AI and Ethics (2025) 5:5783–5793
https://doi.org/10.1007/s43681-025-00825-2

REVIEW

Automation complacency: risks of abdicating medical decision making

Michael I. Saadeh[1] · Joel Janhonen[1,2,3] · Emily Beer[1,4] · Camille Castelyn[1,5] · David N. Hoffman[1,4]

Received: 28 July 2025 / Accepted: 11 August 2025 / Published online: 28 August 2025
© The Author(s) 2025

Abstract
This work investigates automation complacency in relation to decision support systems used in healthcare contexts, especially their impact on clinicians, patients, and the quality of care. While AI and decision support systems can enhance efficiency and outcomes in healthcare, the potential for automation bias risks clinical perils. These include eroded vigilance, impoverished therapeutic relationships, and potentially poorer outcomes regarding overall well-being. This work highlights these concerns to urge actors in the health sector to effectively integrate technology in a way that spares cognitive resources without compromising the essential role of human experts in making medical decisions. To ensure decision support improves patient care, it is crucial to balance computational processing of information with embodied local expertise; we provide a possible starting point for mindful integration. The implementation of systems in the clinical context should encourage vigilance and guard against fatigue and complacency. There is reason to be excited about increasingly efficient and available care. If the risks of automation complacency are avoided, shared time and resources can be used to preserve and promote valuable interactions, insights, and holistic aspects of care.

Keywords Medical decision making · Clinical decision support systems · Automation bias · Clinician-patient relationship · Cognitive resources

1 Introduction

This work investigates automation complacency in healthcare. We explore the implications of such complacency in practice, examining how it affects clinicians, patients, and the overall quality of care. Decision support systems and AI applications that aid in detection, diagnostics, and treatment will likely continue to increase efficiency in healthcare.[1] However, they risk undermining the role of clinicians or at least transforming clinical interactions and relationships. As automation becomes increasingly prevalent in the clinical setting, there is a growing concern that clinicians may overly rely on automated systems, potentially leading to inappropriate outcomes or the impoverishment of care relationships. We cannot reasonably expect that the great promise of artificial intelligence in its application to healthcare delivery will be achieved without rigorous debate about its capabilities and limitations.

Healthcare decision making belongs to patients. The decision support systems at issue here are not making decisions for patients; they are used by clinicians tasked with "deciding" medical questions, including which diagnosis is accurate or what the best treatment would be. Clinicians facilitate patient decisions by employing evidence-based communication skills, interpreting diagnostic tests, recognizing and

✉ Joel Janhonen
joel.janhonen@bioetiikka.fi

Michael I. Saadeh
msaadeh98@gmail.com

Emily Beer
eab2276@columbia.edu

Camille Castelyn
camille.castelyn@up.ac.za

David N. Hoffman
dnh2101@columbia.edu

1 Innovative Bioethics Forum, New York, USA

2 The Finnish Institute of Bioethics, Tampere, Finland

3 University of Turku, Turku, Finland

4 Columbia University, New York, USA

5 University of Pretoria, Pretoria, South Africa

[1] This paper includes imaging technologies (including those that the FDA regulates as "software as a medical device") in its definition of clinical decision support systems.

mitigating cognitive biases, and adopting a patient-centered approach [1]. But now, emerging technology in the form of decision support systems is taking on some aspects of this facilitation and decision making. Meta-analysis of over 100 controlled trials shows that computerized decision support, including AI-assisted diagnosis, prediction, disease detection, and treatment, improves both targeted processes of care and patient outcomes [2]. However, such tools should not gain a decisive role in the decision making process but rather provide support when their implementation is justified. This should occur in conjunction with the approval and discussion between the patient and the medical staff.

Remaining vigilant when designing, applying, and engaging with decision support systems is vital for putting a check on automation bias and consequent complacency. Automation bias may risk abrogation of trust as a result of diagnostic errors, insufficient patient treatment plans, deference of clinician attention, and overly generalized or even biased datasets. This may occur as clinicians seek to spare cognitive effort, whether deliberate or not, thus producing an increased bias toward automation tools and complacency in clinicians. If this technology takes the front seat in the patient-clinician relationship, worse patient outcomes, contextually oblivious actions, and impoverished care may ensue [3].

1.1 Context

Recent reviews summarize recommendations to improve equity [4, 5], mitigate algorithmic bias [6, 7], and integrate AI ethically [8, 9]. While the importance of keeping clinicians in the loop is widely recognized, nuances can be found in the task-specific needs for tailored oversight and proposals to make human-AI workflows more continuous or collaborative with adjustable confidence thresholds and hand-off points—rather than passively acting on recommendations. Beyond performance metrics, strategies and measures to ensure transparency, accountability, and cultivation of active patient and community involvement are being broadly considered.

There is already an international consensus on the ethical desirability of maintaining vigilant human oversight and appropriate control throughout the lifecycle of AI-enabled clinical systems [10]. Our contribution about personal, relational, and contextual concerns aligns with this, while adding depth and enriching the discussion with issues such as holistic appreciation, relations, trust, and dignity. This paper concerns the use of AI systems marketed to clinicians as decision support systems—or medical devices, in the case of imaging—that carry risk for automation complacency. Table 1 presents examples of such systems with related advantages and risks.

There are many other applications for AI in healthcare that also leverage AI's primary strength—the processing of large swaths of data that could not be accomplished manually by a human. These include things like drug discovery, pandemic modeling, and research that requires analysis of enormous data sets. Because our focus is on point-of-care, uses of AI that do not involve decision making in the clinical setting are outside the scope of this paper.

1.2 Decision support systems in the delivery of healthcare

Each particular clinical circumstance and its relevant AI call for noting that "artificial intelligence" is not one unitary phenomenon. As a starting point, discussions of AI in healthcare must recognize important differences between broad categories of AI models: for instance, traditional AI is explainable, while machine learning tools, and especially deep learning tools, continuously adapt, learn from errors, and refine themselves at the expense of explainability. The main tradeoff here is between adaptability and transparency. Clinicians may need to differentiate between being able to interpret the recommended action, explain why the tool has chosen this course of action, or describe how the AI tool came to the decision it did [11]. These various levels may add to the understanding of the level of decision support the tool is offering and to what extent cognitive bias may be at play. Dynamic machine learning tools are less predictable and more opaque; hence, they come with additional ethical considerations.

In 2019, the Nuffield Foundation published an ethics roadmap for researching outcomes of technologies that perform algorithm-driven decision making [12]. The document identifies four sets of insightful trade-offs that are likely to also describe tensions that will be encountered in the clinical context. The benefits are increased diagnostic accuracy,

Table 1 Decision support and automation systems

	Imaging interpretation	Assessment of clinical lab findings	Indications for further care/treatment
Domain	Radiology, ultrasound interpretation	Blood chemistry, enzyme levels, electrocardiogram	Speech therapy, physical therapy, discharge planning
Proposed advantages	Early disease detection, improved treatment planning, less invasive	Assess probability of sudden death (in a highly repeatable manner)	Improved clinical efficiency, care coordination, personalized communication
Complacency risk	Generalization / overfitting dataset	Insufficient attention to diagnostic accuracy	Incorrect assumptions, incomplete case management, sociocultural risk

personalization of medicine, increased efficacy (ensuring a minimum level of quality and access for all), and convenience. The risks include inequality due to embedded biases or lack of representation, threats to solidarity and social participation, challenges to privacy, and autonomy over personal information. This roadmap considered greater convenience, or the reduction of difficulty and effort, to be in possible tension with the promotion of dignity and self-actualization, if things and interactions central to human flourishing would become hollow and devoid of meaning.

The clinical merits and ethical propriety of decision support systems are based on their providing support rather than supplanting decision making. When the systems act as decision substitution and the clinical expertise is sidelined, decision substitution is masquerading as (and being marketed as) decision support. Where the replacement of decisions made by individual humans (decision substitution) is misrepresented as AI decision support, we lose the ability to define and assign responsibility for the consequences of the patient care decisions they produce. Vendors that promote AI tools for healthcare may describe their tools as decision support, whereas practitioners may use them for decision substitution due to the inherent productivity pressures of the healthcare environment. Clinical users are likely tempted to adopt the recommendations of the algorithm with little or no human assessment. It is all well and good for a vendor to say that its product can be used only as a decision support tool, but if the algorithm that produces the "recommendation" is so complex and opaque that no human actor could, within a reasonable clinical time frame, assess all factors that the AI tool used to reach its recommendation in order to genuinely adopt the algorithm's recommendation as valid, the actual result is decision substitution. This circumstance is made more problematic with AI tools that base their outputs on a dynamic model and evolving dataset, as compared to those that rely upon a static and validated clinical body of source data.

1.3 Automation bias

Automation bias and consequent complacency are present in the healthcare industry and impact patient outcomes. The use of decision support systems in clinical healthcare is becoming normalized in a growing number of care tasks [13]. This makes related biases an increasingly relevant topic in healthcare delivery. Automation bias is the tendency to defer to decision support systems, replacing one's own investigative means and information-collection process, as well as one's use of judgment steeped in medical expertise. Automation complacency, implying action or lack thereof, arises when an operator or interactor of an automated system incorrectly assumes it is running smoothly and fails to account for future problems that may arise from this sort of negligence or "low index of suspicion" [14]. The complacency trap is often exacerbated by mechanical functions of routine jobs, like time pressure and workload. Another factor that may lead to automation complacency is the design of the systems involved in decision making. Therefore, it is possible to mitigate automation complacency by increasing user accountability, improving worksite conditions, or decreasing cognitive demand without attempting to somehow cancel our innate automation bias [15].

Automation complacency is problematic because it can lead to over-reliance on technological decision support. Reduced vigilance by human operators may compromise patient safety, quality of care, and clinical interactions; awareness of the underlying bias is therefore critical. If clinicians and patients were active in questioning automated systems' recommendations, they could provide a check on this bias, reducing automation complacency. While advisable, this may be insufficient, as sustainably overcoming inherent cognitive biases remains challenging, especially if the system design and clinical circumstances incentivize it.

Over-reliance and complacency, while unique, arise similarly as they relate to human attention and mental effort. In general, people try to refrain from cognitive effort. [16] If a method or mental shortcut is available that makes decision making more efficient and gives the right answer most of the time, people have generally found it worth adopting, at least in evolutionary terms [17]. This is understandable, given the immense pressures that clinicians are under, due to fatigue, understaffing, limited resources, and competing agendas. However, technological options for sparing effort can undermine human vigilance. Then again, a prolonged cognitive overload would manage the same by producing anxiety and fatigue that makes one adopt so-called avoidance behaviors [18]. The benefits of decision automation in sparing cognitive resources are explored briefly later. There appears to be a need to strike a neutral balance between complacency and avoidance.

Automation should not be allowed to sideline human intuition and the unique competencies of clinicians. In time, people come to embody the practical know-how of their profession [19]. Once integrated, procedures become automatic and do not need continual deliberation. But importantly, decisions that need critical deliberation should not be automated, mentally or technologically. The attentional integration model proposes that the reallocation of attentional resources away from decision making results in decreased situational awareness and learned complacency [14]. Hence, when clinicians become complacent and direct their attention elsewhere, there may be insufficient monitoring of automation output. As complacency is partly "learned carelessness," experts are not better shielded from it than

amateurs. However, since personal and situational factors are known to modulate the outcome, there might be means to mitigate overreliance on automated systems.

1.4 Automation bias in diagnosis and clinical care

Automation bias has been demonstrated in the clinical context. Sutton and colleagues reviewed the literature on benefits and risks related to clinical decision support systems [20]. These systems could come to undermine the quality of patient care. Clinicians will need continued vigilance as technology evolves and experiences of successful and failed implementations accumulate. One important consideration highlighted in their review is the level of fatigue among care providers. It is not guaranteed that automation decreases fatigue, especially if inundated with disruptive alerts or verification tasks that are experienced as a form of unnecessary 'rubber stamping'. Developers should consider the experience and bandwidth of clinicians rather than exacerbate alert fatigue, avoidance, and complacency by primarily worrying about things like potential liability [21].

A 2023 randomized crossover study found that clinicians of all expertise levels were vulnerable to automation bias, even though AI improved their overall diagnostic accuracy and efficiency [22]. While assisted by AI, nearly half of the errors were associated with the misleading effect of this bias. However, all the clinicians in the interview held an optimistic attitude toward their AI-assisted performances, denying having been misled. The authors proposed a strategy of AI suppression that they reported to be effective as a practical method for decreasing errors by clinicians. This again emphasizes the need to consider the clinician-AI interaction rather than merely evaluating the performance of a given algorithm.

Another study concerning the interpretation of electrocardiograms found that while automated diagnosis positively influences decision making, a flawed system significantly reduces the reader's diagnostic accuracy [23]. This indicates a bias that manages to lower the interpreter's confidence and interpretation accuracy when an automated diagnosis is incorrect. Instead of suppression, authors recommended using multiple predictions corresponding to a kind of differential diagnosis that would increase the chance of providing the correct answer, as well as function to keep complacency in check. Due to the inherent reliance on automated systems, such a differential could attenuate the disparate levels of confidence and uncertainty.

Researchers from 25 years ago and more have proposed using automation just for monitoring rather than for decision support, when effort from the human individual should not be substituted or undermined [17].

2 Clinical perils of relying on decision support systems

Exploring prospective perils at the individual, interpersonal, and systemic levels will follow, but we should start with what is evident in light of history. The dangers of automation bias when using decision support systems are especially grave if the tools make inappropriate or poor decisions. The use of AI-powered decision support systems in healthcare is in an early, transitional phase. But technological transitional phases are not unprecedented. The development of X-rays as a clinical tool is a useful analog. The discovery that X-rays could produce a radiographic image of the interior of the human body was received with much anticipation. Understandably, the prospect of avoiding what was known as "exploratory surgery" was an enormous temptation for the practitioners of the day. X-ray has demonstrated its enormous value in the ensuing decades, but missteps occurred in the early years with catastrophic consequences.

2.1 Historical analog

One example is the use of radiation to treat an "enlarged" thymus gland. When clinicians began to investigate the underlying causes of sudden infant death syndrome (SIDS), they inferred a correlation between the death of infants in the first several months of life and the size of the infant's thymus gland, which was thought to obstruct the infant's airway. This conclusion was grounded in the clinicians' understanding of the normal size of a thymus gland. Clinical assumptions about normal thymus gland size were based on observation of the thymus glands of the deceased newborns that were available for anatomical dissection. What was not known then, but was later determined, was that the infant cadavers used in anatomical studies did not have normal-sized thymus glands but unusually small ones. The cadavers were sourced from people experiencing poverty; chronic stress from disease and malnutrition shrank their thymus glands. This mistaken inference about what was clinically normal and what was abnormal, produced from well-intentioned but faulty anatomical studies, led to the conclusion that the enlarged thymus was the source of airway obstruction in SIDS babies and the mistaken belief that irradiation to shrink the thymus could redress the cause of asphyxia. [24].

Radiation of the thymus gland dramatically elevated the risk of thyroid and breast cancer in the thousands of patients who received this treatment before it was eventually identified as harmful and contraindicated [25]. The clinicians who initiated the practice of thymus radiation did not know that their conclusions were based on invalid underlying assumptions.

This example is relevant for two different reasons. First, it demonstrates the damage caused by hastily and excitedly adopting new technology and overusing it before proving its safety. X-rays were enthusiastically adopted as both a method of diagnosis and treatment before clinicians understood the terrible risks of thymus gland radiation. Second, the thymus gland example is akin to relying on a dataset that is not generalizable or overfitting (creating a complex model too closely based on the training data). Today's circumstances bear some resemblance, as we contemplate the appropriate level of scrutiny for data underlying decision making and need to be vigilant over our evolving relationship with medical technology. Likewise, we must evaluate whether a decision support or a decision substitution algorithm is making clinically valid determinations. Without oversight over implementation and operation, systems risk being inaccurate, useless, and inappropriate. The duty of care dictates what information we must consider in assessing the reliability of an artificial intelligence system, whether static, dynamic, or a model continuously learning from accumulating data.

2.2 Recent examples that raise concern

The dangers of automation bias and consequent complacency when using decision support systems are especially grave if these systems are making clinically inappropriate decisions. As with X-ray pelvimetry and thymus irradiation, the danger of using AI tools that have not been adequately tested before clinical implementation is manifested in two recent circumstances. The first is the use of AI for reading mammograms. Some anticipate that AI will completely replace humans in the reading of mammograms due to their greater accuracy, speed, and reduced cost. [26]. However, a recent study of a commercially available FDA-approved algorithm for breast cancer detection found considerable differences in error rates depending on a patient's race and age. For example, a black patient was 45% more likely to receive a false positive than a white patient. The FDA imposes no formal requirement for diversity in validation datasets, and many commercial developers withhold dataset details under trade-secret and intellectual-property protections [27]. This raises significant concerns about whether currently available algorithms can be ethically adopted at scale. As the authors noted, if this algorithm's outputs were adopted, black patients would have a significantly higher recall rate than white patients, which could "worsen healthcare disparities and decrease the benefits of AI assistance."

The second circumstance where AI algorithms have been found to be problematic—due to the inaccurate or incomplete nature of the underlying data relied upon to "charge the system"—is the use of AI in hospital case management and discharge planning. This problem has been the subject of wide-ranging discussion at professional conferences where AI tools are purporting to give thorough discharge planning "recommendations," leaving the clinician unable to see what data did and did not contribute to the algorithmic determination.

For example, if a discharge planning algorithm concludes that a patient's inpatient treatment is no longer clinically necessary and that the patient is safe to discharge home, that decision might be entirely valid when considering only the data available to the algorithm in the patient's medical record and other sources. But the algorithm may not have access to all relevant information, which can lead to poor outputs. For example, an algorithm may determine that a patient who cannot walk on level surfaces can be discharged to receive home physical therapy, even though the patient lives in a walk-up apartment building with stairs at the front entrance and more stairs internally to reach the patient's bedroom. [28]. This very circumstance has produced a lawsuit against United Healthcare for using an AI algorithm to generate decisions that consistently select the least expensive discharge plan. [29, 30].

Mammography and discharge planning demonstrate the need for clinicians to provide oversight. For mammography, access to the data and a transparent model would provide a way to check accuracy. And in the discharge context, oversight would ensure a holistic appreciation of a patient's needs, including data not available to the algorithm. Oversight in each case would incorporate human judgment in the decision to rely on the algorithm's recommendation. Clinicians may also need a new skill set that enables them to evaluate the decision support systems they use.

2.3 Personalized and holistic care requires a person who cares

As a working assumption, general clinical care should be organized to optimize patient outcomes with precise and effective treatments by meeting individual medical needs. The use of decision support systems, like most aspects of clinical care, should maximize the clinical value from resources. A positive clinical outcome and accurate diagnosis are also primary goals for any clinician. However, it is equally important to acknowledge the various factors contributing to patient well-being. Patient preferences, relevant and effective communication, and the general emotional empathy of clinicians can all contribute to successful clinical care [31].

An automated system's accurate diagnosis is only one part of a patient's well-being. Thus, the implementation of a human-in-the-loop at all decision stages of clinical care is critical to combating automation complacency. A

human-in-the-loop approach can ensure personal sensitivities are accounted for and patient preferences are best met. Clinicians' expertise steeped in experience may be nearly irreplaceable for appropriately including idiosyncratic human factors in medical decision making. Successful compilation of a preference profile is still worlds apart from meaningfully conversing with the patient or having a conversation to inform care. For example, a patient who is diagnosed according to an automated system's diagnostic recommendations may disagree with the treatment plan despite its accuracy or statistical merits. Perhaps the patient's quality of life would be reduced if a rigorous treatment plan were advised to aggressively treat a disease. While the automated system may "argue" or compute that a certain treatment plan is necessary due to survival calculus, the patient may disagree if their personal calculus differs, which may be a combination of morals, pain preferences, additional disorders, and external factors. More importantly, only a person can properly assess their own needs as they are experiencing pain or discomfort in the face of sickness. The question of patient safety is most central to the debate about an automated system's introduction into healthcare [32]. So, while the importance of an accurate diagnostic recommendation is a primary goal of successful patient outcomes, the patients also need to be the primary voice in their treatment plan.

2.4 Undermining the clinician-patient relationship

Positive patient-clinician relationships are associated with better health outcomes [33]. For example, a study on patients with multiple chronic conditions found that participants who reported better quality patient-provider relationships also "perceived less treatment and self-management burden, reported less distress, felt greater confidence in their ability to self-manage" [34]. Another study found that the primary factor influencing whether a cancer patient decides to report medication management concerns and safety events is their perceived relationship with their clinicians. Patients reported that perceived good relationships were the result of mutual regard and trust [35]. In psychotherapy, research demonstrates that, after accounting for patient characteristics, the therapeutic relationship between patient and provider can account for "as much, and probably more, of the outcome variance as particular treatment methods" [36].

The potential impact of decision support systems on the patient-clinician relationship must be evaluated. If clinicians rely improperly on decision support systems, this could have a negative impact on the relationship. For instance, a clinician who adopts a suggested diagnosis without critical analysis may be unable to satisfactorily explain to the patient why that diagnosis is appropriate. If a clinician signs off on an inappropriate patient discharge by recklessly adhering to a recommendation of a discharge planning algorithm (like in the example above, where a mobility-impaired patient is deemed safe for discharge to an unsuitable home environment), this may undercut the patient's trust and respect for their clinicians. Similarly, a clinician who adopts a suggested course of treatment without considering alternatives may be unprepared to discuss treatment options with the patient and may not be able to tailor the treatment to the patient's concerns. All these examples could be avoided if clinicians appropriately use these systems for decision support rather than decision making. Learning the extent to which these problems may occur requires evaluation in the clinical environment.

Many patients are uncomfortable with the use of AI in their healthcare and do not believe it will lead to better health outcomes [37]. To the extent that the use of AI systems is disclosed to a patient, as it often should be, the patient may respond according to their perception of AI. A patient's distrust of AI could result in less trust in a clinician, damaging the patient-clinician relationship and potentially lowering the patient's confidence in the quality of their care. Patient- or clinician-facing AI applications that have conversational ability or other human-like qualities pose unique risks [38]. But both algorithm-driven decision making and digitalized therapeutic interactions remove the human touch from healthcare. The objective of efficient and personalized medicine should not cause care to become impersonal.

As others have noted, patient safety and benefit cannot be adequately assured by evaluating AI performance outside of the clinical setting; we must also evaluate the clinician-AI interaction [22]. But we must go further than simply looking at things like diagnosis error rates. A clinician's use or misuse of automated systems may have indirect impacts as well, including on the patient-clinician relationship.

2.5 Navigating sociocultural landscapes on autopilot

Clinical decisions are not only isolated events but repeating features that occur in continuums of care. As complex decision trees branch out from each fork in the road, practical management of healthcare can appear daunting. To limit variability and optimize outcomes, interventions are sequenced into "clinical pathways" that signify the normal course of clinical decisions. These conceptual tools help to streamline and manage the quality of treatment and care processes specific to a given diagnosis, procedure, or symptom [39].

TheWHO and the OECD importantly emphasize that such standardized coordination of care involves more than algorithmic decision making. These recommendations need

to become embedded into the institutional environment, its local structures, and culture. The steps taken in the course of treatment do not occur in a vacuum but within a complex healthcare context [40].Clinical pathways should be considered dynamic; continually reshaped by the needs of the target population, evolving medical knowledge, as well as the available competencies and resources. Pathways to guide care can only be crafted by first knowing the terrain by heart.

Regardless, attempts to deploy decision support systems to pathway modeling are underway [41, 42]. Prior work has raised ethical concerns on discrimination, transparency, patient autonomy, and moral responsibility [43]. The concern about automation bias—that clinicians uncritically follow clinical pathways with minimal cognitive effort—also applies [44]. Increased emphasis on the algorithm should not come at the expense of contextual factors, de-humanizing or de-localizing the processes of care. Intangible human considerations are easily left out of AI models, as these are challenging to describe mathematically, which could decrease their emphasis if AI is relied on to calculate these pathways. In contrast to a task like analyzing an image, developing clinical pathways requires real-world understanding.

Understanding sociocultural factors is an important part of healthcare delivery as a whole. Human clinicians are well placed for this task, but future AI applications may attempt to perform a social calculus. One approach to utilizing culture-sanctioned social metrics is described by Khan and colleagues [45]. Figuring out which medical solutions would be considered inappropriate based on socioeconomic or cultural factors is highly risky as well and may exacerbate inequities. Yet for some applications, the capacity to factor in social considerations is necessary for arriving at appropriate decisions. Potential dynamic interactions with patients from diverse backgrounds would benefit from being influenced by relevant social signals [46], especially for patient-facing medical decision support applications.[2]

[2] There has already been at least one instance where an AI-powered support tool for clinicians was transformed into a patient/consumer-facing application [51]. Patient-facing tools are beyond the scope of this paper. But we must acknowledge that these tools can evolve quickly. To the extent that we accept a decision support tool because it is properly used by clinicians, such a tool may easily find a new user or context. This may create new or magnified ethical concerns, along with new potential benefits. As we argue throughout, it is not simply the accuracy of a tool or algorithm that must be evaluated, but the implications and context of its real-world use.

3 Beyond clinical perils

3.1 Optimizing cognitive resources

The tendency to spare effort has come about for a good reason. Doctors and other clinicians should arguably save their cognitive resources for the most important aspects of their jobs, especially for tasks that escape automation. One such category of tasks is empathetic caregiving, which is an essential part of caring professions and expends some of the available attention and effort. The critically important emotional labor often depletes cognitive resources, especially if there is a constant need to project positive emotions to patients [47]. Practitioners may benefit from relying on automation when possible. This reliance can be considered to protect against extraneous effort, wasting time, and cognitive attention, allowing them to focus on patient communication.

3.2 Guaranteeing minimum threshold of care

Advanced and abundant medical decision support would most likely increase the overall amount of available healthcare. Besides adding to the quantity of provided services, a general upside of automation is the uniformity in the quality of output. For example, a production line assembling cars is unencumbered by bad Mondays, prejudiced attitudes, or scattered attention, resulting in neglect that could later prove catastrophic. When a specific task is automated, the outcomes become more predictable, and explainable errors need only happen once. In a practical sense, relying on an Excel worksheet that has been in use for years, not to make unprecedented errors, is the appropriate mindset. Trusting this automated accountant with minimal suspicion makes much more sense than trusting a human.

Regardless of how advanced automation gets, patients should not be considered and treated as malfunctioning biological machines—repaired and maintained onboard a 'clinical assembly line.' While understanding the mechanisms and functions of the body is critical, clinicians are more than mere mechanics of the body. It is worth reflecting on how clinical pathways and quality differ and should differ from production lines. Nevertheless, when correctly applied, automation in the healthcare context could perceivably guarantee a minimum but uniform level of consideration for all. Therefore, if we care about improving productivity and accessibility, arguably any competencies that can be automated should be, not to substitute human expertise with AI, but to complement skill sets and address shortages. It could be widely beneficial to have AI-based support available in mundane everyday medical situations that do not require or even merit a human professional,

certainly not a second opinion. Assuming such applications would not be designed to increase engagement by evoking strong emotions and so end up needlessly driving distressed people to their clinicians.

3.3 Personalized support for preventative medical decisions

Healthcare expenditures per person in the US are greater than those in other industrialized countries, despite less optimal health outcomes based on metrics like life expectancy. Increasing expenditure does not seem to translate into better outcomes [48]. This has led some to argue that AI offers an opportunity to shift from a defensive battle against disease to instead taking the initiative, enabling the sector to focus more on prevention and supporting overall health [49]. The clinician-patient relationship need not compete with an accompanying user-facing health coach that is well informed about their situation and always available for advice. Such personalized guidance and motivation could be the ultimate 'decision support system', not primarily to help diagnose or address an acute problem but to make the right personal choices to avoid them in the first place. Yet, private health systems could use wellness and prevention technology to substitute rather than complement existing services to maximize profits.

4 Implications for practice

In this paper, we have shown how and why automation bias may result in complacency during clinicians' use of decision support systems. Automation bias is present in diagnostics and clinical care, as it is in other fields and consumer use. Because people tend to trust technology, or they lack some combination of time, attention, and other resources to be appropriately suspicious of it, these circumstances carry implications for clinical practice. Clinicians' inherent tendency to defer to software purposed to support decision calls for designing even better and more reliable AI. It also calls for educating, training and raising clinicians' awareness toward understanding the implications of the tools they may use now and in the future.

A useful resource, for example, is "Top 10 must-knows for clinicians using AI models" by Abgrall and colleagues [11]. A guide for rethinking clinical reasoning, by Corrao and Argano, emphasises awareness of the decision making process and the ability to switch between automated non-critical thinking and the more costly and slow deliberative mode. [1] While complete debiasing may be impossible, systems should be designed to activate this shift to more effortful engagement, rather than letting users slip into unreflective "autopilot mode." Additionally, during the design process, a differentiation should be made between the sharp or blunt end of healthcare making. The sharp end of healthcare decision making refers to operative decisions, whereas the blunt end has to do with management and administrative tasks. [50].

Some tasks warrant merely automated monitoring, whereas others best benefit from decision support that facilitates dynamic workflow or differential diagnosis based on multiple lower-level predictions—leaving the final interpretation to the clinician [17, 23]. For mindful integration in clinical practice, a gradient of task-appropriate AI use cases should be compiled; we contribute a possible starting point for this deliberation in Fig. 1.

Human experts must retain control over the implementation and use of the decision support systems to ensure that decisions ultimately have human insight and accountability. If clinicians serve as the final check on the validity of medical recommendations, yet are overly influenced by AI and perhaps not properly informed of its limits and deficits, then invalid recommendations may ensue. AI integration ought to be sensitive to human needs and the subtleties of good care.

Many patients rely on the clinicians' specialized medical knowledge and would not want their diagnosis or care passed to AI. Human-tasks—like breaking bad news—are no place to spare attention or effort. Special emphasis should be placed on patient dignity and self-actualization. The AI-related value trade-offs listed in the Nuffield roadmap should include 'connection' as one of the tensions, to encourage attention on the implications for human interactions and relationships [12, 38]. Because practicing medicine is not just a science but an art, dealing with the very processes of life, guidance about it needs to be based on experience and subsequently attained understanding of these processes. Something that lacks these experiences can only ever be a tool for navigation, not a guide to rely on.

Optimism about the benefits of medical automation should not divert attention from concerns about its de-humanizing or de-localizing impact and challenge to responsibility and dignity. While some may prefer automated decision making and note that AI is often right and the problem of human error appears insoluble, others prefer to rely on the clinician for context, empathy, expertise, and certainty that all options were explored.

5 Conclusion

Automation bias may surface as clinicians' interaction increases with decision support systems in a clinical setting. This can be exacerbated by a variety of challenges that divert

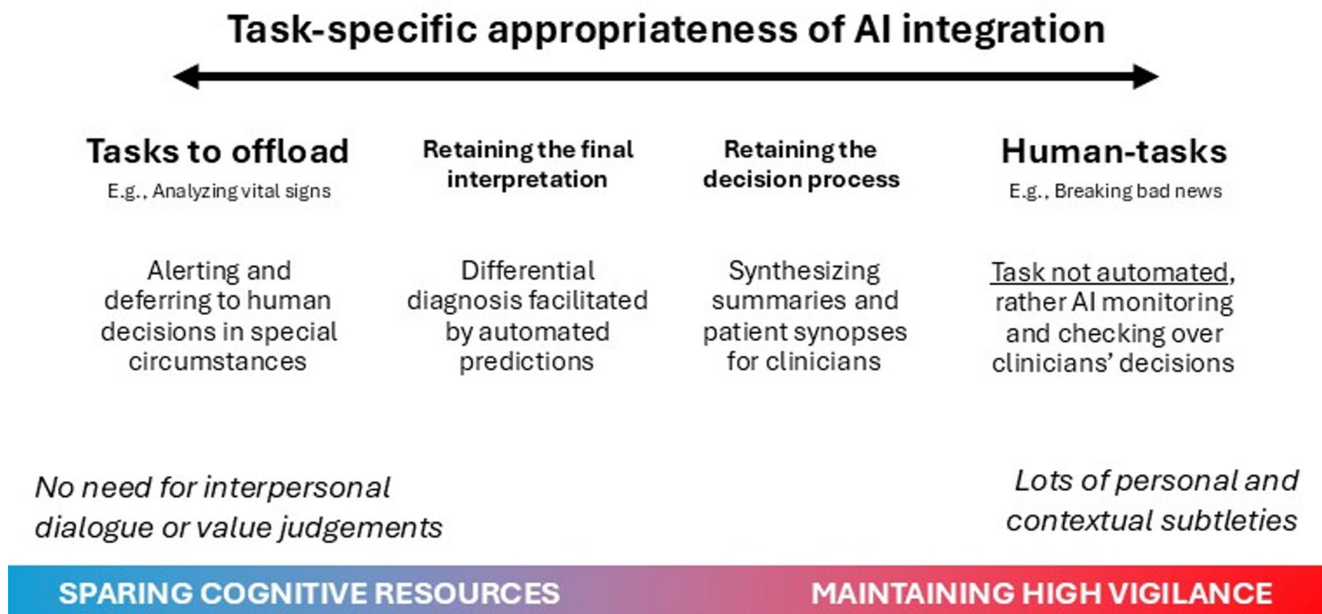

Fig. 1 Ethical gradient of clinical AI

resources or attention in such a way that clinicians further rely on these tools and automation complacency occurs. We call for greater attention to the challenge of distinguishing between automatable functions—where interpersonal dialogue, value judgements, and contextual subtleties play no role—and tasks that do now and always require a human in the loop. Our proposed rethinking of these interactions may shift clinicians' use of automation systems from a superficial and shallow "checking of the boxes" to a recommendation for using the saved time and resources to deepen and enrich therapeutic relationships.

Acknowledgements The authors are grateful to the Innovative Bioethics Forum (https://innovativebioethicsforum.com/) for providing a collaborative platform to explore and develop the ideas presented in this work. We extend special thanks to Anne Zimmerman for her continued support throughout.

Author contributions M.I.S., J.J., and E.B. drafted the initial version. C.C. and D.H. revised the text and contributed to the work throughout. All authors read and approved the final manuscript.

Funding Open Access funding provided by University of Turku (including Turku University Central Hospital). No funding was received to assist with the preparation of this manuscript.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval and consent to participate Not applicable.

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/.

References

1. Corrao, S., Argano, C.: Rethinking clinical decision-making to improve clinical reasoning. Front. Med. (2022). https://doi.org/10.3389/fmed.2022.900543
2. Kwan, J., Lo, L., Ferguson, J., Goldberg, H., Diaz-Martinez, J., Tomlinson, G., Grimshaw, J., Shojania, K.: Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. BMJ (Clin. Res. Ed.) (2020). https://doi.org/10.1136/bmj.m3216
3. Gala, D., Behl, H., Shah, M., Makaryus, A.: The role of artificial intelligence in improving patient outcomes and future of healthcare delivery in cardiology: a narrative review of the literature. Healthcare **12**(4), 481 (2024)
4. Berdahl, C., Baker, L., Mann, S., Osoba, O., Girosi, F.: Strategies to improve the impact of artificial intelligence on health equity: scoping review. JMIR AI (2023). https://doi.org/10.2196/42936
5. Liu, M., Ning, Y., Teixayavong, S., Liu, X., Mertens, M., Shang, Y., Li, X., Miao, D., Liao, J., Xu, J., Ting, D., Cheng, J., Ong,. Teo, T., Tan, N., RaviChandran, F., Wang, L., Celi, M., Ong, Liu, N.: A scoping review and evidence gap analysis of clinical AI fairness. Npj Digit. Med., **8**(1), 481
6. Sasseville, M., Ouellet, S., Rhéaume, C., Sahlia, M., Couture, V., Després, P., Paquette, J.-S., Darmon, D., Bergeron, F., Gagnon,

M.-P.: Bias mitigation in primary health care artificial intelligence models: scoping review. J. Med. Internet Res. (2025). https://doi.org/10.2196/60269
7. Hasanzadeh, F., Josephson, C., Waters, G., Adedinsewo, D., Azizi, Z., White, J.: Bias recognition and mitigation strategies in artificial intelligence healthcare applications. npj Digit. Med. (2025). https://doi.org/10.1038/s41746-025-01503-7
8. Mennella, C., Maniscalco, U., De Pietro, G., Esposito, M.: Ethical and regulatory challenges of AI technologies in healthcare: a narrative review. Heliyon (2024). https://doi.org/10.1016/j.heliyon.2024.e26297
9. Elhaddad, M., Hamam, S.: AI-driven clinical decision support systems: an ongoing pursuit of potential. Cureus (2024). https://doi.org/10.7759/cureus.57728
10. Health Ethics & Governance: Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization (2021)
11. Abgrall, G., Holder, A., Dagdia, Z., Zeitouni, K., Monnet, X.: 2024 Should AI models be explainable to clinicians? Crit. Care 28(1), 301
12. Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., Cave, S.: Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. Nuffield Foundation, London, UK (2019)
13. Patient Safety Network. Clinical Decision Support Systems: [Online]. Available: https://psnet.ahrq.gov/primer/clinical-decision-support-systems. [Accessed 27 September 2024] (2019)
14. Parasuraman, R., Manzey, D.: Complacency and bias in human use of automation: an attentional integration. Hum. Factors **52**(3), 381–410 (2010)
15. Goddard, K., Roudsari, A., Wyatt, J.: Automation bias: a systematic review of frequency, effect mediators, and mitigators. J. Am. Med. Inform. Assoc. **19**(1), 121–127 (2012)
16. Rasmussen, J.: Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE Trans. Syst. Man Cybern. **13**(3), 257–266 (1983)
17. Skitka, L., Mosier, K., Burdick, M.: Does automation bias decision-making? Int. J. Hum.-Comput. Stud. **51**(5), 991–1006 (1999)
18. B. da Silva Cezar and A. Gastaud Maçada, 2023 "Cognitive Overload, Anxiety, Cognitive Fatigue, Avoidance Behavior and Data Literacy in Big Data environments J *Information Processing & Management,* 60(6):103482,.
19. Gordon, D.: Clinical science and clinical experience: changing boundaries between art and science in medicine. In: Examined, B. (ed.) Dordrecht, pp. 257–298. Kluwer Academic Publishers, The Netherlands (1988)
20. Sutton, R., Pincock, D., Baumgart, D., Sadowski, D., Fedorak, R., Kroeker, K.: An overview of clinical decision support systems: benefits, risks, and strategies for success. npj Digit. Med. **3**, 17 (2020)
21. Patient Safety Network, "Alert Fatigue," 2019. [Online]. Available: https://psnet.ahrq.gov/primer/alert-fatigue. [Accessed 27 September 2024].
22. Wang, D.-Y., Ding, J., Sun, A.-L., Liu, S.-G., Jiang, D., Li, N., Yu, J.-K.: Artificial intelligence suppression as a strategy to mitigate artificial intelligence automation bias. J. Am. Med. Inform. Assoc. **30**(10), 1684–1692 (2023)
23. Bond, R., Novotny, T., Andrsova, I., Koc, L., Sisakova, M., Finlay, D., Guldenring, D., McLaughlin, J., Peace, A., McGilligan, V., Leslie, S., Wang, H., Malik, M.: Automation bias in medicine: the influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms. J. Electrocardiol. **51**(6), S6–S11 (2018)
24. Ritterman, J.: To err is human: can American medicine learn from past mistakes? Perm. J. **21**(3), 16–181 (2017)
25. Hildreth, N., Shore, R., Dvoretsky, P.: The risk of breast cancer after irradiation of the thymus in infancy. N. Engl. J. Med. **321**(19), 1281–1284 (1989)
26. Zeng, A., Houssami, N., Noguchi, N., Nickel, B., Marinovich, M.L.: Frequency and characteristics of errors by artificial intelligence (AI) in reading screening mammography: a systematic review. Breast Cancer Res. Treat. **207**, 1–13 (2024)
27. Nguyen, D., Ren, Y., Jones, T., Thomas, S., Lo, J., Grimm, L.: Patient characteristics impact performance of AI algorithm in interpreting negative screening digital breast tomosynthesis studies. Radiology (2024). https://doi.org/10.1148/radiol.232286
28. Jaffe, S.: Feds rein in use of predictive software that limits care for medicare advantage patients. 5 Oct 2023. [Online]. Available https://kffhealthnews.org/news/article/biden-administration-software-algorithms-medicare-advantage/
29. Estate of Gene B. Lokken, The et al. v. UnitedHealth Group, Inc. et al. (2023)
30. Ross, C., Herman, B.: UnitedHealth faces class action lawsuit over algorithmic care denials in Medicare Advantage plans. 14 Nov 2023. [Online]. Available https://www.statnews.com/2023/11/14/unitedhealth-class-action-lawsuit-algorithm-medicare-advantage/
31. Comer, R., Comer, J.: Abnormal Psychology. Worth Publishers, New York (2020)
32. Academy of Medical Royal Colleges. Academy of Medical Royal Colleges. Artificial intelligence in healthcare. Academy of Medical Royal Colleges. Academy of Medical Royal Colleges, London, UK (2019).
33. Kamimura, A., Higham, R., Rathi, N., Panahi, S., Lee, E., Ashby, J.: Patient–provider relationships among vulnerable patients: the association with health literacy, continuity of care, and self-rated health. J. Patient Exp. **7**(6), 1450–1457 (2020)
34. Eton, D., Ridgeway, J., Linzer, M., Boehm, D., Rogers, E., Yost, K., Rutten, L.F., St Sauver, J., Poplau, S., Anderson, R.: Healthcare provider relational quality is associated with better self-management and less treatment burden in people with multiple chronic conditions. Patient Prefer. Adherence **11**, 1635–1646 (2017)
35. Bunni, D., Walters, G., Hwang, M., Gahn, K., Mason, H., Manojlovich, M., Gong, Y., Jiang, Y.: Oncology patients' willingness to report their medication safety concerns from home: a qualitative study. Support. Care Cancer **32**, 352 (2024)
36. Norcross, J., Lambert, M.: Psychotherapy relationships that work III. Psychotherapy **55**(4), 303–315 (2018)
37. Tyson, A., Pasquini, G., Spencer, A., Funk, C.: 60% of Americans would be uncomfortable with provider relying on AI in their own health care. 22 Feb 2023. [Online]. Available https://www.pewresearch.org/science/2023/02/22/60-of-americans-would-be-uncomfortable-with-provider-relying-on-ai-in-their-own-health-care/
38. Zimmerman, A., Janhonen, J., Beer, E.: Human/AI relationships: challenges, downsides, and impacts on human/human relationships. AI Ethics **4**(4), 1555–1567 (2024)
39. Roeder, N., Hensen, P., Hindle, D., Loskamp, N., Lakomek, H.: Clinical pathways: effective and efficient inpatient treatment. Chirurg **74**(12), 1149–1155 (2003)
40. Rotter, T., Baatenburg de Jong, R., Evans Lacko, S., Ronellenfitsch, U., Kinsman, L.: Clinical pathways as a quality strategy. In: Improving Healthcare Quality in Europe: Characteristics, Effectiveness and Implementation of Different Strategies, Copenhagen, Denmark, European Observatory on Health Systems and Policies, pp. 309–329 (2019)
41. Qu, G., Liu, Z., Cui, S., Tang, J.: Study on self-adaptive clinical pathway decision support system based on case-based reasoning. Lect. Notes Electr. Eng. **269**, 969–978 (2014)

42. Wenk, J., Voigt, I., Inojosa, H., Schlieter, H., Ziemssen, T.: Building digital patient pathways for the management and treatment of multiple sclerosis. Front. Immunol. **15**, 3389 (2024)
43. Hill, E.R., Mitchell, C., Brigden, T., Hall, A.: Ethical and legal considerations influencing human involvement in the implementation of artificial intelligence in a clinical pathway: a multistakeholder perspective. Front. Digit. Health **5**, 1139210 (2023)
44. Grissinger, M.: Understanding Human Over-Reliance On Technology. Pharmacy and Therapeutics **44**(6), 320–375 (2019)
45. Khan, S., Streater, J., Bhatia, T., Fiore, S., Bölöni, L.: Learning social calculus with genetic programing. In: Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference (2013).
46. Janhonen, J.: Socialisation approach to AI value acquisition: enabling flexible ethical navigation with built-in receptiveness to social influence. AI Ethics **5**(1), 527–553 (2023)
47. Lee, L., Madera, J.: A systematic literature review of emotional labor research from the hospitality and tourism literature. Int. J. Contemp. Hosp. Manag. **31**(7), 2808–2826 (2019)
48. Peter, G.: Peterson Foundation. Why are Americans paying more for healthcare? 3 Jan 2024. [Online]. Available https://www.pgpf.org/blog/2024/01/why-are-americans-paying-more-for-healthcare
49. Alonso, V.: Scott Galloway—unpacking the AI burst and current market conditions. 6 Oct 2023. [Online]. Available https://www.nbforum.com/newsroom/events/nordic-business-forum-2023/scott-galloway-unpacking-the-ai-burst-and-current-market-conditions/
50. Nemeth, C., Nunnally, M., O'Connor, M., Klock, P., Cook, R.: Getting to the point: developing IT for the sharp end of healthcare. J. Biomed. Inform. **38**(1), 18–25 (2005)
51. Butcher, M.: Ada Health built an AI-driven startup by moving slowly and not breaking things. [Online]. Available https://techcrunch.com/2020/03/05/move-slow-and-dont-break-things-how-to-build-an-ai-driven-startup/. [Accessed 5 September 2024] (2020)

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