Artificial Intelligence Review (2025) 58:356
https://doi.org/10.1007/s10462-025-11352-1

AI-induced Deskilling in Medicine: A Mixed-Method Review and Research Agenda for Healthcare and Beyond

Chiara Natali[1,2] · Luca Marconi[1] · Leslye Denisse Dias Duran[3] · Federico Cabitza[1,4]

Accepted: 4 August 2025

Abstract

The integration of Artificial Intelligence (AI) in healthcare is reshaping clinical practice, offering both opportunities for enhanced decision-making and risks of skill degradation among medical professionals. This growing impact calls for a comprehensive evaluation of its effects on medical expertise. This study presents a mixed-method literature review, combining systematic analysis with narrative synthesis to examine AI-induced deskilling and upskilling inhibition-the erosion of medical expertise and the reduction of opportunities for skill acquisition due to AI-driven decision support systems. Anchoring the discussion in the core medical competencies outlined by the *Federation of Royal Colleges of Physicians of the UK-Practical Assessment of Clinical Examination Skills* (PACES-MRCPUK), the systematic review identifies key vulnerabilities in physical examination, differential diagnosis, clinical judgment, and physician-patient communication. The narrative review explores broader themes related to Human–AI Interaction and the Impact of AI on Human Skills in Organizations. In response to concerns about the *Second Singularity*-a scenario in which decision-making autonomy is increasingly ceded to AI, weakening human oversight-this review advocates for a research agenda that prioritizes longitudinal studies, real-time monitoring of AI's impact, and the development of frameworks to mitigate skill erosion, ensuring the preservation of professional autonomy and the safeguarding of the irreplaceable elements of human judgment in medicine and beyond.

Keywords Deskilling · Artificial intelligence · Healthcare · Human–AI interaction · Narrative review · Systematic review

1 Introduction

The rapid integration of Artificial Intelligence (AI) in healthcare decision-making processes offers significant opportunities to enhance diagnostic accuracy, improve efficiency, and potentially transform patient outcomes. Yet, as AI systems become more deeply embed-

Chiara Natali and Luca Marconi contributed equally to this work.

Extended author information available on the last page of the article

ded in daily practice, growing concerns have emerged about their long-term impact on medical expertise-particularly the gradual erosion of physicians' critical skills and clinical judgement.

Gary Klein's notion of the second singularity offers a powerful lens through which to understand this risk: a turning point where repeated delegation to intelligent systems leads not to liberation, but to atrophy-an irreversible loss of professional know-how as skills go unused and unrefined over time (Klein 2022). Beyond individual cognitive effects, entire professions and organizations risk being undermined by system fragility (Sparrow and Hatherley 2019): when automation gradually erodes human expertise, critical processes become increasingly vulnerable as reliance on AI becomes entrenched and the necessary skills to operate without it are forgotten.

In healthcare, the risk of system embrittlement is particularly pronounced due to the critical importance of continuously practising and updating clinical skills-the competencies essential for physicians to ensure consistently high standards of patient care (Michels et al. 2012).

This paper focuses on two interrelated yet conceptually distinct threats: deskilling, the degradation of previously acquired competencies due to reduced practice, and upskilling inhibition, the suppression of opportunities to develop new or advanced skills-especially among trainees-due to over-reliance on AI systems.

Deskilling, also known as *skill loss*, *skill erosion* (Rinta-Kahila et al. 2023), *skill decay* or *skill fade* (de Andres Crespo et al. 2025), refers to the degradation, or loss of previously acquired professional skills, typically resulting from reduced practice or diminished necessity due to automation or technological substitution. This aligns with the classical notion of deskilling introduced by Braverman (1974), which describes the fragmentation of work into increasingly standardized tasks requiring less human expertise. In clinical settings, this trend manifests as a shift from hands-on, experience-driven decision-making to an oversight role in which physicians validate AI-generated recommendations rather than independently diagnosing and treating patients. The result is a progressive disengagement from complex cognitive and procedural tasks, reducing both technical proficiency and the nuanced clinical judgment that underpins effective medical practice. Beyond the loss of practical skills, this transformation also risks diminishing professional confidence and autonomy (Bainbridge 1983), reinforcing dependence on AI even in situations where human expertise remains essential.

Rather than the loss of established expertise, upskilling inhibition denotes the hindrance or suppression of opportunities for acquiring new skills, often because individuals are not adequately exposed to tasks that promote learning, challenge, or progressive responsibility. Medical training relies on a gradual increase in complexity and autonomy, where novices-students, residents, and early-career specialists-develop expertise by engaging with progressively difficult cases under supervision (Beane 2019). Over-reliance on AI may disrupt this learning trajectory by limiting exposure to challenging decision-making scenarios. When AI systems consistently provide solutions, trainees may miss critical opportunities to develop diagnostic acumen, problem-solving skills, and confidence in independent judgment. In the long run, this inhibition may result in a generation of clinicians who are less prepared to operate without AI assistance, exacerbating concerns about the sustainability of human expertise in medicine.

Both deskilling and upskilling inhibition phenomena present significant risks to clinical effectiveness, patient safety, and physician autonomy, as they impact critical competencies such as clinical judgement, physical examination, differential diagnosis, and effective patient communication. However, it is important to acknowledge that the relationship between AI and human skills in medicine is context-dependent and complex. AI implementation does not uniformly result in skill erosion; rather, it can stimulate a transformation in medical competencies (Aslam and Hoyle 2022). Physicians may find their roles evolving, focusing increasingly on overseeing AI-driven processes, validating algorithmic outputs, and integrating AI recommendations within broader patient management strategies. Moreover, AI systems themselves can function effectively as training tools, actively facilitating skill acquisition through adaptive tutoring systems or context-sensitive decision aids designed explicitly to maintain and even enhance clinical judgement and diagnostic capabilities.

Despite the significance of these dynamics, there remains a lack of systematic evidence on how, when, and for whom AI-induced deskilling occurs in healthcare. Its gradual and often imperceptible nature makes it difficult to measure, and its most affected domains-such as tacit knowledge, heuristics, and interpersonal skills-are inherently challenging to quantify (Anichini et al. 2024). Existing perspectives on AI's impact on medical skills remain divided. Pessimists argue that AI adoption could diminish clinical reasoning and patient interaction, reinforcing the "use it or lose it" principle of skill retention (Kleim and Jones 2008; Hoff, 2011). Optimists counter that AI serves as an augmentative tool, much like stethoscopes or imaging technologies, enhancing rather than replacing physician capabilities (Obermeyer and Lee 2017; Verghese et al. 2018). However, conclusive evidence remains limited, with most existing accounts remaining either speculative or fragmented across disciplines. The field lacks a cohesive vocabulary, a shared set of mechanisms, and a structured research agenda to investigate this phenomenon with the rigour it demands.

This paper addresses this gap by systematically examining the phenomenon of AI-induced deskilling and upskilling inhibition in medical literature, identifying their mechanisms, manifestations, and implications for clinical expertise, medical training, and healthcare system resilience. Based on this analysis, we propose a research agenda to guide future empirical inquiry and policy design.

To do so, we adopt a mixed-method approach combining a systematic literature review-anchored in the PACES-MRCPUK framework of clinical competencies and aligned with PRISMA guidelines-with a narrative synthesis that integrates broader insights from human–AI interaction, organisational behaviour, and ethical design.

The structure of the paper is as follows: in Sect. 2, we detail our methodology, distinguishing between the systematic and narrative review components. Section 3 presents findings from the systematic review, structured by clinical skill domains, followed by thematic results from the narrative review, centred on the topics of human–AI interaction and organisational integration. Section 4 synthesises these insights to propose a structured research agenda that highlights key areas for empirical investigation and practical intervention aimed at mitigating AI-induced deskilling risks. Section 5 outlines the study's limitations. Finally, Sect. 6 summarises our conclusions, reaffirming the need to safeguard and cultivate professional autonomy and the critical skills integral to healthcare and other high-stakes professions in the age of AI.

2 Methods

The methodological approach presented in this study integrates systematic analysis using the PRISMA reporting format (Page et al. 2021) with a narrative synthesis that captures the broader conceptual and theoretical dimensions of this issue. This two-pronged methodology categorizes existing research into seven clinical skill domains, based on the Practical Assessment of Clinical Examination Skills examination of the *Federation of Royal Colleges of Physicians of the UK* (PACES MRCP(UK)), designed to test the knowledge and skills of trainee doctors in core clinical skills (Elder et al. 2011). This examination is a comprehensive approach with a marked practical component that aims to ensure that physicians are prepared for real-world clinical competencies, including history taking, physical examination, communication, and decision making. As the aim of the systematic review is to identify the concerns regarding the phenomenon of deskilling in the studies included, the PACES methodology serves as an empirical foundation to analyze these concerns and make a classification according to established and broadly accepted clinical standards regarding clinical skills.

The narrative review examined broader implications for human–AI interaction and organisational skill dynamics, offering a comprehensive view of how AI is reshaping clinical skills in healthcare. Our structured research agenda advocates conceptual clarification, rigorous empirical investigation, and wide-reaching, actionable interventions to mitigate unintended consequences in medicine and other high-stakes, high-expertise professions. In the following sections, we detail our review methodology, present key findings, and discuss their implications for the future of medical training and practice, informing strategies that balance technological progress with the vital preservation of human judgment, autonomy, and professional resilience.

As stated in the introduction, although the phenomenon of deskilling is not a novel concept, its potential implications for the health workforce, health organizations, and health institutions remain unclear. Consequently, our goal is to contribute to the ongoing debate about the unintended consequences of AI implementation in clinical settings, and thus, the present hybrid systematic-narrative methodology is designed to identify, analyze, and categorize the key themes and concerns that are articulated in the current scientific literature. To the best of our knowledge, no hybrid study has been conducted on this particular topic.

2.1 Systematic review methodology

This systematic review is structured according to the PRISMA reporting guidelines (Page et al. 2021) to ensure a structured and evidence-based search of the literature. Searches were carried out in three databases: PubMed, Scopus and Web of Science. The search strategy was structured into three groups of terms, which were organized into a query string using the appropriate Boolean operators. First, we used the terms "deskilling", "de-skilling", "loss of skills", "skills degradation", "skills erosion", "skills atrophy", "skills displacement" and "skills decline". Second, we added the following terms to narrow the search for the medical sector "medicine" OR "healthcare" OR "health care" OR "clinical". Finally, we used the terms "artificial intelligence", "machine learning", "deep learning", and "decision support" to focus on articles discussing AI applications. Table 1 represents the identified inclusion and exclusion criteria for our systematic review.

Table 1 Inclusion and exclusion criteria for the systematic review

Criteria	Description
Inclusion (IC1)	Studies completely written in English
Inclusion (IC2)	Studies with a central focus on healthcare or medicine and artificial intelligence
Inclusion (IC3)	Studies directly and significantly dealing with the theme of deskilling
Exclusion (EC1)	Papers merely mentioning the notion of deskilling without elaboration
Exclusion (EC2)	Papers representing commentaries without specific relevance
Exclusion (EC3)	Papers outside the specific domain of AI in healthcare or medicine

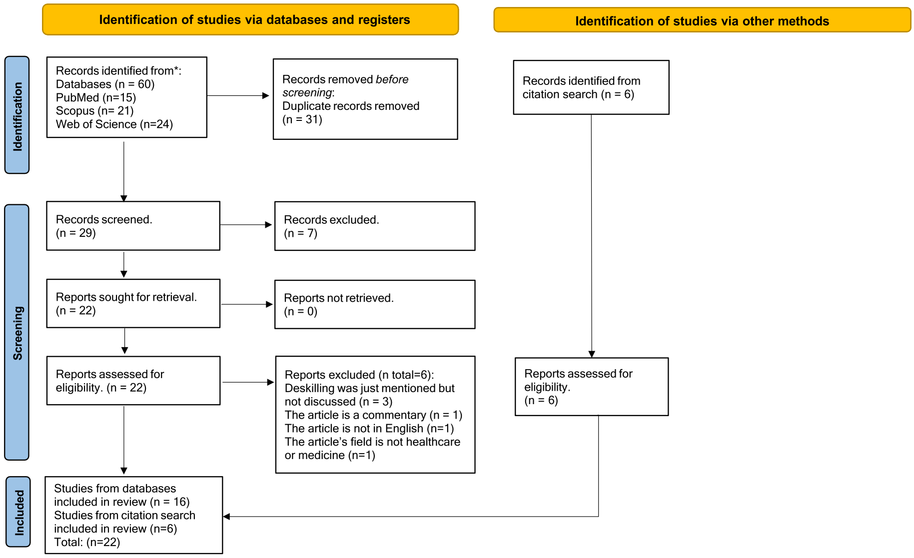

Fig. 1 PRISMA diagram

The articles and sources resulting from the query search were extracted in RIS format and then collected in Zotero for efficient management. Of the 60 sources initially identified, 31 duplicates were removed. The remaining database for abstract screening consisted of 29 articles. After the screening of the title and abstract, seven articles that did not meet the inclusion criteria were removed. The final eligibility assessment was conducted on the 22 sources and, after a full-text review, 6 articles that did not meet the inclusion criteria were excluded. From the remaining 16 sources, a cross-citation search was conducted to find articles that were not found in the database search but were considered potentially relevant to the systematic review. Six new sources were identified. They were screened and assessed for eligibility using the same criteria and it was confirmed that all were published in peer-reviewed journals. All of the six sources were included in the study (Fig. 1).

2.2 Narrative review methodology

The narrative review, complementing the systematic review, provides an in-depth exploration of AI-induced deskilling and upskilling inhibition in medicine. It seeks to understand this phenomenon from the perspectives of diverse professional stakeholders involved in AI development and deployment in healthcare. Through this comprehensive approach, the review analyzes the relationship between AI and medical professionals' expertise, addressing both its challenges and potential mitigative strategies. It emphasizes the importance of fostering a culture of reskilling and adaptation, highlighting the crucial balance between technological integration and human skill preservation in the medical field.

The narrative review component of this study intends to capture a broader conceptual, organizational, and ethical perspectives on AI-induced deskilling and the inhibition of upskilling. To ensure comprehensive thematic coverage, a broad exploratory search strategy was employed, focusing on the intersection of artificial intelligence, clinical skills, and healthcare transformation. Searches were conducted using Google Scholar, which was considered appropriate for identifying a diverse range of sources beyond the scope of traditional clinical and biomedical databases-such as theoretical contributions, interdisciplinary discussions, and commentaries relevant to human–AI collaboration, professional expertise, and the transformation of human skills in clinical contexts.

Keyword combinations included terms such as "AI and deskilling", "AI and clinical skills", "human–AI interaction in healthcare", "AI and clinical decision-making", "AI and human expertise", "AI in medical training", and "AI in medical education and training". The search process was iterative and snowball-based, incorporating both forward and backward citation tracking from relevant studies identified in the systematic review and through additional exploratory readings.

A total of 62 articles were included in the narrative synthesis. These were selected based on their conceptual relevance, diversity of perspectives (including empirical studies, conceptual papers, reviews, and opinion pieces), and their coverage of both macro- and micro-level issues related to human–AI interaction and organizational integration. Inclusion criteria required that articles explicitly address the impact of AI on human expertise, professional identity, or clinical decision-making in healthcare or comparable high-stakes domains. Studies that did not substantively engage with themes such as the transformation of clinical roles, training, or decision-making were excluded.

Each selected article was read in full and systematically analyzed using qualitative thematic synthesis. Key themes were extracted through an inductive process and subsequently mapped onto a categorization framework developed by the authors, consisting of two primary dimensions and seven analytical categories. Several articles from the systematic review were intentionally retained in the narrative synthesis to provide cross-dimensional insights, thereby ensuring conceptual continuity and enabling a more nuanced contextual interpretation. This layered approach facilitated the articulation of emergent patterns, associated risks, and potential mitigation strategies related to AI-driven transformations in clinical skills and expertise.

2.2.1 Scope and categorization

The categorization framework, which underpins our analysis, provides a comprehensive view of the subject and is structured around two primary dimensions—*Human–AI Interaction* and *AI Adoption and Integration*—encompassing a total of seven categories. The following list presents the complete categorization framework, detailing these two dimensions along with their corresponding categories:

1. Human–AI interaction

   - Impact of AI on Human Skills in Organizations
   - Deskilling Types and AI Influence on Clinical Decision Making
   - Physician Expectations and Deskilling Concerns
   - Patient-Centered Considerations, Perspectives and Safety

2. AI Adoption and Integration

   - Incorporation of Domain Knowledge in AI Models
   - Training and Education
   - Ethical and Legal Implications

The categories cover various aspects, from clinicians' interaction with AI in medical workflows to AI design, application, and integration within organizational medical processes. They also extend to the ethical, social, and legal implications of AI adoption and human–AI collaboration in the evolving landscape of medical skills and processes.

Overall, this framework provides a detailed analysis of the relationship between AI and healthcare professionals' skill dynamics, considering implications both organizational and ethical perspectives.

3 Results

3.1 Systematic review results

The included articles spanned academic research using qualitative analysis from semi-structured interviews: Aquino et al. (2023), Hallowell et al. (2023), Akudjedu et al. (2023), Chen et al. (2021), Stogiannos et al. (2025) and Parchmann et al. (2024), i.e. studies dealing with individuals' attitudes, beliefs, and opinions about the implementation and impact of AI in medical settings; a mini-review (Koplin et al. 2025); a rapid review (Ruskin et al. 2020); an essay (Rafner et al. 2022); a mixed methods review (Smith and Baumann 2020); review articles using conceptual analysis (Nakagawa et al. 2023; Dias Duran 2021; Duran et al. 2023; Kayaduvar and Ünal 2023; Monteith et al. 2022; Panesar et al. 2020; Sparrow and Hatherley 2019; Choudhury and Chaudhry 2024; Lu 2016; Levy et al. 2019) and opinion pieces with critical appraisal methods (Cabitza et al. 2017a; Kashou et al. 2024).

Some articles focused on a specific branch of medicine as follows: radiology (Akudjedu et al. 2023; Chen et al. 2021; Stogiannos et al. 2025), neurosurgery (Panesar et al. 2020), anesthesiology (Ruskin et al. 2020; Duran et al. 2023), oncology (Sparrow and Hatherley 2019), cardiology (Kashou et al. 2024), pathology (Nakagawa et al. 2023), fertility medicine (Koplin et al. 2025), geriatrics (Parchmann et al. 2024), psychiatry (Monteith et al. 2022), ophthalmology (Levy et al. 2019) and diagnosis of rare diseases (Hallowell et al. 2023).

Nine papers (Dias Duran 2021; Kayaduvar and Ünal 2023; Levy et al. 2019; Lu 2016; Nakagawa et al. 2023; Rafner et al. 2022; Smith and Baumann 2020; Aquino et al. 2023) established a definition of deskilling that featured a main element: deskilling as the reduction or decrease of existing skills due to the introduction of technology, in particular, artificial intelligence. Some papers (Cabitza et al. 2017; Lu 2016; Ruskin et al. 2020; Sparrow and Hatherley 2019; Aquino et al. 2023; Levy et al. 2019) also noted some consequences and causes of the deskilling phenomenon, and in particular, most of the papers identified clinician over-reliance on AI technology as problematic. While some articles explicitly indicating over-reliance as the cause of the deskilling phenomenon (Cabitza et al. 2017), others actually considered over-reliance as a potential consequence of deskilling itself (Sparrow and Hatherley 2019).

Through systematic analysis, 17 concerns regarding the phenomenon of deskilling were identified as presented in Table 2. These themes were classified into categories following the classification of clinical skills assessment presented by Elder et al. (2011) and Elder (2018) and based on the PACES examination. Clinical skills are generally understood to be the set of competencies that medical students and practicing physicians must develop, refine, and continually update in order to provide a standard of care to patients. According to a study conducted by Michels et al. (2012), “The term [clinical skills] is perceived to include physical examination skills, practical procedures, communication skills, and treatment/therapeutic skills, with or without integration across these domains.”

The seven categories of skills according to PACES are physical examination, identifying physical signs, clinical communication, differential diagnosis, clinical judgment, managing patient concerns, and maintaining patient well-being. The descriptors of these skills proposed by Elder et al. (2011) were initially used as a framework to analyze the concerns identified in the systematic review. Through the analysis, it was determined that for the purposes of classifying the concerns identified in the studies, the skills “identifying physical signs" and “physical examination", as well as the skills “managing patient concerns" and “clinical communication" could be grouped into two single categories as they encompassed similar aspects found throughout the analysis of the studies included in the review. Furthermore, analysis of these concerns revealed that of the 17 concerns identified in the articles, 12 could be grouped into the PACES classification, while the remaining 5 did not fit into the established categories. Therefore, two additional categories were formulated to group these concerns: operational and AI-specific. The operational category included concerns about the impact of AI on teamwork in clinical settings or at a systemic level in healthcare institutions. The AI-specific category encompassed the concerns related to the use of AI, such as medical errors as a result of automation bias or human–model interaction, dependence on AI tools that might lead to carelessness and the lack of ability of clinicians to understand and challenge the AI’s outputs. The full list of concerns mapped onto the PACES classification and the two additional categories are presented in Tables 2.

Table 2 Concerns identified in the literature on AI-induced deskilling in medicine, linked with specific clinical skills identified by Elder et al. (2011), as well as Organizational and AI-specific skills

Clinical skills	Concerns	References
Physical examination and identifying physical signs	Deterioration of physical examination skills	Hallowell et al. (2023), Levy et al. (2019), Lu (2016), Monteith et al. (2022), Rafner et al. (2022), Ruskin et al. (2020)
	Deterioration or loss of manual skills	Duran et al. (2023), Hallowell et al. (2023), Rafner et al. (2022), Ruskin et al. (2020), Smith and Baumann (2020)
	Reduction in skill development	Aquino et al. (2023), Chen et al. (2021), Monteith et al. (2022)
Clinical communication and management of patients' concerns	Worsening of the communication between physicians and patients	Aquino et al. (2023), Levy et al. (2019), Lu (2016), Parchmann et al. (2024)
	Deterioration of the relationship between physicians and patients	Akudjedu et al. (2023), Dias Duran (2021), Levy et al. (2019), Lu (2016), Rafner et al. (2022)
Differential diagnosis	Decrease in diagnostic accuracy	Cabitza et al. (2017), Kashou et al. (2024), Rafner et al. (2022), Smith and Baumann (2020)
	Clinical knowledge and interpretation	Choudhury and Chaudhry (2024), Hallowell et al. (2023), Koplin et al. (2025), Levy et al. (2019), Nakagawa et al. (2023), Rafner et al. (2022), Smith and Baumann (2020), Sparrow and Hatherley (2019)
Clinical judgment	Poorer clinical reasoning and judgment	Aquino et al. (2023), Cabitza et al. (2017), Koplin et al. (2025), Levy et al. (2019), Lu (2016), Parchmann et al. (2024), Rafner et al. (2022), Smith and Baumann (2020)
	Unwillingness to provide a definitive clinical assessment	Cabitza et al. (2017), Dias Duran (2021), Monteith et al. (2022), Rafner et al. (2022)
	Poorer clinical decision making	Aquino et al. (2023), Dias Duran (2021), Hallowell et al. (2023), Koplin et al. (2025), Levy et al. (2019)
Maintaining patient welfare	Deterioration of moral skills of physicians	Dias Duran (2021), Hallowell et al. (2023), Parchmann et al. (2024), Stogiannos et al. (2025)
	Undermining patient safety	Levy et al. (2019), Rafner et al. (2022)
Organizational	Negative impact on the system's integrity	Monteith et al. (2022), Rafner et al. (2022), Smith and Baumann (2020), Sparrow and Hatherley (2019)
	Negative impact on the clinical team's situational awareness and teamwork	Panesar et al. (2020), Smith and Baumann (2020)
AI specific	Medical error derived from HCI	Panesar et al. (2020), Ruskin et al. (2020)
	Dependence on AI tools, decreased motivation for in-depth research and learning	Akudjedu et al. (2023), Chen et al. (2021), Choudhury and Chaudhry (2024), Nakagawa et al. (2023), Ruskin et al. (2020)
	Inability to understand and challenge AI's outputs	Akudjedu et al. (2023), Choudhury and Chaudhry (2024), Kashou et al. (2024), Lu (2016), Panesar et al. (2020), Ruskin et al. (2020)

3.1.1 Physical examination and identifying physical signs

Physical examination is one of the pillars of the diagnostic process, as it allows the physician to confirm or reject a potential diagnosis, assess the extent of an illness or disease, and provides information necessary in order to devise a treatment plan (Simpkin et al. 2017).

Six of the articles (Hallowell et al. 2023; Levy et al. 2019; Lu 2016; Monteith et al. 2022; Rafner et al. 2022; Ruskin et al. 2020) directly assessed the deterioration of physician skills as a concern derived from deskilling due to the implementation of AI, while another three (Duran et al. 2023; Smith and Baumann 2020) pointed out the loss of manual skills in general, which also includes examination skills, and Aquino et al. (2023); Chen et al. (2021); Monteith et al. (2022) remarked on the reduction of skill development. Ruskin et al. (2020) presented a discussion about the potential effects of AI in anesthesiology and highlighted the consequences of losing manual skills: “Skill at performing a task manually will atrophy as practitioners become reliant upon an automated system; as a result, manual skills may diminish in humans who rely heavily on automation to manage system failures", thus subscribing to Kleim and Jones (2008)’s principle of Experience-Dependent Neural Plasticity, “use it or lose it."

Smith and Baumann (2020) compared the lessons learned of professionals in different areas where automation has taken place. Their study highlighted that manual skills can deteriorate once the clinician has become dependent on technology, even if these skills had been acquired with care through training in medical school. These skills appeared “to degrade over time from lack of use of manual methods, reducing the ability to detect, assess and respond to situations when automation is degraded or unavailable”. This has problematic implications: In the event of technological failure, lack of proficiency in manual examination techniques could result in a critical inability to diagnose patients accurately, significantly increasing the risk of medical errors.

3.1.2 Clinical communication and managing patients’ concerns

Effective communication is an essential part of healthcare, both in providing direct patient care and in facilitating efficient clinical and operational processes within healthcare organizations. It includes, but is not limited to, the ability to gather information from a patient in a structured and systematic manner through history-taking, the ability to effectively convey information to a patient or family member, and the ability to communicate key information to another healthcare professional on the team. Clinical communication is closely related to the ability to effectively manage the concerns of patients, family members and carers because it involves the ability to respond accurately and sensitively to their questions, concerns or challenges, and to demonstrate empathy in such interactions, which requires active listening, empathy, and willingness to achieve shared decision making.

Two main concerns were identified in the literature. First, Aquino et al. (2023), Levy et al. (2019), Lu (2016) and Parchmann et al. (2024) discussed the potential negative effects of deskilling on doctor–patient communication. Lu (2016) argue that a risk of implementation with AI is that physicians would become preoccupied with reading the results of assessments made by the AI tool instead of focusing on direct interaction with patients. As a result, diagnostic reasoning would be negatively affected: “The embodiment and intersubjectivity of doctor–patient interactions are critical to reasoning because in clinical settings, diagnostic reasoning is developed based on more training, immediate pattern recognition of patient symptoms, and features obtained from face-to-face communication”. Similarly, Parchmann et al. (2024) further emphasizes that introducing AI tools to clinical workflows could create a model of human–computer interaction that might neglect the important human aspects in medical decision-making. They argue that patient communication and emphatic skills in the delivery od care require human competence. If replaced by automated systems, which lack these human qualities the principle of beneficence is at risk of being infringed. Second, Akudjedu et al. (2023), Dias Duran (2021), Levy et al. (2019), Lu (2016) and Rafner et al. (2022) highlighted the deterioration of the relationship between the patient and physician. Dias Duran (2021), for instance, discusses the fiduciary nature of the physician–patient relationship and the importance of communication within it. This includes, for example, the positive impact of making relevant decisions with compassion, appropriately informing the patient about these decisions in order to obtain consent and protect their autonomy.

3.1.3 Differential diagnosis

Making a diagnosis is a complex process that requires a range of skills, from factual knowledge about a wide variety of diseases, conditions and syndromes to a keen eye and manual skills to examine the patient. The goal of diagnosis is to determine the causes and consequences of a particular abnormal health condition to develop a treatment plan that will return the patient to a state of health. However, in many cases, a group of signs and symptoms may be indicative of several diagnoses. Thus, the method of differential diagnosis aims to determine which one is the correct one. Elder et al. (2011) defines this skill as the ability to use clinical information gathered from the clinician's own assessment of a patient and various forms of clinical testing to identify the most likely cause of the symptoms. This skill requires the clinician to be able to perform this analysis without extensive preparation and under conditions where information might be limited and time might be critical. Regarding this clinical skill, Cabitza et al. (2017), Kashou et al. (2024), Rafner et al. (2022) and Smith and Baumann (2020) remarked on the possibility of AI tools to affect the diagnostic accuracy of human clinicians, and Choudhury and Chaudhry (2024), Hallowell et al. (2023), Koplin et al. (2025), Levy et al. (2019), Nakagawa et al. (2023), Rafner et al. (2022), Smith and Baumann (2020) and Sparrow and Hatherley (2019) discussed the decrease in clinical knowledge and in the ability to analyze and interpret diagnostic information.

Aquino et al. (2023) conducted a study where interviews were conducted with professionals in different disciplines, including healthcare. One group of participants expressed concerns about losing control over the clinical process due to skill erosion. In particular, one participant criticized the idea of AI reducing the case load. Contradicting one of the most positively perceived benefits of introducing AI in clinical contexts—namely, the automation of repetitive tasks and subsequent decrease in workload—the participant argued that a clinician's expertise is not necessarily improved by the number of years of experience, but by the quantity and difficulty of cases they handle. A reduction of case load would thus reduce the chances of clinicians to acquire expertise in baseline cases and also in complex ones. This is an example of concerns related to the topic of upskilling inhibition. Furthermore, the study conducted by Hallowell et al. (2023), that examined the views of different stakeholders regarding the use of AI tools—such as computational phenotyping—to diagnose rare diseases, found that interviewees remarked on the importance of diagnostic skills as irreplaceable in the clinical process despite of the portended accuracy of AI tools: "Some argued that the input of human expertise was essential to accurate diagnostic decision-making arguing that in addition to facial phenotypes there were a range of other phenotypic features and subjective aspects of human–decision making that were necessary aspects of diagnosis in this context." Nakagawa et al. (2023), in turn, argued that if clinicians stop exercising their learned diagnostic skills on a regular basis, this would have an particularly devastating impact on future clinicians: “if AI triages and manages all the simpler cases automatically; in this scenario, pathologists may lose familiarity with the bread-and-butter cases, resulting in future generations of pathologists who are no longer facile with the majority of lesions.”

3.1.4 Clinical judgment

Clinical judgment is a critical pillar of any medical decision-making process. This skill involves the process of making decisions based on a combination of knowledge, critical thinking, and reasoning. The assessment of this skill, according to Elder et al. (2011), focuses on evaluating the ability of clinicians to investigate a clinical problem based on information gathered about the patient, to present relevant results to the patient, and to subsequently treat one or more aspects of the problem. Clinical judgment is therefore a necessary prerequisite for clinical decision making. There are three potential negative consequences of deskilling related to clinical judgment skills: Poorer clinical reasoning and judgment (Aquino et al. 2023; Cabitza et al. 2017; Koplin et al. 2025; Levy et al. 2019; Lu 2016; Parchmann et al. 2024; Smith and Baumann 2020), unwillingness to provide a definitive clinical assessment (Cabitza et al. 2017; Dias Duran 2021; Monteith et al. 2022; Rafner et al. 2022), and poorer decision-making (Hallowell et al. 2023)

Levy et al. (2019) argue that in the field of ophthalmology, the use of deep learning in combination with optical coherence tomography (OCT) has increased diagnostic accuracy but also created a new set of challenges. Among them, they cite the deskilling of new professionals as an undesirable side effect of over-reliance on automation: “(...) as ophthalmologists in training begin to examine patients and learn about various pathologies, they may be inclined to diagnose, treat, or make decisions based solely on OCT and similar sophisticated technologies, even without thoroughly examining the patient." Similarly, both (Cabitza et al. 2017; Dias Duran 2021) point out that if clinicians become overly reliant on the outputs of AI tools, this could, over time, undermine their professional confidence to make a definitive clinical assessment unaided and to make decisions based on that assessment. Dias Duran (2021) emphasizes that this could become a perpetuating cycle based on the assumption that because AI tools can achieve high accuracy rates, therefore they are always more reliable for complex clinical cases.

Finally, Hallowell et al. (2023) touch upon the topic of the new clinical skills that clinicians would be required to develop to work in environments where AI tools become the standard. In other words, the necessity of reskilling or upskilling of medical professionals. They report that: “AI tools were perceived as making the diagnostic process less reliant on clinical judgment, thus undermining clinicians’ expertise and potentially compromising their ability to validate algorithmic output and identify algorithmic bias." This highlights the need to find a balance between meaningful AI involvement that aligns with the needs and interests of the clinicians, in accordance with contextual requirements and conditions and reinforcing the role of the clinicians so they are able to adapt to the changes brought by the technologies and can harness their benefits. While clinical judgment is not a skill that can easily be measured unless in practice and closely related with the clinical outcome, the ability of physicians to make decisions in a timely, efficient, and accurate manner is essential to clinical practice. The recurrence of concerns connected with clinical judgment seems to highlight that clinicians are similarly concerned about the deterioration of interpersonal or character-based skills as they are with technical ones such as manual skills.

3.1.5 Maintaining patient welfare

Maintaining patient's welfare points to a broader set of skills that has at its core the relationship between clinicians, patients, and family members. According to Elder et al. (2011), it means treating "(...) a patient or relative respectfully and sensitively and in a manner that ensures their comfort, safety and dignity". Four articles expressed concern in the potential deterioration of the moral skills of physicians (Hallowell et al. 2023; Dias Duran 2021; Parchmann et al. 2024; Stogiannos et al. 2025), and three emphasized on the concern that deskilling could undermine the safety of patients (Aquino et al. 2023; Levy et al. 2019; Rafner et al. 2022).

Levy et al. (2019) noted that deskilling could lead to decreased patient trust, as well as increased stereotyping of patients, as also highlighted by Dias Duran (2021). Lack of trust from patients can make the diagnostic process more difficult as the patient may not feel confident disclosing all relevant information to their physician and as result obtaining consent for procedures or treatment might not be possible. As remarked by Dias Duran (2021), trust is an indispensable part of the fiduciary relationship of physicians and patients, even more so, when confronted with the introduction of complex technologies.

Similarly, Aquino et al. (2023) identify three attitudes of interviewees regarding the degree of automation and its consequences: Even attitudes that support automation of some aspects of the clinical process make clear that tasks that require empathy or a human touch should not be automated: "This means that decisions about automation in health care should be limited by consideration of the relational or caring aspects of the clinical task in question." In this study, patient safety refers to the emotional, psychological, and moral well-being of patients, which can be as important as physical or medical safety.

3.1.6 Organizational

During the literature review, we identified concerns that did not fit into the clinical skills framework established in Elder et al. (2011). Rather than focusing on deskilling as a phenomenon of individual clinicians, some concerns raised focused on the systemic implications of workforce deskilling on organizations. Several papers (Sparrow and Hatherley 2019; Monteith et al. 2022; Rafner et al. 2022) suggested that systems could become more fragile, and (Smith and Baumann 2020; Panesar et al. 2020) touched on the negative impact on team situational awareness and teamwork.

Rafner et al. (2022) conceptualized deskilling as a socio-technical phenomenon, i.e. where social and technical aspects need to be considered together, and emphasized that deskilling due to AI implementation also needs to be understood from an organizational perspective. According to them: "(...) the utilization of disruptive advanced technologies requires consideration from multiple perspectives taking into account the longer-term and the potential short-term gain". In a similar fashion, Smith and Baumann (2020) touch upon the possibility of an organization's integrity being compromised as a result of lack or decreased training opportunities available to the health workforce, especially regarding manual methods, and Sparrow and Hatherley (2019) discuss at length the likelihood of system fragility, where some processes might end up relying entirely on AI technology: “the profession as a whole may lose a skill if no one remembers how to perform a task that doctors use to perform before the new technologies arrived." These perspectives reinforce the argument that deskilling and upskilling inhibition must be studied and addressed not as an isolated cognitive phenomena but acknowledging the systemic factors that play a role and the consequences at the organizational level, that thus affect institutions and potentially, healthcare sectors as a whole.

3.1.7 AI-specific

The last group of concerns identified in the systematic review were related to the functioning of the technology itself, or derived specifically due to its implementation. There are three types of concerns: First, Panesar et al. (2020); Ruskin et al. (2020) highlighted the potential of medical error derived from Human–Computer interaction (HCI). Panesar et al. (2020) frames the issue of technical errors in the context of the deterioration of clinical skills development. According to the authors, in an unexpected situation where the model or equipment malfunctions or is unavailable, a deskilled professional is more likely to make an error, which could create the need for additional supervision, increase the likelihood of medical errors, or have a negative impact on the overall clinical workflow.

Second, Akudjedu et al. (2023); Chen et al. (2021); Nakagawa et al. (2023); Monteith et al. (2022); Ruskin et al. (2020) remark on the increased dependence of clinicians on AI tools as they become increasingly ubiquitous within clinical workflows. This dependence goes hand in hand with automation bias, defined as the tendency of a user to attribute authority or excessive trust to an AI tool over advice from other sources. Monteith et al. (2022) argues that a consequence of such a phenomenon is that users, in this case clinicians, follow incorrect advice despite information to the contrary or even personal training. Similarly, Ruskin et al. (2020) suggests that “automation bias may lead a clinician to over-rely on alerts and prescribe medications only when suggested by clinical decision support or computerized provider order entry." Panesar et al. (2020) point out that a consequence of such dependence and the loss of critical skills could come to the detriment of patients or other clinicians in the case of unexpected scenarios, such as a system malfunction or missing equipment.

Finally, Kashou et al. (2024); Lu (2016); Choudhury and Chaudhry (2024) speak about the problematic implications of poor or lack of understanding about the functioning of the AI tools and the resulting inability to identify when could be appropriate or even necessary to challenge AI’s outputs. Choudhury and Chaudhry (2024), for instance, argue that the successful integration of highly complex models such as LLMs would depend largely on the epistemic difference between professionals with more experience and clinical judgment and those who would use the tools precisely because they lack them.

3.1.8 Visualizations

To visually present the results of the systematic review, we used the tools Litmaps and VOSviewer. Figure 2, generated via Litmaps, shows the distribution of sources included in the review according to momentum, i.e. the number of citations (updated to February 2025) on the Y-axis and the year of publication on the X-axis. This visualization maps direct cita-

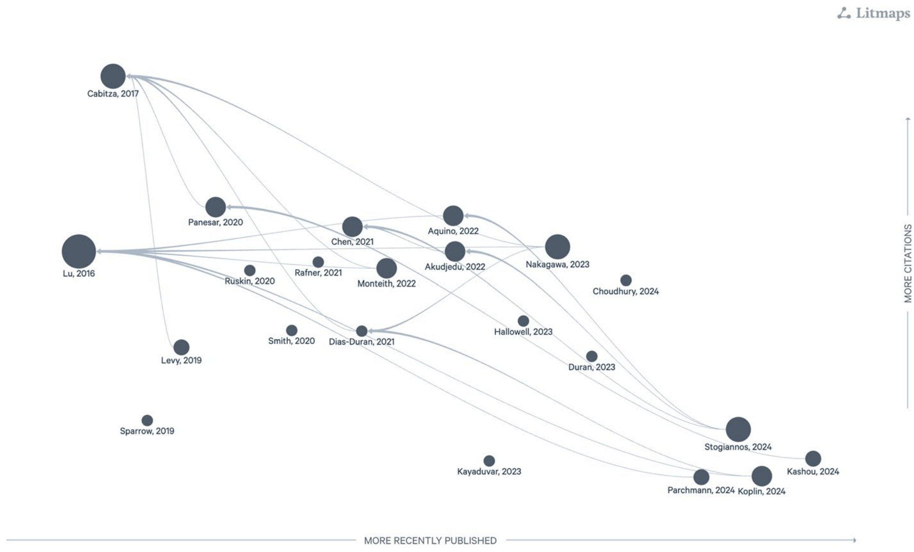

Fig. 2 Network visualization generated on Litmaps according to momentum: Citation count on the y-axis and publication date on the x-axis. The size of the node represents the connectivity, i.e. the number of citations of each article within this map

tion links among publication, highlighting how studies build upon one another. Given that the criteria of the systematic review only included papers that meaningfully discussed the phenomenon of deskilling due to the implementation of AI models in the medical sector, this visualization shows that the topic of deskilling of medical professionals has not yet been widely researched within the time frame of the review.

To analyze the interconnections and core concepts within the reviewed literature, we employed VOSviewer's Network Visualization feature to generate two distinct visual representations: bibliographic coupling (Fig. 3) and concept occurrence (Fig. 4).

The bibliographic coupling network (Fig. 3) reveals the extent to which publications share common references, offering insight into thematic coherence and intellectual clusters within the literature. This method allows for the identification of conceptually related works, even if they do not cite each other directly.

The text-based network visualization (Fig. 4) further illustrates the conceptual landscape of the reviewed literature. This approach identifies key terms and their relationships based on text extracted from the title and abstract fields of 21 retrieved documents, excluding structured abstract labels and copyright statements. Using full counting, we selected terms with a minimum of five occurrences and applied a relevance score to retain the 60% most relevant terms, resulting in 13 selected items. The network reveals how concepts cluster together, highlighting thematic structures and key research topics. Notably, the term deskilling (Cluster 1) exhibits strong connectivity, with 10 links, a total link strength of 58, and eight occurrences, underlining its central role in discussions on AI's impact on professional expertise

By complementing the citation and bibliographic coupling analyses (Figs. 2 and 3) with this term co-occurrence analysis, we gain a more comprehensive understanding of the research landscape—in terms of direct academic influence as well as underlying conceptual

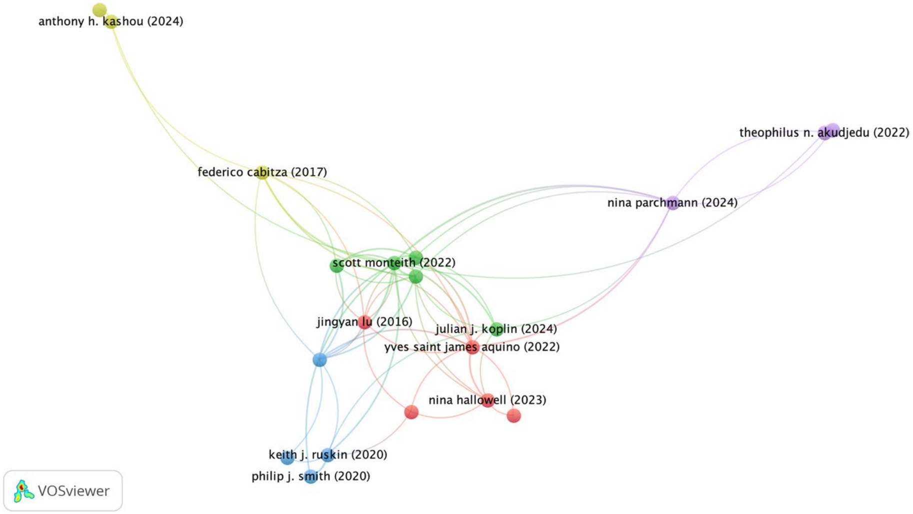

Fig. 3 Network visualization of bibliographic coupling generated by VOSviewer. Each node represents a document, with edges indicating shared references. Thicker lines denote stronger coupling strength, reflecting the number of cited references two publications have in common. Out of 22 documents analysed, VOSviewer retrieved 21, displaying the largest connected set (n = 20), grouped into five clusters. The network comprises 61 links with a total link strength of 89, illustrating conceptual relationships between studies

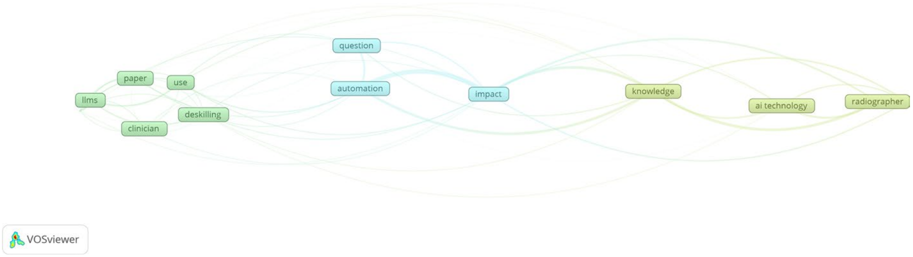

Fig. 4 Network visualization of term co-occurrence generated by VOSviewer. Text data were extracted from the title and abstract fields of 21 retrieved documents, applying full counting with a minimum occurrence threshold of five. Based on relevance scoring, 13 terms were selected, grouped into three clusters. The network consists of 54 links with a total link strength of 620. Notably, *Deskilling* (Cluster 1) has 10 links, a total link strength of 58, and eight occurrences, highlighting its prominence in discussions on AI's impact on expertise

commonalities, tracing the intellectual structure of the literature and the core concepts shaping current discourse.

3.2 Narrative review results

The outcome of this narrative review maps the academic discourse on AI-induced deskilling and upskilling inhibition in medicine. It highlights the current challenges, open questions, and future directions in this evolving field. Our analysis aims to provide healthcare professionals, policymakers, and AI developers with a comprehensive understanding of the implications of AI in healthcare, guiding more informed and ethically sound decisions in the integration of AI technologies in medical practices.

In what follows, we will present each pillar of our categorization framework by outlining their specific importance for depicting the current scenario of deskilling and upskilling inhibition issues, with the aim to contribute to identify, categorize and tackle the broad set of issues, open challenges and future directions of AI-induced deskilling and skill evolution and perception in the medical and clinical domain. A summary of the most salient themes is presented in Table 3.

Table 3 Salient cross-cutting themes emerging from the narrative review, and associated concerns or opportunities with representative references

Theme	Concerns & opportunities	References
Over-reliance and critical thinking erosion	Automation bias, diminished clinical judgement, and reduced diagnostic reasoning due to habitual deferral to AI outputs	Campbell et al. (2020), Monteith et al. (2022), Cabitza et al. (2017a), Green 2019), Wessel (2023), Kapoor et al. (2019), Cabitza (2021), Lu (2016), Banerjee et al. (2021)
Professional role transformation	Shift from clinician to supervisor; redefinition of professional identity; loss of autonomy; devaluation of tacit knowledge	Mosch et al. (2022), Sambasivan and Veeraraghavan (2022), Kundu (2021), Rafner et al. (2022), Aslam and Hoyle (2022)
Training and education challenges	Need to revise curricula; ensure skill preservation; promote AI literacy while maintaining human competencies	Lu (2016), Zulkipli et al. (2023), Banerjee et al. (2021), Zhang et al. (2023), Rao (2023), Vallor (2015)
Ethical and moral deskilling	Decline in ethical sensitivity and moral judgement; challenges to human accountability and responsibility gaps	Vallor (2015), Dias Duran (2021), Hallowell et al. (2023), Iqbal et al. (2022), Da Silva et al. (2022), Gerke et al. (2020), Drabiak et al. (2023)
Socio-technical vulnerabilities	Increased system fragility; technical dependency; inability to verify or override AI failures; broader societal and legal consequences	Sparrow and Hatherley (2019), Panesar et al. (2020), Aquino et al. (2023), Hoff (2011), Tsai et al. (2003), Drabiak et al. (2023), Cabitza et al. (2017a), Morley et al. (2020)
Human–AI Synergy and mitigation potential	Recognition of hybrid intelligence; potential for skill enhancement and resilience through thoughtful integration and role delineation	Rafner et al. (2022), Aslam and Hoyle (2022), Nelson et al. (2020), Kundu (2021), Mofatteh (2021)

3.2.1 Human–AI interaction

Human–AI interaction is the dimension that captures the nature and degree of interaction between human actors and AI systems in the healthcare context. It ranges from low to high, depending on the level of automation (Parasuraman et al. 2000), collaboration, and coordination of human and AI agents. It also reflects the balance of power, control, and responsibility between human and AI actors. This dimension is important for our framework, as it reveals how AI affects the skills, roles, and identities of healthcare workers, such as physicians, nurses, and technicians, as well as the expectations, perspectives, and outcomes of healthcare consumers, such as patients, caregivers, and insurers. Human–AI interaction also influences the quality, safety, and ethics of healthcare delivery, as well as the innovation, learning, and adaptation of healthcare practice.

3.2.1.1 Impact of AI on human skills in organizations This category focuses on the effects of AI on the cognitive, psychomotor, and social skills of healthcare workers, as well as the challenges and opportunities for reskilling, upskilling, and deskilling in the age of AI.

Morandini et al. (2023) conducted a narrative review analyzing the research and practice on the impact of AI on human skills in organizations, shedding light on the phenomenon of upskilling and reskilling. The authors then examine how the introduction of AI impacts the skills required by workers and how AI can support the upskilling and reskilling of employees. The paper discusses the implications for human resource management and organizational development, which is crucial in understanding how AI may affect the skills of healthcare professionals and the potential for upskilling inhibition in the medical field.

Furthermore, Christopher J et al. (2019) highlighted key challenges in delivering clinical impact with artificial intelligence, emphasizing the difficulties in understanding the inner workings of AI models and the implications of their decisions. This is essential in comprehending the potential deskilling effects of AI in medicine and the barriers to upskilling due to the complexity of AI technologies.

In addition, Malik et al. (2019) provided an overview of artificial intelligence in medicine, offering insights into the practical implementation of AI technologies in the medical field. Following (Schemmer et al. 2021), understanding the practical aspects of AI in medicine is crucial for evaluating its impact on human skills and the potential for deskilling and upskilling inhibition. The authors highlight the drawbacks of AI-based automated decision-making, including deskilling of knowledge workers. They propose Intelligent Decision Assistance as a new class of DSS, which would support knowledge workers without influencing them through automated decision-making, using techniques of explainable AI (Schemmer et al. 2021).

Aslam and Hoyle (2022) discuss the broad range of skills that clinicians should develop or refine to fully embrace the opportunities that AI technology will bring. The authors highlight the need for an awareness to identify AI systems that might already be in place and the need to properly assess the utility of their outputs to correctly incorporate the AI system into clinical workflows. They also stress the need for clinicians to cultivate those human skills that are beyond the capabilities of AI systems, and which should be just as important as ever, with the aim to create an harmonious synergy between educated human and machine interaction for the best possible patient experience and care.

Rafner et al. (2022) examine the impact of AI on the skills and competencies of workers in various domains, including healthcare. They argue that AI can both de-skill and up-skill workers, depending on the level of human–AI interaction, the nature of the tasks, and the design of the AI systems. The paper also discusses the concept of Hybrid Intelligence, first proposed in Dellermann et al. (2019), as a form of human-centered AI that aims to achieve mutual learning and collaboration between humans and machines: Hybrid Intelligence can help workers to re-skill, that is, to acquire new skills and competencies that are relevant and valuable in the age of AI. Combining human and artificial intelligence can mitigate the risks of deskilling while maximizing the benefits of AI. This approach suggests a symbiotic relationship between humans and AI, where each compensates for the other's limitations (Rafner et al. 2022).

Moreover, Rao (2023) emphasized the urgent need for healthcare workforce upskilling and ethical considerations in the era of AI-assisted medicine. This reference is particularly relevant as it directly addresses the phenomenon of upskilling inhibition in the medical domain due to the integration of AI technologies.

3.2.1.2 Deskilling types and AI influence on clinical decision-making This category addresses the causes and consequences of deskilling for workers, organizations, and society, as well as the specific implications of AI-induced deskilling, especially in the medical domain where human judgment and empathy are essential.

In the medical domain, the influence of AI on deskilling has been a topic of significant interest. Various dimensions of deskilling have been identified, including technical deskilling, decision-making deskilling, moral deskilling, and semiotic deskilling.

Technical deskilling refers to the reduction in the skills required to perform certain tasks due to the automation and AI-driven tools (Dias Duran 2021). Decision-making deskilling involves the reliance on AI algorithms for clinical decision-making, potentially leading to a decrease in critical thinking and diagnostic skills among healthcare professionals (Green 2019; Wessel 2023). Social deskilling involves the erosion of interpersonal skills and patient interaction as AI systems take over certain aspects of patient care (Winter and Carusi 2022b). Moral deskilling refers to the potential reduction of human users' ability to make appropriate moral judgments and decisions due to over-reliance on technological developments, as envisioned by Vallor (2015). Semiotic deskilling refers to weakening of the human ability to interpret situations, which can exacerbate medical over-use (Cabitza 2021). Therefore, it is strictly related to semiotic desensitization, which is the progressive decrease of responsiveness and sensitivity of physicians with respect to heterogeneous signs deriving from the patients they interact with, in favor of data provided by electronic patient records, registries, and decision support systems (Cabitza 2017).

In the literature, some possible ways to prevent or mitigate different deskilling types are discussed, as in case of moral reskilling or upskilling, which involve the learning of new moral skills or enhancing existing ones through education, training, or reflection (Vallor 2015). These dimensions of deskilling have implications for healthcare workers and organizations, including concerns about job satisfaction, patient safety, and the overall quality of care.

3.2.1.3 Physician expectations and deskilling concerns This category explores the expectations and concerns of physicians regarding the use of AI in their practice, as well as the barriers and facilitators for the integration of AI in their workflow. It also analyzes the impact of AI on the professional identity, autonomy, and satisfaction of physicians.

Overall, the integration of AI in medical practice has raised concerns and expectations among physicians. Physicians' expectations include the potential for AI to predict clinical events, improve diagnostic accuracy, and reduce the burden of disease (Maassen et al. 2021). However, there are concerns about the deskilling of clinicians due to increased reliance on AI (Campbell et al. 2020). There are also concerns that the implementation of certain AI systems may lead to the deskilling of ward doctors and nurses in relation to their ability to care for deteriorating patients (Bunch et al. 2023). Furthermore, Hoff, (2011) emphasized that deskilling among primary care physicians is influenced by both their actions and the pressures in their work context, suggesting that the impact of AI on deskilling or upskilling may be multifaceted.

To identify and analyse the issues related to deskilling concerns among healthcare professionals and stakeholders, Aquino et al. (2023) conducted a qualitative study exploring the views of professionals involved in the development, deployment and regulation of healthcare AI. The paper aims to understand how these professionals perceive the potential benefits and risks of AI-enabled automation in healthcare work, and what values and priorities they have for healthcare work and its relationship to AI. The authors used a constructivist grounded theory approach to analyze data from 72 semi-structured interviews with experts from various disciplines and sectors. The research led to identify three main issues that shape the professional perspectives on AI and deskilling: the extent of automation, the quality of clinical skills, and the model of healthcare work. The paper argues that these issues reflect different values and assumptions about the nature, purpose and value of healthcare work, as well as different expectations about the role of technology in improving or harming clinical outcomes (Aquino et al. 2023). All these findings stress the importance of addressing physicians' concerns and expectations regarding the integration of AI in medical practice to ensure effective and ethical use of AI technologies.

3.2.1.4 Patient-centered considerations, perspectives and safety Patient perspectives on AI-induced deskilling are crucial in understanding the impact of AI on healthcare delivery. This category explores the views and experiences of patients regarding the use of AI in healthcare, in the context of AI-induced deskilling and upskilling inhibition of healthcare workers. It also examines the factors influencing patients' trust, acceptance, and control of AI, as well as the potential benefits and risks of AI for healthcare quality and safety, particularly concerning its potential relationship to deskilling.

Nelson et al. (2020) found that most patients are receptive to the use of AI for skin cancer screening within the framework of human–AI symbiosis, indicating a positive attitude towards AI integration in medical diagnosis. Nevertheless, the study shows that patients are aware of the potential human deskilling induced by AI, as it was reported among the perceived risks (Nelson et al. 2020).

Lennartz et al. (2021) conducted a questionnaire study to investigate patients' opinions on the use of AI in different aspects of the medical workflow and the level of control and supervision under which they would deem the application of AI in medicine acceptable, shedding light on patient-centered considerations over the control of AI-induced deskilling in medical processes. According to the obtained results, patients would trust physicians over AI in most clinical capabilities, while notable exceptions are related to basing treatment decisions on the most current clinical knowledge (Lennartz et al. 2021).

Winter and Carusi (2022a) reported both professional and patient expectations related to AI-supported diagnosis of pulmonary hypertension, finding that concerns related to deskilling deriving from over-reliance are common among patients and professionals.

Monteith et al. (2022) discuss various challenges related to the implementation of AI in psychiatry. The authors point out that the introduction of AI in psychiatry may lead to deskilling through automation bias and workflow changes. Therefore, over-reliance on AI recommendations is pointed out as a crucial point, leading to a reduction in critical thinking and decision-making skills. Additionally, workflow changes driven by AI implementation may lead to a shift in the skill set required for certain tasks, potentially resulting in deskilling. Furthermore, the authors also highlight the possible issues related to safety, especially when thorough human oversight is needed, as the progressive automation of complex tasks in the overall workflow could lead to negative impacts for safety (Monteith et al. 2022).

3.2.2 AI adoption and integration

This dimension captures the organizational and institutional aspects of AI adoption and integration in the healthcare context, as regards the potential or actual relationships with deskilling. It also reflects the degree of acceptance, readiness, and trust of AI systems among healthcare stakeholders, assessing how such factors influence the presence or risk of deskilling. Findings span across the themes of domain knowledge incorporation in AI, the evolving role of education and training, as well as ethical and legal implications.

3.2.2.1 Incorporation of domain knowledge in AI models This category focuses on the integration of domain knowledge, such as medical knowledge, into AI models. In our framework, this category can help to identify and understand the actual trade-off between the utility of incorporating specific domain knowledge into the AI models and systems used in the healthcare field, and their influence on the emergence, presence, or maintenance of deskilling or upskilling inhibition.

Overall, the incorporation of domain knowledge into AI models has raised concerns about AI-induced deskilling, especially in areas where domain experts are essential. In a broad perspective, Sambasivan and Veeraraghavan (2022) have highlighted how AI development, particularly in low-resource contexts, often results in the deskilling of local domain experts: in these scenarios, domain experts are frequently relegated to the role of data collectors, performing standardized tasks to create datasets. This process often overlooks the nuanced and in-depth knowledge these experts possess. As a result, there's a growing argument for re-envisioning the role of domain experts in AI development, advocating for their involvement as full-range technical experts who co-create datasets and models. This approach recognizes the value of their expertise beyond mere data collection, emphasizing a partnership that leverages their deep domain knowledge for more effective and contextual AI solutions (Sambasivan and Veeraraghavan 2022).

The incorporation of domain knowledge in AI models can lead to deskilling in medicine due to the potential over-reliance on AI systems, which may result in reduced critical thinking and decision-making skills among healthcare professionals: Campbell et al. (2020) studied the application of AI for diagnosis and management of glaucoma in adults, reporting concerns that deskilling could be stimulated by increasing reliance on automated image analysis, leading to lower clinicians' ability to make decisions based on clinical signs. Kapoor et al. (2019) stress the risky limitation of incorporating AI and automated decision systems in medicine, since it could boost deskilling by reducing the physicians' ability to make informed decisions, form critical opinions according to the whole set of patients' signs and symptoms. As a case study, they report the study of Tsai et al. (2003), which shows that the diagnostic accuracy of internal medicine residents decreased when electrocardiograms were incorrectly annotated by a computer-aided system.

3.2.2.2 Education and training This category examines the needs and methods for evolving medical education and training in the age of AI, in order to prepare and equip healthcare workers with the necessary knowledge, skills, and attitudes to work with AI systems.

Education and training can help to prevent or overcome the potential deskilling or upskilling inhibition of healthcare workers, by providing them with more competencies and evolving technical skills, as well as more innovation, adaptation, and resilience. As AI takes over more routine, data-intensive tasks, the risk of an impressive shift of needed skills for healthcare professionals necessitates a reevaluation of medical training and practice, where the role of the healthcare professional might evolve from being the primary diagnostician to more of a supervisor or manager of patient care, integrating AI insights into a holistic treatment approach (Kundu 2021).

Lu (2016) discusses the impact of medical technology on health care, especially on the cognitive skills and communication abilities of doctors, and considers the advancement of education and the evolution of training to physicians an essential pillar in managing the many wide-ranging impacts of technology in the medical field. Overall, the paper argues that technology can negatively affect doctor-patient interactions, physical examination skills, and clinical knowledge development. According to this study, three possible approaches to medical education and professional development to help doctors better understand and use technology in their practice. The approaches are: (1) providing training on using medical technology, emphasizing the importance of not undervaluing medical technology while promoting its optimal use, (2) helping students recognize the importance of clinical evidence, stressing the need for healthcare professionals to focus on evidence from patient history and physical examinations in clinical decision-making, and (3) helping doctors establish autonomy while interfacing with medical technology, which involves building confidence in their medical judgment, especially as they gain experience (Lu 2016). This study underlines the importance of maintaining a balance between delegating tasks to machines and ensuring that human doctors do not lose their critical role and skills in the medical process. It urges doctors to take control of when, how, and why to use technology, rather than letting technology dictate these aspects.

Banerjee et al. (2021) performed a survey of trainee doctors' perceived impact of AI technologies on clinical training and education at UK NHS postgraduate centers in London in 2020: the respondents expressed mixed perceptions of AI's impact on their training, with concerns about reduced development in clinical judgment and practical skills due to reliance on AI systems. However, they also acknowledged AI's potential to improve training in research and audit.

Also, there are proposals for integrating AI-based research projects into medical education, focusing on the importance of ethical use and critical understanding of AI tools (Zulkipli et al. 2023).

Importantly, this evolution of the medical training curricula has to be carefully designed and implemented, since the misuse of conversational agents and generative AI could negatively impact the educational targets and the clinicians' skills: Zhang et al. (2023) express the concern of the over-reliance on ChatGPT as the primary source of information in training exams and assignments, leading to cheating and risky impact on the trainee's curricula.

Overall, while AI has the potential to upskill medical professionals by providing them with tools for more efficient and accurate work, there's a simultaneous risk of deskilling, particularly in areas requiring deep clinical judgment and decision-making. Balancing these two aspects is crucial for the future of medical education and training (Zhang et al. 2023).

3.2.2.3 Ethical and legal implications This category examines the ethical and legal issues and challenges that arise from the use of AI in healthcare, such as the impact on human dignity, rights, and values, the responsibility and liability for AI decisions and actions, and the regulation and governance of AI systems. It is relevant for our framework, as it reveals how AI can affect the potential deskilling or upskilling inhibition of healthcare workers, by influencing their professional identity, autonomy, and satisfaction, as well as their ethical awareness and judgment.

The integration of AI in medicine raises significant ethical and legal concerns, particularly regarding the potential deskilling of healthcare providers. Overall, the potential deskilling of clinicians due to increased reliance on AI is a significant concern, since the inclusion of AI in medical processes may lead to changes in job profiles of physicians and demands for new categories of medical professionals (Mosch et al. 2022).

The ethical and legal challenges of AI in medicine extend to patient privacy, access to high-quality data, and the over-reliance on AI by healthcare professionals (Iqbal et al. 2022).

Furthermore, the use of AI in healthcare has the potential to undermine the patient-provider relationship, contribute to the deskilling of providers, and introduce algorithmic bias that may be challenging to detect (Da Silva et al. 2022; Gerke et al. 2020).

Gerke et al. (2020) also underline the possible implications of transparent algorithms in the medical processes, which could lead to transformations of the diagnostic processes beyond the skills of clinicians (and also patients) to understand.

Over-reliance has concrete ethical implications in medicine: Iqbal et al. (2022) reported the concrete risk that it could lead to neurosurgeons not learning, acquiring and mastering the needed surgical skills and techniques. Also (Mofatteh 2021) expresses similar concerns, while reporting both potentially positive and negative implications of the use of AI in neurosurgery: on the one hand, he agrees that AI could lead to over-reliance and discouragement in acquiring the learning skills required to become experts of surgical techniques; on the other hand, it points out the possibility of creating a virtuous skills synergy between AI and human skills, to make them collaborate positively and expand the techniques of neurosurgeons, avoiding deskilling.

Similarly, Panesar et al. (2020) discuss both the potential benefits and the risks associated with the increasing use of AI in neurosurgery, highlighting the implications of deskilling in the field. According to the study, this deskilling could be due to factors like over-reliance on AI, poor understanding of AI processes, overconfidence, and a lack of necessary vigilance in an automated clinical workflow. A key relevant consideration reported is related to manual skills: as regards specifically neurosurgery, the authors speculate that maintaining a core emphasis on manual skills may lessen de-skilling: though, in practice this could be difficult to uphold, due to external constrains and the ethical duty to provide patients with the best possible treatment available (Panesar et al. 2020).

Furthermore, the ethical, legal, and societal implications of using AI systems in breast cancer care have been reviewed, highlighting the need to address these implications: according to Carter et al. (2020), in case of breast cancer evaluation, deskilling could be avoided by including AI as a support or further check to existing human readers' decision, rather than replacing readers, as well as by revealing AI outputs only after human decisions have been made.

Morley et al. (2020) reveal how the risk of AI-induced deskilling of practitioners, due to the over-reliance on AI and machine diagnostics, not only is a problem for the individuals (both the clinicians and the patients involved in the diagnostic process), but it also introduces a broader societal effect, since it could lead to repeated ethical concern of misdiagnosis or missed diagnosis, potentially affecting hundreds or even thousands of people during time.

Moreover, serious disruptions of performance or inefficiencies may occur in case the exploited technology fails or breaks down (Cabitza et al. 2017a).

Drabiak et al. (2023) extend the scenario of the ethical, societal and legal implications of AI-induced deskilling as a consequence of over-reliance on AI and machine learning: they foresee a scenario where eventually the human medical know-how will decrease to a point where physicians has only the possibility to perform simple checks on the machine, while losing the ability to identify, select and correct AI errors. In a broader social and ethical perspective, such a situation will lead to a society where AI will totally assume the responsibility for healthcare outcomes. On the other hand, according to the authors, deskilling will also impact the continuous improvement of AI algorithms, making it harder to train new and improved systems: since the legal standards of care evolve alongside AI intervention, inclusion and adoption in medical processes, it will be necessary to make them evolve to both incentivize clinicians to use technology, but avoiding stimulating over-reliance on it, to avoid misuse and potential individual, social and legal problems.

4 Discussion and research agenda

To our knowledge, this literature review is the first of its kind to investigate the under-studied phenomena of AI-induced deskilling and upskilling inhibition in the medical field using a mixed-method approach. Leveraging systematic and narrative review methodologies, this study sheds light on how AI reshapes clinical roles, workflows, and decision-making processes in terms of deskilling dynamics, presenting both challenges and opportunities for skill acquisition, enhancement and maintenance for healthcare professionals.

Given the frontier nature of the concepts of AI-induced deskilling and upskilling inhibition, the structured approach of the systematic tackled a carefully curated corpus of

22 papers, presenting a structured exploration of this niche literature. This was achieved through the application of the PRISMA methodology, prioritizing the highest quality of evidence and relevance, in a commitment to depth over breadth. The review revealed concerns over AI's influence on clinical judgment, including its potentially deskilling influence in the practices of differential diagnosis and physical examinations. Furthermore, it addresses the challenges AI poses to clinical communication, patient relationships, and overall clinical workflow, underlining the comprehensive range of AI's potential unintended consequences on professional medical skills.

The narrative review widened the scope of the research, considering a wide-ranging set of 62 papers. This comprehensive approach enabled us to examine a wide range of implications and challenges, organizing the discourse into two primary macro-dimensions: *Human–AI Interaction* (Sect. 3.2.1) and *AI Adoption and Integration* (Sect. 3.2.2).

Under *Human–AI Interaction*, we explored the impact of AI on organizational roles and human skills, identified various types of deskilling and AI's influence on clinical decision-making processes, and assessed physicians' expectations alongside their apprehensions regarding deskilling. Patient-centered considerations and perspectives on safety further enriched our understanding of AI's role in healthcare from a multi-stakeholder perspective.

The second macro-dimension, *AI Adoption and Integration*, scrutinized the critical incorporation of domain knowledge into AI models, highlighting the imperative of aligning AI advancements with educational and training frameworks. This segment also confronted the ethical and legal implications of AI in healthcare.

The systematic and narrative reviews converge on several key themes, highlighting shared concerns and implications regarding AI-induced deskilling and upskilling inhibition in healthcare. First, both reviews underline the pervasive issue of clinician over-reliance on AI systems, a tendency linked to automation bias, reduced critical thinking capabilities, diminished diagnostic reasoning, and declining clinical judgement. This over-reliance is problematic as it leads to a progressive under-utilisation of valuable human expertise and decision-making abilities essential in medical practice.

Additionally, the reviews highlight the possibility of a profound transformation of professional roles, with clinicians increasingly relegated to supervisory roles rather than active decision-makers. This shift has important implications for professional identity, autonomy, and the retention of tacit knowledge, potentially altering fundamental aspects of clinical practice and professional self-conception.

Training and curricular adaptation emerge consistently as another critical theme. There is an urgent need for medical education to integrate AI literacy comprehensively, while carefully preserving foundational clinical competencies, including manual skills and clinical reasoning. This balance is crucial to prevent the inadvertent displacement of essential human expertise, ensuring that healthcare professionals can effectively interpret, manage, and oversee AI-driven processes.

Ethical and moral deskilling is also emphasised, characterised by declining ethical sensitivity, weakened moral judgement, and growing challenges related to accountability. The opacity and complexity of automated decision-making processes complicate the clear attribution of responsibility, potentially diluting clinicians' ethical engagement and moral agency in patient care.

Moreover, the reviews caution against significant socio-technical vulnerabilities arising from high levels of AI dependency. Increased reliance on AI technologies introduces organisational fragility, disrupts teamwork and situational awareness within clinical teams, and generates broader legal and safety implications. These vulnerabilities necessitate careful consideration of organisational resilience strategies and clear guidelines on human oversight and accountability.

On an optimistic note, both reviews recognise the potential benefits associated with hybrid intelligence as thoughtfully designed human–AI collaborations capable of enhancing rather than diminishing clinical skills. Hybrid intelligence frameworks offer practical pathways for mitigating deskilling risks through reflective interactions, complementing human skills, and fostering professional adaptability and continuous learning.

Despite these valuable insights, significant gaps remain evident in the current literature. Particularly, there remains a conflation between genuine skill erosion and the phenomenon of upskilling inhibition. Additionally, there is an absence of clearly defined metrics and structured approaches to empirically assess deskilling and upskilling inhibition in practice. Addressing these gaps through targeted research and practical intervention frameworks will be essential for maximising the benefits and minimising the risks associated with the integration of AI in healthcare.

4.1 Ethical implications of deskilling in expert work

AI-driven deskilling raises ethical concerns that extend beyond immediate performance gaps to normative considerations, touching on professional integrity and public trust. In medicine, the loss of clinician abilities due to AI reliance has prompted questions about what ought to be done to safeguard patient welfare in light of these emerging risks and responsibilities. These concerns are not unique to healthcare. Other high-stakes fields—law, aviation, finance—face parallel risks as experts increasingly lean on AI. Thus, exploring the ethical dimension of deskilling means scrutinising the consequences of having a deskilled workforce and determining what should be done to mitigate harm and maintain trust in care delivery. Below, we examine four dimensions of ethical implications (technical skills, moral skills, individual responsibility/autonomy, and organisational capacity) using healthcare as a primary context and drawing comparisons to these domains.

4.1.1 Technical skills and expertise

Heavy reliance on AI can erode the technical skills and domain expertise of human professionals, raising ethical issues around competence and safety. Most of the articles included in the systematic review focused on deskilling in medicine from a technical or clinical perspective, and among them, the main concerns were the deterioration or loss of manual skills and of specialized clinical skills. Among the most frequently reported concerns were the deterioration or loss of manual skills and the erosion of specialised clinical competencies. A comprehensive review of empirical evidence indicates that trainees' retention of technical skills, such as procedural accuracy and speed, declines markedly when opportunities for hands-on training are reduced (de Andres Crespo et al. 2025).

This degradation in practitioners' manual dexterity, diagnostic acumen, or decision-making capabilities, whether driven by over-reliance on AI or a lack of practical engagement, poses significant risks for healthcare delivery in terms of quality and acceptability (Weidener et al. 2024), resulting in suboptimal or even harmful patient outcomes. This scenario represents a breach of core bioethics principles (Beauchamp and Childress 1994), as it undermines beneficence (the obligation to act in the patient's best interest) and violates non-maleficence (the duty to "do no harm"). Every patient entrusts their well-being to healthcare professionals under the assumption that those professionals possess the requisite skills – an assumption enshrined in the clinician's ethical duty of care. If AI-induced deskilling leads to a loss of critical clinical competencies, it directly challenges that duty of care and the covenant of trust between patient and provider.

Similar patterns appear in other fields, such as aviation: here, extensive automation has been linked to a decline in pilots' manual flying and monitoring skills. The U.S. Federal Aviation Administration thus issued new guidance urging airlines to ensure pilots regularly practice manual control–a move reflecting a recognition that the fundamental skills to uphold passenger safety cannot be substituted by autopilots, and must be practiced continuously (Mooty 2022).

4.1.2 Moral skills, responsibility and autonomy

Beyond technical know-how, AI-induced deskilling of professionals may also involve practitioners losing their moral and ethical skills—a phenomenon described as moral deskilling (Vallor 2015). In medicine, clinicians cultivate moral competencies over years, as they are closely intertwined with how they communicate sensitive medical information to patients, how they manage their concerns about the diagnosis and prognosis, and how they respect the patient's autonomy in making decisions about their own care. Over-reliance on AI decision aids can deskill practitioners in these areas: if a diagnostic AI provides an "explanation" or treatment recommendation, clinicians may begin to trust the machine's judgement even on value-laden decisions, or communicate with patients in a more perfunctory manner.

Over time, this diminished ethical reasoning capacity could mean that a doctor is less prepared to recognise when an AI's suggestion conflicts with a patient's best interests or personal values. Likewise, communication competence can suffer—for instance, if physicians come to lean on AI-driven decision aids, they might not practice the nuanced conversations needed to explain risks and alternatives to patients. This can harm the informed consent process, as patients may receive less comprehensive or less human-centred explanations about their care options. Thus, clinicians with poorly developed moral skills may be unable to safeguard the patients' welfare, which may ultimately have a negative impact on their emotional and psychological well-being, may deteriorate the trusting clinician-patient relationship, and may risk dehumanizing the clinical encounter.

This concern is similarly present in other high-expertise domains. Judges and lawyers, for instance, face pressures to use AI risk assessments or legal analytics; if they begin to treat algorithmic outputs as inherently objective, they might stop exercising their full moral judgment on matters of justice and equity, as in the moral complacency observed in the infamous case of judge reliance on the highly discriminatory COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) system (Kaas 2024). In the military context, where life-and-death decisions are increasingly aided by AI, Vallor (2013) observed that deskilling can even erode virtues like courage and mercy. Ultimately, preserving moral competence is essential for professionals to fulfill their roles as ethical agents and to prevent a slow dehumanization (Woodruff et al. 2024) of high-stakes decision processes.

AI-induced deskilling also complicates individual responsibility and autonomy. In healthcare, if a clinician becomes dependent on an AI's outputs, their ability to exercise independent judgment may wane. This raises the question: who is accountable when something goes wrong? A deskilled practitioner might argue that an adverse outcome is the fault of a misleading algorithm, while patients still rely on the clinician's duty of care. Indeed, over-reliance on technology can blur the allocation of responsibility and accountability for errors. Ethically and legally, current frameworks expect a human professional to be ultimately responsible, but that assumption is strained when the human's role has shifted to merely overseeing an AI. Scholars have described this as creating "moral crumple zones" (Elish 2019), where a human operator ends up absorbing blame for system failures even if the system's complexity limited their control. Notable examples come from aviation, where documented conflict between the pilot and automation resulted in loss of life as pilots are left "out of the loop" (Evjemo and Johnsen 2019) of the complex autopilot system, a phenomenon also investigated as "automation surprise" (Dehais et al. 2015).

Ensuring that the human remains "in-the-loop"-and meaningfully in control (Santoni de Sio and Van den Hoven 2018)—mitigates the risk that responsibility becomes so diffused that no one takes ownership of outcomes. In high—stakes domains, such ambiguity can undermine trust: patients, clients, and the public expect clear accountability for critical decisions. In law, judges might experience a loss of decisional autonomy if, say, guidelines or predictive algorithms strongly recommend certain sentencing decisions (Kaas 2024); similarly, in finance, credit officers might find themselves deferring too readily to algorithmic risk scores when evaluating life-altering decisions such as mortgage approvals. Both may feel pressured to follow the AI's output, potentially "abdicating" some of their responsibility in a manner akin to the complacent pilot (Kaas 2024). Thus, ethical AI deployment must address the "responsibility gap" problem head on—through clarity in roles, audit trails of human–AI decision processes, and professional standards that emphasize human oversight.

4.1.3 Organisational capacity and resilience

Finally, AI-induced deskilling can weaken the organisational capacity of institutions by eroding collective expertise, teamwork, and the training pipeline of future experts. In healthcare, if many clinicians rely uncritically on AI, a hospital could see a general decline in staff's procedural skills and clinical intuition. Over time, this not only impacts individual performance but also the functioning of teams. A skilled healthcare team relies on each member's competence and the ability to communicate and trust one another's judgments. When deskilling sets in, clinicians may begin to doubt colleagues' abilities or struggle to coordinate effectively. For example, junior doctors may find that senior physicians-who perhaps leaned on AI for years-are less able to mentor them in nuanced hands-on techniques or complex decision-making. Such organisational deskilling undermines the integrity of the profession in the long run. The same dynamic is a risk in law firms, airlines, and financial institutions. Law firms traditionally operate on an apprenticeship model where senior lawyers train junior associates through supervised practice. If automation shortcuts much of the junior lawyers' work (for instance, document review or drafting) and seniors themselves begin to rely on AI outputs, the firm's reservoir of human legal expertise may shrink over time. In aviation, an analogous concern is that an entire airline crew corps could lose its edge in handling emergencies. Commercial aviation has already seen incidents attributed partly to pilots' diminished manual flying proficiency in an automated cockpit environment (Mooty 2022). If most flights are on autopilot and simulators don't fully recreate surprise scenarios, the airline as an organisation becomes less resilient to rare crises. On a societal scale, Kim and Scheller-Wolf (2022) argue that broad deskilling trends could strain social cohesion, safety nets and wellbeing, destabilising economic and social systems.

Across technical, moral, personal, and organisational dimensions, AI-induced deskilling poses ethical challenges in fields such as medicine, law, aviation, finance, and beyond. These domains deal with irreducible uncertainties and value-laden decisions that require human wisdom which must not be left to atrophy in the age of AI. This informs the need to conduct in-depth investigations on this phenomenon and develop strategies (from better interface design to policy and training interventions) to mitigate deskilling and ensure the active engagement of professionals-technically sharp, morally attuned, and firmly in control.

4.2 Research agenda for healthcare (and beyond)

Effectively addressing the ethical concerns highlighted above demands a coherent research agenda that prioritises conceptual clarity, methodological rigour, and actionable recommendations. Future research should rigorously examine how skills evolve under conditions of increased AI integration, acknowledging that competencies in healthcare-and comparably, other high-stakes fields-may transform rather than simply diminish. Our proposed agenda, which is presented in Table 4, aims to achieve this with a double focus: (1) conceptual clarification and empirical investigation, and (2) practical interventions. Although informed by the specific phenomenon of deskilling within medicine, the proposed strategies and insights are deliberately formulated to be applicable to broader high-expertise contexts, drawing meaningful parallels with fields such as aviation, law, and finance discussed in previous sections.

4.2.1 Conceptual clarification and empirical investigation

Achieving conceptual clarity is a crucial first step in advancing research on AI-induced deskilling and upskilling inhibition. These two phenomena-one reflecting the erosion of existing expertise and the other a diminished opportunity to acquire new skills-should be recognized as distinct yet interconnected, affecting users of varied expertise levels in different ways. Establishing a shared understanding will enable more precise investigations and the development of targeted investigations and interventions to mitigate skill erosion and ensure the appropriation of necessary competencies. More light should be shed also on the instances where expertise, rather than vanishing, may instead be shifting in nature, a phenomenon termed *re-skilling* (Rafner et al. 2022). In such scenarios, clinicians might trade some manual skills for new proficiencies in interpreting and guiding AI recommendations, in turn focusing more deeply on patient-facing tasks. Clarifying whether we are observing outright skill loss versus a transformation of skill sets-or a suppression of skill gain-has tangible implications for how we design interventions. Distinguishing deskilling from a failure to upskill, as well as characterizing on a case-by-case fashion the desirability of *shifting* skills, is essential for developing training programs and policies that address the correct problem: preserving crucial traditional skills, promoting complementary new skills, and

Table 4 Research agenda on ai deskilling evaluation and mitigation

Research domain	Key focus areas and strategies	References
Conceptual clarification	Distinguishing between AI-induced deskilling, upskilling inhibition, and re-skilling phenomena; Clarifying skill transformations versus outright loss or inhibited skill gain	Vallor (2015), Green (2019), Dias Duran (2021), Cabitza (2021), Rafner et al. (2022), Winter and Carusi (2022a), Wessel (2023)
Empirical investigation	Real-time monitoring and longitudinal studies to detect practitioner–AI interaction shifts and evaluate long-term impacts on competencies and outcomes	Tsai et al. (2003), Campbell et al. (2020)
Qualitative and mixed-methods approaches	Employing qualitative methods (interviews, focus groups, case studies) and mixed-method designs to capture dimensions of professional identity, job satisfaction, trust, and safety	Nelson et al. (2020), Lennartz et al. (2021), Banerjee et al. (2021), Kundu (2021), Maassen et al. (2021), Winter and Carusi (2022a), Mosch et al. (2022), Bunch et al. (2023), Aquino et al. (2023)
Mitigating publication bias	Emphasising failure analyses and transparent reporting; Extending evaluation metrics beyond accuracy to include skill retention, cognitive engagement, and autonomy	Dickersin et al. (1987), Natali et al. (2024)
Hybrid intelligence and AI design	Developing AI tools that actively engage practitioners, reinforcing professional reasoning and preventing skill atrophy through context-aware and frictional interactions; Developing explicit collaboration protocols guiding practitioners on effective AI utilization for decision-making tasks	Dellermann et al. (2019), Schemmer et al. (2021), Cabitza et al. (2021), Miller (2023), Cabitza et al. (2024, 2025)
Interdisciplinary collaboration	Encouraging clinician involvement in AI development, fostering interdisciplinary collaborations, and continuous feedback loops for refining AI applications	Mosch et al. (2022), Iqbal et al. (2022), Sambasivan and Veeraghavan (2022)
Educational initiatives and complementary human-centric skills	Integrating comprehensive AI education into professional curricula, continuous professional development programs, ethical workshops to enhance practitioner autonomy and decision-making; Fostering non-replicable AI skills such as empathy, ethical judgment, and complex decision-making to sustain essential human dimensions at work	Vallor (2015), Kundu (2021), Aslam and Hoyle (2022), Rao (2023), Zhang et al. (2023)
Cross-domain insights	Adapting strategies and governance principles from different high-stakes fields (healthcare, law, finance, aviation, military) contexts	Carrel (2018), Ruan (2020), Browning (2024), Amer et al. (2024), Odonkor et al. (2024), Vallor (2013), Golfetti et al. (2021), Talib et al. (2025)
Policy and governance	Advocating for evidence-based policies addressing AI-induced skill dynamics, promoting ethical AI integration, and safeguarding practitioner autonomy	Lu (2016), Drabiak et al. (2023)

ensuring that the introduction of AI in healthcare actually augments the overall expertise of medical professionals.

Then, there should be a concerted effort to conduct *real-time monitoring of AI applications in situ* (e.g. in hospital wards, courtrooms, cockpits). Such monitoring would not only provide a more current snapshot of AI's impact but also allow for the capture of emergent trends and potential issues in their infancy. For example, by tracking clinician–AI interactions and decision outcomes in real time, subtle shifts-such as delays in clinician response or reduced use of independent judgment-can be detected before they culminate in significant skill decay or patient safety events. Such monitoring efforts should be coupled with longitudinal studies are crucial to map the trajectory of AI's influence on skills, practices, and outcomes over extended periods. These studies can reveal patterns that might be invisible in one-off assessments, offering a richer understanding of AI's long-term effects on expertise development. This includes identifying whether initial over-reliance on AI stabilizes or worsens over time, and how the presence of AI alters the career-long skill progression of workers. For instance, does a junior lawyer who frequently uses an AI research assistant develop the same analytical skills as one trained without it? Only by combining real-time data with longitudinal follow-ups can we discern transient adaptation effects from genuine deterioration of skills, thereby informing timely interventions.

Methodologically, a *mixed-methods research approach* is essential to complement these quantitative measures and provide a more holistic view of AI's role in skill development. Quantitative performance metrics alone cannot capture shifts in professional identity, trust, or autonomy that accompany AI adoption. Qualitative methodologies could complement them by gathering deeper insights on the experiences and perceptions of healthcare professionals and patients. In-depth interviews, focus groups, and case studies can explore how practitioners perceive and adapt to AI in their work, including the impact on their professional identity, job satisfaction, and patient trust. For example, a longitudinal study might quantify changes in diagnostic accuracy and speed when doctors use an AI tool, while interviews with those doctors reveal nuances about their trust in the tool or changes in their decision rationale. By triangulating data in this way, researchers can validate findings across methods and uncover latent effects that single-method studies might miss, providing a more robust evidence base for policy and practice recommendations.

However, a significant obstacle to balanced research on the topic of AI-induced deskilling is publication bias (Dickersin et al. 1987): the tendency to favour studies that reject the null hypothesis. In the context of AI in healthcare, this results in disproportionate emphasis on AI's successes, while negative or inconclusive findings-such as AI-induced errors, automation bias, and disruptions to clinical workflows-are under-reported, leading to unrealistic expectations about AI's capabilities and obscuring its potential risk. To address this, future research must *prioritize failure analyses*, documenting cases where AI contributes to skill degradation, misdiagnosis, or patient safety risks. *Transparent reporting* should be encouraged in journals and conferences, for example by requiring AI impact assessments that present both benefits and risks for a more balanced understanding of AI's role in healthcare. Additionally, *evaluation metrics* must extend beyond accuracy and efficiency (Natali et al. 2024), incorporating human-centered factors such as skill retention, cognitive engagement, and professional autonomy to develop more robust assessment frameworks.

4.2.2 Practical interventions

Future research should not only watch for harms but also evaluate how AI might be intentionally leveraged as a solution to the deskilling problem. A combination of thoughtful design and robust governance is required to prevent AI-induced deskilling in high-expertise domains. Design-centred innovations can make AI a tool for continued learning and engagement rather than a source of complacency. Meanwhile, organisational policies, training regimes, and regulatory frameworks create the professional environment and incentives for experts to remain skilled, vigilant, and morally grounded.

An emerging theme in Hybrid Intelligence (Dellermann et al. 2019) is the design of AI technologies that function as interactive training tools or skill augmenters rather than passive automation devices. For example, following Explainable AI techniques, AI systems should provide transparent reasoning or feedback that the human can learn from. Examples include support systems that not only provide recommendations but also prompt (or even provoke) practitioners with relevant, if also dissenting, questions or explanations (Miller 2023; Sarkar 2024; Reingold et al. 2024; Cabitza et al. 2025), tailored on the user's skill level, thereby reinforcing clinicians' reasoning processes. In another fashion, AI diagnostic tool could intentionally abstain from giving a definitive interpretation in certain cases (Campagner et al. 2019), prompting clinicians to perform their own analysis before revealing the AI's suggestion. Through similarly *frictional* approaches (Cabitza et al. 2024), the AI becomes a learning tool, ensuring the human professional remains critically engaged. We advocate for monitoring studies to explicitly evaluate such constructive use cases: Are AI-driven tutoring systems improving clinicians' proficiency over time? Do context-aware decision aids help maintain diagnostic acumen? By gathering data on these questions in real clinical settings, the field can identify which design features successfully counteract skill atrophy.

Clear interaction protocols can be built into the design of AI systems, defining how responsibilities are shared between the human expert and the AI. Rather than ad-hoc use, practitioners would follow established guidelines for when to trust, when to verify, and how to integrate AI input. For example, in radiology practice a "double-reading" protocol might be instituted: the radiologist and the AI each interpret an image independently and then compare results, a process shown to maintain diagnostic vigilance (Cabitza et al. 2021). Cabitza et al. (2023) names this a *human-first* approach, highlighting how human–AI collaboration protocols can leverage AI for support while keeping ultimate judgment in human hands. These design-centred rules and affordances make the human's active participation an integral part of the technology's use, thereby sustaining expertise and team resilience (e.g. a flight crew or surgical team maintaining robust communication and cross-checking even with AI support).

Effective AI integration into healthcare and other high-expertise fields requires practitioners' active involvement in AI development and implementation. Encouraging interdisciplinary collaboration among experts, AI developers, ethicists, and patients can ensure that AI tools align with clinical realities and ethical standards (Mosch et al. 2022; Iqbal et al. 2022). Regular feedback mechanisms with end-users are crucial for continuous refinement, particularly in complex diagnostic scenarios.

However, even the best interface design will not fully prevent deskilling without supportive governance, training, and organisational policies. Governance and organisational interventions create an environment in which human expertise is continuously cultivated and safeguarded as AI is integrated. These interventions span education, policy/regulation, and institutional practices to ensure that professionals remain capable, ethical, and ready to step up when AI is unavailable or incorrect.

Professional training curricula and continuing education should be updated to reflect the presence of AI while actively counteracting skill erosion. This involves two aspects: (1) teaching practitioners about AI—its limitations, failure modes, and proper use—to foster informed usage, (2) reinforcing core domain skills that AI might otherwise overshadow, and (3) identifying and strengthening skills that whose "human touch" makes them complementary to AI. For example, medical schools can incorporate modules on interpreting AI outputs and require students to diagnose cases without AI assistance to ensure they can function without decision support, as well as putting more emphasis on the development of skills such as moral judgment and empathy.

Organisations can institute requirements that professionals regularly practice tasks without AI, maintaining proficiency. Aviation provides a clear precedent, with pilots being encouraged to log a minimum number of manual flying hours to combat automation-induced skill fade (Mooty 2022). Healthcare institutions could similarly require clinicians to periodically perform clinical reasoning exercises unaided by AI. These policies act as a form of practice and proficiency audit, ensuring that if the AI systems fail or produce dubious outputs, the human expert is prepared to intervene competently. Importantly, aviation also demonstrates that technology itself can be leveraged to counteract deskilling. Flight simulators provide structured opportunities to maintain proficiency in low-frequency, high-stakes tasks. Similarly, in healthcare domains such as surgery, where hands-on opportunities may be limited, simulation-based training leveraging VR technology has been recommended as a means to preserve technical competencies and prevent skill decay (de Andres Crespo et al. 2025).

Beyond technological solutions, future studies should also emphasise interventions focused on user training and novel modes of interaction with AI, such as *Human–AI Collaboration Protocols* which stipulate how practitioners should utilise AI tools for specific clinical tasks (Cabitza et al. 2021). Medical curricula must integrate comprehensive AI education, encompassing both practical usage and critical engagement with AI's capabilities and limitations. Continuous professional development programmes focusing on critical thinking and decision-making should accompany AI training. Specialised workshops on ethical AI use in clinical practice will enhance clinicians' ethical and professional autonomy and emphasis should be placed on developing complementary human-centric skills such as empathy, ethical judgment, and complex decision-making. These skills remain vital for patient-centered care, preserving, and enhancing the irreplaceable human elements of medical practice.

Moreover, high-expertise fields can learn governance strategies from each other, with each field implementing proactive measures rather than learning only from its own mistakes. Potential insights for healthcare to could be found in lessons learned from law (Carrel 2018; Ruan 2020; Browning 2024), management (Amer et al. 2024), finance and accounting (Odonkor et al. 2024), the military (Vallor 2013), and aviation (Golfetti et al. 2021; Talib et al. 2025). This could inspire new strategies, interaction models, and governance principles that are transferable to healthcare settings.

Ultimately, the integration of AI into healthcare represents a balancing act between *efficiency* and *expertise*, *automation* and *autonomy*, *technological precision* and *human intuition*. While AI offers unprecedented opportunities to enhance medical practice, future research must be guided by a commitment to describing, promoting and preserving the irreplaceable elements of human clinical judgment, patient communication, and ethical reasoning. The risks of deskilling are not an inevitable consequence of AI implementation but rather a challenge that must be anticipated and managed through responsible governance, strategic training initiatives, and evidence-based policy interventions. This represents a defining priority, and a role to play, for future research in human–AI interaction in the medical field.

5 Limitations

While our review provides a substantive analysis of the current state of AI in healthcare, it is not without its limitations. First, concerning the use of the PACES framework, we recognize that this tool is primarily designed to evaluate skills acquired by medical students in structured, controlled learning environments. As such, it may not fully capture the complexity and nuance of clinical competencies in real-world settings. We adopted this framework because it offers one of the closest measurable proxies for clinical skill assessment currently available. However, future research would benefit from empirical studies that examine the dynamics of deskilling, skill inhibition, reskilling, and upskilling within the context of medical AI. Such investigations would provide valuable insights and further enrich the discourse in regards to these phenomena.

Second, the extensive and time-consuming effort of producing literature reviews is generally subject to being outpaced by the concurrent and rapid evolution of AI technologies and their applications in healthcare organizational processes. Additionally, the phenomenon of publication bias—wherein the literature may disproportionately reflect successful AI implementations over those that are less successful, or that yield negative outcomes—presents a challenge to obtaining a balanced view of AI's efficacy and risks. This bias can lead to an overly optimistic portrayal of AI in healthcare, potentially overlooking the critical lessons that could be drawn from less favorable results. To address these limitations, we proposed in Sect. 4 an agenda for future research on the theme of deskilling in medicine, encompassing the conceptualization and differentiation of the concepts of deskilling and upskilling inhibition; real-time monitoring of AI applications in healthcare; emphasis on longitudinal studies; wider adoption of qualitative methodologies and mixed-methods research.

6 Conclusions

Critical voices have cautioned how the repeated delegation of decision-making to machines may erode human expertise, to the point that certain skills may even be irreversibly lost in a "second singularity" Klein (2022), unless proactive measures are taken. In healthcare, this manifests as a growing dependence on AI-driven recommendations, progressively narrowing clinicians' opportunities to refine their judgment through real-world experience. Without deliberate interventions, the medical profession risks entering a phase where AI's authority is unquestioned, and human oversight becomes increasingly performative rather than substantive.

Compounding this concern is the phenomenon of *epistemic sclerosis* (Cabitza 2021), a form of knowledge crystallization driven by AI's reliance on historical data and past interpretations. As AI systems continuously reinforce pre-existing diagnostic patterns, they risk solidifying medical knowledge rather than expanding it, limiting the ability of clinicians to challenge prevailing norms or recognize novel cases. This epistemic rigidity could be particularly harmful in complex, ambiguous, or low-data scenarios, where true innovation and clinical breakthroughs rely on the ability to critically reinterpret existing knowledge rather than passively accept algorithmic outputs.

However, it is also important to recognise a dialectic tension: rather than simply disappearing, some medical competencies may evolve and shift in nature as AI is introduced. For example, as more routine diagnostic tasks become automated, clinicians might assume greater responsibility for oversight-validating AI-generated recommendations and critically appraising algorithmic outputs and, even more crucially, cultivating those skills that escape AI (Aslam and Hoyle 2022), such as holistically considering disparate and *undatafiable* (Anichini et al. 2024) contextual information for the cultivation of humane and empathetic patient management strategies. In this view, expertise is not unilaterally eroded by AI but can evolve, with meta-decision skills of supervision and interpretation gaining prominence alongside traditional clinical acumen.

Our analysis highlighted the profound challenges and significant opportunities that AI presents within the clinical workflow and organizational processes, effectively posing itself as a double-edged sword: a potent and compelling tool for enhancing clinical efficiency and decision-making, while simultaneously posing the risk of eroding the skills supporting the indispensable human elements of healthcare.

A central contribution of this review is its structured categorization of AI-induced deskilling, framed within the PACES-MRCPUK framework (Elder et al. 2011; Elder 2018). By systematically evaluating concerns across key clinical skill domains-including physical examination, differential diagnosis, clinical judgment, and patient communication-our findings reveal the breadth of AI's unintended consequences. The structured approach highlights specific vulnerabilities, such as the deterioration of physical examination skills, the weakening of diagnostic reasoning, and the erosion of physician-patient communication, reinforcing the urgency of implementing strategies that preserve critical expertise. Among such strategies, a promising avenue is to leverage AI itself as a training tool to counteract deskilling. Adaptive AI tutors or decision-making simulators could provide clinicians with guided practice and feedback, helping them maintain and even enhance their competencies in parallel with AI integration. In this way, the very technologies that pose a risk to skills can also serve to foster continuous learning and the resilience of medical expertise.

Furthermore, AI-induced deskilling is socio-technical (Rafner et al. 2022). It does not occur in isolation, but interacts with organizational and systemic factors, affecting team-based decision-making, interprofessional collaboration, and institutional knowledge retention. The restructuring of clinical roles to accommodate AI tools can create unforeseen challenges, such as fragmented workflows, reduced team situational awareness, and over-specialization in AI validation rather than direct diagnostic engagement and skills. Left unaddressed, these shifts could weaken professional autonomy and entrench new dependencies that fundamentally reshape medical expertise.

Our work holds significant implications for the interface between technology and human behavior within healthcare settings, with implications extending to other high-stakes, high-expertise fields. The integration of technology into professional practice is a journey of adaptation, potentially marked by significant gains in capability and efficiency. Yet, it simultaneously invokes reflection on what it means to possess a skill, to exercise professional judgment, and to maintain autonomy in an age of algorithmic influence on the decision-making process. The decision to responsibly embrace technology, while recognizing the potential for skill loss, requires respecting and upholding the enduring significance of human expertise even at a time of technological hype, aligning AI's trajectory with the ethos of patient-centred care.

Acknowledgements Chiara Natali gratefully acknowledges the financial support provided by the Federal Commission for Scholarships for Foreign Students in the form of the Swiss Government Excellence Scholarship (ESKAS No. 2024.0002) for the academic year 2024–2025. Federico Cabitza acknowledges funding support provided by the Italian project PRIN PNRR 2022 InXAID—Interaction with eXplainable Artificial Intelligence in (medical) Decision making. CUP: H53D23008090001 funded by the European Union—Next Generation EU.

Author contributions All authors contributed to drafting, revising, and finalizing the manuscript. CN and LM co-led project coordination. CN led conceptual development, synthesising insights from the literature review to frame the study's broader implications, ensure coherence across sections, and establish connections to Human–AI Interaction. LM led the methodological development of the review process and contributed additional insights to the Research Agenda. LDDD executed the literature review methodology, collecting and synthesising data in the manuscript, and co-authored the Ethical Implications section with CN. CN and LDDD prepared figures and tables (CN: Figures 3–4, Table 3; LDDD: Figures 1–2, Table 1, CN&LDDD: Tables 2–3–4). FC conceived the original idea and provided strategic guidance, supervision, and editorial oversight throughout the research process.

Funding Chiara Natali: Swiss Government Excellence Scholarship (ESKAS No. 2024.0002). Federico Cabitza: PRIN PNRR 2022 InXAID project, Interaction with eXplainable Artificial Intelligence in (medical) Decision making, CUP: H53D23008090001, funded by the European Union—Next Generation EU.

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

Declarations

Conflict of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

Akudjedu TN, Torre S, Khine R, Katsifarakis D, Newman D, Malamateniou C (2023) Knowledge, perceptions, and expectations of artificial intelligence in radiography practice: a global radiography workforce survey. J Med Imaging Radiat Sci 54(1):104–116

Amer M, Hilmi Y, El Kezazy H (2024) Big data and artificial intelligence at the heart of management control: towards an era of renewed strategic steering. In: The international workshop on big data and business intelligence. Springer, pp 303–316

Anichini G, Natali C, Cabitza F (2024) Invisible to machines: designing ai that supports vision work in radiology. In: Computer supported cooperative work (CSCW), pp 1–44

Aquino YSJ, Rogers WA, Braunack-Mayer A, Frazer H, Win KT, Houssami N, Degeling C, Semsarian C, Carter SM (2023) Utopia versus dystopia: professional perspectives on the impact of healthcare artificial intelligence on clinical roles and skills. Int J Med Inf 169:104903

Aslam TM, Hoyle DC (2022) Translating the machine: skills that human clinicians must develop in the era of artificial intelligence. Ophthalmol Ther 11(1):69–80

Bainbridge L (1983) Ironies of automation. In: Analysis, design and evaluation of man–machine systems. Elsevier, pp 129–135

Banerjee M, Chiew D, Patel KT, Johns I, Chappell D, Linton N, Zaman S (2021) The impact of artificial intelligence on clinical education: perceptions of postgraduate trainee doctors in London (UK) and recommendations for trainers. BMC Med Educ 21(1):1–10

Beane M (2019) Learning to work with intelligent machines. Harv Bus Rev 97(5):140–148

Beauchamp TL, Childress JF (1994) Principles of biomedical ethics. Oxford University Press, New York

Braverman H (1974) Labor and monopoly capital: the degradation of work in the twentieth century. Monthly Review Press, New York

Browning JG (2024) No "robot lawyers" just yet: the role of continuing legal education in fulfilling the duty of technological competence. J Leg Educ 72(3):11

Bunch J, Jones D, Psirides A (2023) Are we deskilling or reskilling our hospital ward clinicians? Intern Med J 53(4):640–643

Cabitza F (2017) Breeding electric zebras in the fields of medicine. CoRR. https://doi.org/10.48550/arXiv.1701.04077

Cabitza F (2021) Cobra AI: Explor e some unintended consequences. Perspectives on dependable AI, machines we trust. MIT, Cambridge, p 87

Cabitza F, Campagner A, Sconfienza LM (2021) Studying human–AI collaboration protocols: the case of the Kasparov's law in radiological double reading. Health Inf Sci Syst 9:1–20

Cabitza F, Campagner A, Ronzio L, Cameli M, Mandoli GE, Pastore MC, Sconfienza LM, Folgado D, Barandas M, Gamboa H (2023) Rams, hounds and white boxes: investigating human–AI collaboration protocols in medical diagnosis. Artif Intell Med 138:102506

Cabitza F, Natali C, Famiglini L, Campagner A, Caccavella V, Gallazzi E (2024) Never tell me the odds: investigating pro-hoc explanations in medical decision making. Artif Intell Med 150:102819

Cabitza F, Famiglini L, Fregosi C, Pe S, Parimbelli E, La Maida GA, Gallazzi E (2025) From oracular to judicial: enhancing clinical decision making through contrasting explanations and a novel interaction protocol. In: Proceedings of the 30th international conference on intelligent user interfaces, pp 745–754

Cabitza F, Rasoini R, Gensini GF (2017) Unintended consequences of machine learning in medicine. JAMA 318(6):517–518

Campagner A, Cabitza F, Ciucci D (2019) Three-way classification: ambiguity and abstention in machine learning. In: Rough sets: international joint conference, IJCRS 2019, Debrecen, Hungary, 17–21 June 2019, proceedings. Springer, pp 280–294

Campbell CG, Ting DS, Keane PA, Foster PJ (2020) The potential application of artificial intelligence for diagnosis and management of glaucoma in adults. Br Med Bull 134(1):21–33

Carrel A (2018) Legal intelligence through artificial intelligence requires emotional intelligence: a new competency model for the 21st century legal professional. Ga St UL Rev 35:1153

Carter SM, Rogers W, Win KT, Frazer H, Richards B, Houssami N (2020) The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. Breast 49:25–32

Chen Y, Stavropoulou C, Narasinkan R, Baker A, Scarbrough H (2021) Professionals' responses to the introduction of AI innovations in radiology and their implications for future adoption: a qualitative study. BMC Health Serv Res 21:1–9

Choudhury A, Chaudhry Z (2024) Large language models and user trust: consequence of self-referential learning loop and the deskilling of health care professionals. J Med Internet Res 26:e56764

Christopher JK, Karthikesalingam A, Suleyman M, Corrado G, King D (2019) Key challenges for delivering clinical impact with artificial intelligence. BMC Med 17:1–9

Da Silva M, Horsley T, Singh D, Da Silva E, Ly V, Thomas B, Daniel RC, Chagal-Feferkorn KA, Iantomasi S, White K et al (2022) Legal concerns in health-related artificial intelligence: a scoping review protocol. Syst Rev 11(1):1–8

de Andres Crespo M, Lykoudis PM, Myint F, Berlingieri P (2025) Surgery and technical skill decay. Int J Surg 111(5):3399–3413

Dehais F, Peysakhovich V, Scannella S, Fongue J, Gateau T (2015) "automation surprise" in aviation: real-time solutions. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems, pp 2525–2534

Dellermann D, Ebel P, Söllner M, Leimeister JM (2019) Hybrid intelligence. Bus Inf Syst Eng 61(5):637–643

Dickersin K, Chan S, Chalmersx T, Sacks H, Smith H Jr (1987) Publication bias and clinical trials. Control Clin Trials 8(4):343–353

Drabiak K, Kyzer S, Nemov V, El Naqa I (2023) Ai and machine learning ethics, law, diversity, and global impact. Br J Radiol 96:20220934

Duran LDD (2021) Deskilling of medical professionals: an unintended consequence of ai implementation? Giornale di filosofia 2(2):47–59
Duran H-T, Kingeter M, Reale C, Weinger MB, Salwei ME (2023) Decision-making in anesthesiology: will artificial intelligence make intraoperative care safer? Curr Opin Anesthesiol 36(6):691–697
Elder A (2018) Clinical skills assessment in the twenty-first century. Med Clin 102(3):545–558
Elder A, McManus C, McAlpine L, Dacre J (2011) What skills are tested in the new paces examination? Ann Acad Med Singapore 40(3):119
Elish MC (2019) Moral crumple zones: Cautionary tales in human-robot interaction. In *Engaging Science,Technology, and Society*
Evjemo T, Johnsen S (2019) Lessons learned from increased automation in aviation: the paradox related to the high degree of safety and implications for future research. In: 29th European safety and reliability conference
Gerke S, Minssen T, Cohen G (2020) Ethical and legal challenges of artificial intelligence-driven healthcare. Artif Intell Healthc 26:295–336
Golfetti A, Napoletano L, Cichomska K (2021) A framework to understand current and future competences and occupations in the aviation sector. In: Transformation of transportation. Springer, Cham, pp 213–226
Green BP (2019) Artificial intelligence, decision-making, and moral deskilling. Markkula Center for Applied Ethics website
Hallowell N, Badger S, McKay F, Kerasidou A, Nellåker C (2023) Democratising or disrupting diagnosis? Ethical issues raised by the use of AI tools for rare disease diagnosis. SSM Qual Res Health 3:100240
Hoff T (2011) Deskilling and adaptation among primary care physicians using two work innovations. Health Care Manage Rev 36(4):338–348
Iqbal J, Jahangir K, Mashkoor Y, Sultana N, Mehmood D, Ashraf M, Hafeez MH (2022) The future of artificial intelligence in neurosurgery: a narrative review. Surg Neurol Int 13:536
Kaas MH (2024) The perfect technological storm: artificial intelligence and moral complacency. Ethics Inf Technol 26(3):49
Kapoor R, Walters SP, Al-Aswad LA (2019) The current state of artificial intelligence in ophthalmology. Surv Ophthalmol 64(2):233–240
Kashou AH, Noseworthy PA, Anavekar NS, Rowlandson I, May AM (2024) Bridging ecg learning with emerging technologies: advancing clinical excellence. J Electrocardiol 86:153765
Kayaduvar M, Ünal C (2023) Decision-making processes in an artificially intelligent healthcare sector: can algorithms beat the physicians? In: Management in the digital era: different perspectives. Nova Publishers, Hauppauge
Kim TW, Scheller-Wolf A (2022) Technological unemployment, meaning in life, purpose of business, and the future of stakeholders. In: Business and the ethical implications of technology. Springer, pp 13–31
Kleim JA, Jones TA (2008) Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. J Speech Lang Hear Res51(1):S225–S239
Klein GA (2022) Snapshots of the mind. MIT, Cambridge
Koplin JJ, Johnston M, Webb AN, Whittaker A, Mills C (2025) Ethics of artificial intelligence in embryo assessment: mapping the terrain. Hum Reprod 40(2):179–185
Kundu S (2021) How will artificial intelligence change medical training? Commun Med 1(1):8
Lennartz S, Dratsch T, Zopfs D, Persigehl T, Maintz D, Große Hokamp N, Pinto dos Santos D (2021) Use and control of artificial intelligence in patients across the medical workflow: single-center questionnaire study of patient perspectives. J Med Internet Res 23(2):e24221
Levy J, Jotkowitz A, Chowers I (2019) Deskilling in ophthalmology is the inevitable controllable? Eye 33(3):347–348
Lu J (2016) Will medical technology deskill doctors? Int Educ Stud 9(7):130–134
Maassen O, Fritsch S, Palm J, Deffge S, Kunze J, Marx G, Riedel M, Schuppert A, Bickenbach J (2021) Future medical artificial intelligence application requirements and expectations of physicians in German university hospitals: web-based survey. J Med Internet Res 23(3):e26646
Malik P, Pathania M, Rathaur VK et al (2019) Overview of artificial intelligence in medicine. J Fam Med Primary Care 8(7):2328
Michels ME, Evans DE, Blok GA (2012) What is a clinical skill? Searching for order in chaos through a modified Delphi process. Med Teach 34(8):e573–e581
Miller T (2023) Explainable ai is dead, long live explainable ai! hypothesis-driven decision support using evaluative AI. In: Proceedings of the 2023 ACM conference on fairness, accountability, and transparency, pp 333–342
Mofatteh M (2021) Neurosurgery and artificial intelligence. AIMS Neurosci 8(4):477
Monteith S, Glenn T, Geddes J, Whybrow PC, Achtyes E, Bauer M (2022) Expectations for artificial intelligence (AI) in psychiatry. Curr Psychiatry Rep 24(11):709–721

Mooty WL (2022) Advisory circular: flightpath management. Technical report. Department of transportation, Federal Aviation Administration

Morandini S, Fraboni F, De Angelis M, Puzzo G, Giusino D, Pietrantoni L et al (2023) The impact of artificial intelligence on workers' skills: upskilling and reskilling in organisations. Inf Sci Int J Emerg Transdiscipline 26:39–68

Morley J, Machado CC, Burr C, Cowls J, Joshi I, Taddeo M, Floridi L (2020) The ethics of AI in health care: a mapping review. Soc Sci Med 260:113172

Mosch L, Fürstenau D, Brandt J, Wagnitz J, Klopfenstein SA, Poncette AS, Balzer F (2022) The medical profession transformed by artificial intelligence: qualitative study. Digital Health 8:20552076221143904

Nakagawa K, Moukheiber L, Celi LA, Patel M, Mahmood F, Gondim D, Hogarth M, Levenson R (2023) AI in pathology: what could possibly go wrong? In: Seminars in diagnostic pathology, vol 40. Elsevier, pp 100–108

Natali C, Campagner A, Cabitza F (2024) Answering the call to go beyond accuracy: an online tool for the multidimensional assessment of decision support systems. Biostec 2:219–229

Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, Pournamdari AB, Tkachenko E, Barbieri JS, Ko JM et al (2020) Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. JAMA Dermatol 156(5):501–512

Obermeyer Z, Lee TH (2017) Lost in thought: the limits of the human mind and the future of medicine. N Engl J Med 377(13):1209

Odonkor B, Kaggwa S, Uwaoma PU, Hassan AO, Farayola OA (2024) The impact of AI on accounting practices: a review: exploring how artificial intelligence is transforming traditional accounting methods and financial reporting. World J Adv Res Rev 21(1):172–188

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE et al (2021) The prisma 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg 88:105906

Panesar SS, Kliot M, Parrish R, Fernandez-Miranda J, Cagle Y, Britz GW (2020) Promises and perils of artificial intelligence in neurosurgery. Neurosurgery 87(1):33–44

Parasuraman R, Sheridan TB, Wickens CD (2000) A model for types and levels of human interaction with automation. IEEE Trans Syst Man Cybernet Part A Syst Humans 30(3):286–297

Parchmann N, Hansen D, Orzechowski M, Steger F (2024) An ethical assessment of professional opinions on concerns, chances, and limitations of the implementation of an artificial intelligence-based technology into the geriatric patient treatment and continuity of care. GeroScience 46(6):6269–6282

Rafner J, Dellermann D, Hjorth A, Veraszto D, Kampf C, MacKay W, Sherson J (2022) Deskilling, upskilling, and reskilling: a case for hybrid intelligence. Morals Mach 1(2):24–39

Rao D (2023) The urgent need for healthcare workforce upskilling and ethical considerations in the era of ai-assisted medicine. Indian J Otolaryngol Head Neck Surg 75(3):2638–2639

Reingold O, Shen JH, Talati A (2024) Dissenting explanations: leveraging disagreement to reduce model overreliance. Proc AAAI Conf Artific Intell 38:21537–21544

Rinta-Kahila T, Penttinen E, Salovaara A, Soliman W, Ruissalo J (2023) The vicious circles of skill erosion: a case study of cognitive automation. J Assoc Inf Syst 24(5):1378–1412

Ruan N (2020) Attorney competence in the algorithm age. ABAJ Lab Emp L 35:317

Ruskin KJ, Corvin C, Rice SC, Winter SR (2020) Autopilots in the operating room: safe use of automated medical technology. Anesthesiology 133(3):653–665

Sambasivan N, Veeraraghavan R (2022) The deskilling of domain expertise in ai development. In: Proceedings of the 2022 CHI conference on human factors in computing systems, pp 1–14

Santoni de Sio F, Van den Hoven J (2018) Meaningful human control over autonomous systems: a philosophical account. Front Robot AI 5:323836

Sarkar A (2024) Ai should challenge, not obey. Commun ACM 67(10):18–21

Schemmer M, Kühl N, Satzger G (2021) Intelligent decision assistance versus automated decision-making: Enhancing knowledge work through explainable artificial intelligence. arXiv preprint. arXiv:2109.13827:1–10

Simpkin AL, Vyas JM, Armstrong KA (2017) Diagnostic reasoning: an endangered competency in internal medicine training. Ann Intern Med167(7):507–508

Smith PJ, Baumann E (2020) Human-automation teaming: unintended consequences of automation on user performance. In: 2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC), pp 1–9

Sparrow R, Hatherley JJ (2019) The promise and perils of AI in medicine. Int J Chin Compar Philos Med 17(2):79–109

Stogiannos N, O'Regan T, Scurr E, Litosseliti L, Pogose M, Harvey H, Kumar A, Malik R, Barnes A, McEntee MF et al (2025) Lessons on AI implementation from senior clinical practitioners: an exploratory qualitative study in medical imaging and radiotherapy in the uk. J Med Imaging Radiat Sci 56(1):101797

Talib MA, Nasir Q, Dakalbab F, Saud H (2025) Future aviation jobs: the role of technology in shaping skills and competencies. J Open Innov Technol Market Complex 2:100517
Tsai TL, Fridsma DB, Gatti G (2003) Computer decision support as a source of interpretation error: the case of electrocardiograms. J Am Med Inform Assoc 10(5):478–483
Vallor S (2013) The future of military virtue: Autonomous systems and the moral deskilling of the military. In: 2013 5th International Conference on Cyber Conflict (CYCON 2013). IEEE, pp 1–15
Vallor S (2015) Moral deskilling and upskilling in a new machine age: reflections on the ambiguous future of character. Philos Technol 28:107–124
Verghese A, Shah NH, Harrington RA (2018) What this computer needs is a physician: humanism and artificial intelligence. JAMA 319(1):19–20
Weidener L, Fischer M et al (2024) Role of ethics in developing AI-based applications in medicine: insights from expert interviews and discussion of implications. Jmir AI 3(1):e51204
Wessel N-C (2023) Decision-support systems and decision making: managing decisional deskilling in human–DSS interactions in organizations. In: ICDS 2023: The seventeenth international conference on digital society
Winter P, Carusi A (2022) Professional expectations and patient expectations concerning the development of artificial intelligence (AI) for the early diagnosis of pulmonary hypertension (PH). J Respons Technol 12:100052
Winter PD, Carusi A (2022b) (De)troubling transparency: artificial intelligence (AI) for clinical applications. Med Humanit 49(1):17–26
Woodruff A, Shelby R, Kelley PG, Rousso-Schindler S, Smith-Loud J, Wilcox L (2024) How knowledge workers think generative ai will (not) transform their industries. In: Proceedings of the 2024 CHI conference on human factors in computing systems, CHI '24. Association for Computing Machinery, New York
Zhang W, Cai M, Lee HJ, Evans R, Zhu C, Ming C (2023) AI in medical education: global situation, effects and challenges. Educ Inf Technol 29(4):4611–4633
Zulkipli IN, Alam F, Lim M-A (2023) Integrating AI in medical education: embracing ethical usage and critical understanding. Front Med 10:1279707

Authors and Affiliations

Chiara Natali[1,2] · Luca Marconi[1] · Leslye Denisse Dias Duran[3] · Federico Cabitza[1,4]

✉ Chiara Natali
chiara.natali@unimib.it

✉ Leslye Denisse Dias Duran
leslye.diasduran@ruhr-uni-bochum.de

Luca Marconi
luca.marconi@unimib.it

Federico Cabitza
federico.cabitza@unimib.it

1 Department of Informatics, Systems and Communication, University of Milano-Bicocca, Via Sarca 336, 20126 Milan, Italy

2 University of Applied Sciences and Arts of Southern Switzerland, Via La Santa 1, 6900 Lugano, Switzerland

3 Institute of Philosophy I, Ruhr University Bochum, Universitätsstraße 150, 44801 Bochum, North Rhine-Westphalia, Germany

4 IRCCS Ospedale Galeazzi - Sant'Ambrogio, Via Cristina Belgioioso 173, 20157 Milan, Italy