMA*. 2023;330(23):2255. <https://doi.org/10.1001/jama.2023.22557>
 35. Kaspersky Lab. The rise and impact of digital amnesia: why we need to protect what we no longer remember. Kaspersky Lab; 2015. <https://media.kasperskycontenthub.com/wp-content/uploads/sites/100/2017/03/10084613/Digital-Amnesia-Report.pdf>
 36. Dahmani L, Bohbot VD. Habitual use of GPS negatively impacts spatial memory during self-guided navigation. *Sci Rep*. 2020;10(1). <https://doi.org/10.1038/s41598-020-62877-0>
 37. Wiener EL. Beyond the sterile cockpit. *Hum Factors*. 1985;27(1):75-90. <https://doi.org/10.1177/001872088502700107>
 38. Casner SM, Geven RW, Recker MP, Schooler JW. The retention of manual flying skills in the automated cockpit. *Hum Factors*. 2014;56(8):1506-1516. <https://doi.org/10.1177/0018720814535628>
 39. Kosmyna N, Hauptmann E, Yuan YT, et al. Your brain on ChatGPT: accumulation of cognitive debt when using an AI assistant for essay writing task. Published online 2025. <https://arxiv.org/abs/2506.08872>
 40. Gerlich M. AI tools in society: impacts on cognitive offloading and the future of critical thinking. *Societies*. 2025;15(1):6. <https://doi.org/10.3390/soc15010006>

41. Budzyński K, Romańczyk M, Kitala D, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. *Lancet Gastroenterol Hepatol*. 2025;10(10):896-903. [https://doi.org/10.1016/S2468-1253\(25\)00133-5](https://doi.org/10.1016/S2468-1253(25)00133-5)
42. Shojaee P, Mirzadeh I, Alizadeh K, et al. The illusion of thinking: understanding the strengths and limitations of reasoning models via the lens of problem complexity. In: *NeurIPS*. 2025. <https://ml-site.cdn-apple.com/papers/the-illusion-of-thinking.pdf>
43. Messeri L, Crockett MJ. Artificial intelligence and illusions of understanding in scientific research. *Nature*. 2024;627(8002):49-58. <https://doi.org/10.1038/s41586-024-07146-0>
44. Tikhomirov L, Semmler C, McCradden M, et al. Medical artificial intelligence for clinicians: the lost cognitive perspective. *Lancet Digit Health*. 2024;6(8):e589-e594. [https://doi.org/10.1016/S2589-7500\(24\)00095-5](https://doi.org/10.1016/S2589-7500(24)00095-5)
45. Darley JM, Latané B. Bystander intervention in emergencies: diffusion of responsibility. *J Pers Soc Psychol*. 1968;8(4 Pt 1):377-383. <https://doi.org/10.1037/h0025589>
46. Milgram S. Behavioral study of obedience. *J Abnorm Soc Psychol*. 1963;67(4):371-378. <https://doi.org/10.1037/h0040525>
47. Ingham AG, Levinger G, Graves J, Peckham V. The Ringelmann effect: studies of group size and group performance. *J Exp Soc Psychol*. 1974;10(4):371-384. [https://doi.org/10.1016/0022-1031\(74\)90033-X](https://doi.org/10.1016/0022-1031(74)90033-X)
48. Simms A, Nichols T. Social loafing: a review of the literature. *J Manage*. 2014;15
49. Zipf GK. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley; 1949:597
50. Gheza D, Kool W, Pourtois G. Need for cognition moderates the relief of avoiding cognitive effort. *PLoS One*. 2023;18(11):e0287954. <https://doi.org/10.1371/journal.pone.0287954>
51. Fiske ST, Taylor SE. *Social Cognition*. 2nd ed. McGraw Hill Higher Education; 1991
52. Ward AF, Duke K, Gneezy A, Bos MW. Brain drain: the mere presence of one's own smartphone reduces available cognitive capacity. *J Assoc Consum Res*. 2017;2(2):140-154. <https://doi.org/10.1086/691462>
53. Barr N, Pennycook G, Stolz JA, Fugelsang JA. The brain in your pocket: evidence that smartphones are used to supplant thinking. *Comput Hum Behav*. 2015;48:473-480. <https://doi.org/10.1016/j.chb.2015.02.029>
54. Sodhi K, Luthra M, Mehta D. Yerkes-Dodson law for flow: a study on the role of competition and difficulty in the achievement of flow. *Int J Educ Manag Stud*. 2016;6:95. <https://api.semanticscholar.org/CorpusID:151956526>
55. Deutscher C, Ötting M, Langrock R, et al. Very highly skilled individuals do not choke under pressure: evidence from professional darts. Published online 2018. <https://arxiv.org/abs/1809.07659>
56. Day TG, Matthew J, Budd S, et al. Sonographer interaction with artificial intelligence: collaboration or conflict? *Ultrasound Obstet Gynecol*. 2023;62(2):167-174. <https://doi.org/10.1002/uog.26238>
57. Konnikova M. Hazards of automation. Published online 2015. <https://www.newyorker.com/science/maria-konnikova/hazards-automation>
58. Baltzer PAT. Automation bias in breast AI. *Radiology*. 2023;307(4). <https://doi.org/10.1148/radiol.230770>
59. U.S. Food and Drug Administration. Artificial intelligence-enabled device software functions: lifecycle management and marketing submission recommendations—draft guidance for industry and FDA staff. U.S. Department of Health and Human Services; 2025. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/artificial-intelligence-enabled-device-software-functions-lifecycle-management-and-marketing>