Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Physicians Assign Themselves More Responsibility Than the General Population when AI Fails in Healthcare: A Swedish Cross-Sectional Survey

Daniel Andersson

daniel.andersson@gu.se

University of Gothenburg

Sofia Thunberg

Uppsala University

Rasmus Arvidsson

University of Gothenburg

Jonathan Widén

Region Västra Götaland

Ronny Gunnarsson

University of Gothenburg

Carl Wikberg

University of Gothenburg

Charlotte Blease

Uppsala University

Anna Moberg

Linköping University

Peter Nymberg

Malmö, Lund University

Pär-Daniel Sundvall

University of Gothenburg

David Sundemo

University of Gothenburg

Research Article

Keywords:

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11280050/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Physicians Assign Themselves More Responsibility Than the General Population when AI Fails in Healthcare: A Swedish Cross-Sectional Survey

Daniel Andersson[1*], Sofia Thunberg[2], Rasmus Arvidsson[1], Jonathan Widén[3], Ronny Gunnarsson[1,3,4], Carl Wikberg[1], Charlotte Blease[5,6], Anna Moberg[7,8], Peter Nymberg[9,10], Pär-Daniel Sundvall[1], David Sundemo[1,11]

[1*]School of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.
[2]Department of Information Technology, Uppsala University, Regementsvägen 10, Uppsala, 751 05, Sweden.
[3]Research, Education, Development and Innovation, Primary Health Care, Västra Götaland County, Gothenburg, Sweden.
[4]College of Medicine and Dentistry, James Cook University, Cairns, Australia.
[5]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden.
[6]Digital Psychiatry, Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, USA.
[7]Department of Health, Medicine, and Caring Sciences, Linköping University, Linköping, 581 83, Sweden.
[8]Primary Healthcare Center Kärna, Region Östergötland, Kärnabrunnsgatan 10, Linköping, 586 62, Sweden.
[9]Center for Primary Health Care Research, Department of Clinical Sciences, Malmö, Lund University, Jan Waldenströms gata 35, Malmö, 205 02, Sweden.
[10]University Clinic Primary Care, Skåne University Hospital, Region Skåne, Travvägen 27, Helsingborg, 254 83, Sweden.
[11]Center for Digital Health, Sahlgrenska University Hospital, Region Västra Götaland, Gothenburg, Sweden.

*Corresponding author(s). E-mail(s): daniel.andersson@gu.se;
Contributing authors: sofia.thunberg@it.uu.se; rasmus.arvidsson@gu.se;
jonathan.loke.widen@gmail.com; ronny.gunnarsson@gu.se;
carl.wikberg@allmed.gu.se; charlotte.blease@uu.se; anna.moberg@liu.se;
peter.nymberg@med.lu.se; par-daniel.sundvall@gu.se;
david.sundemo@gu.se;

Abstract

Background: The integration of Artificial Intelligence (AI) into clinical decision-making creates uncertainty regarding responsibility and accountability when adverse outcomes occur. Despite this, data regarding stakeholder perspectives remain scarce, especially quantitative data comparing physician and public views on responsibility allocation and transparency. This study aims to evaluate and compare how physicians and the general population in Sweden attribute responsibility for negative medical outcomes when AI is used to support healthcare decisions in primary care, as well as their views on the necessity of transparently disclosing AI use to patients.

Methods: A nationwide, cross-sectional, mixed-methods survey was conducted in Sweden between February and May 2025. The survey targeted two distinct cohorts (N = 500 per group): licensed physicians under 70 years, sampled from the National Board of Health and Welfare register, and adults aged 18 or older sampled from the Swedish population registry. The survey used case vignettes as background for multiple-choice questions, with complementary open-ended free-text responses. The data was evaluated using comparative statistical analysis and reflexive thematic analysis.

Results: Physicians were significantly more likely than the general population to assign responsibility to the physician (92% vs. 74%) and to favor a shared responsibility model involving the physician (48% vs. 35%). In contrast, the general population more frequently selected non-physician actor combinations (26% vs. 8%) and assigned greater responsibility to the patient in a scenario involving direct patient-AI interaction without a physician involved. Neither group broadly assigned responsibility to the AI itself (1.3% physicians vs. 10% general population). The reflexive thematic analysis reinforced these results, identifying a shared accountability framework where all actors share responsibility, but where the physician serves as the ultimate safeguard. A majority of the general population (68%) affirmed the patient's right to always know if AI was used, in contrast to only 23% of physicians who exhibited a polarised distribution.

Conclusion: Respondents endorse a shared responsibility model among actors involved in implementing and using AI for medical decisions. The physician is viewed as the essential human-in-the-loop, bearing ultimate accountability for adverse outcomes of AI-assisted medical decisions according to a vast majority of respondents. Divergent perspectives regarding transparency emphasise the need to clarify informed consent standards for AI deployment in healthcare.

Keywords: artificial intelligence, large language models, generative ai, primary care, responsibility, accountability, transparency

1 Introduction

Expectations regarding the potential of Artificial Intelligence (AI) in healthcare have reached unprecedented levels, with leading experts suggesting that AI may alleviate or even replace a substantial portion of the cognitive work currently performed by clinicians [1, 2]. In a previous report on UK General Practitioners (GP)s' perspectives on AI in primary care, 28% of respondents reported using AI tools in clinical practice [3]. More recent figures suggest continued adoption, with 52% of GPs indicating use of Generative AI in clinical practice [4].

AI systems could transform healthcare at many levels and solve many of its current challenges [5, 6], for instance by reducing costs and workloads [7], improving clinical efficiency and quality [8], and enabling earlier and more accurate diagnoses [9]. However, moving these tools into real-world clinics has proven difficult [10]. A recent systematic review identified six categories of barriers to adoption of AI in healthcare, including ethical, technological, liability and regulatory, workforce, social and patient safety concerns [11]. Among these, the issue of responsibility was identified as one of the most important subthemes.

With the advent of intelligent computer systems that generate or influence decisions for complex health-related tasks, a new type of agent disrupts a clinical decision-making process that has traditionally been grounded in the clinician-patient relationship. This shift has implications for the distribution of responsibility. A survey of 2,108 UK GPs identified professional liability and medico-legal issues as primary concerns, cited by 80% of current AI users and 89% of non-users. Liability concerns were also cited as the top influencing factor regarding the adoption of AI tools in their practice [3]. This sentiment is corroborated by another recent qualitative study, where GPs reported feeling personally responsible for AI-generated content even when tools fail, and that this lack of clear accountability is a major source of professional anxiety [12].

Despite increasing deployment of AI systems in healthcare, their widespread adoption is thus potentially hindered by this unresolved question: who is responsible when AI fails? This study aims to examine this question within the context of the Swedish healthcare system.

2 Background

2.1 The Swedish Healthcare System

Swedish healthcare is a decentralised system funded primarily through general taxation, with regional governments managing resource allocation, service organisation, and quality assurance [13]. Legislation requires healthcare providers to ensure necessary competencies and implement control systems to protect patients from adverse events. Operational managers are formally responsible for implementing routines, while each licensed physician bears independent professional responsibility for their clinical actions.

Healthcare providers must inform patients and provide access to medical records to facilitate informed decision-making [14, 15, 16]. Documentation is predominantly

digital, and patients can access most of their health records through the healthcare system's dedicated website 1177.se.

The Swedish Government has long expressed the ambition to harness the opportunities of digital transformation [17], and the latest digitalisation strategy explicitly encourages AI development and application to improve healthcare [18].

2.2 Responsibility

Responsibility is a complex concept, with overlapping terms such as accountability, blameworthiness, liability, role-responsibility, and causality [19]. This study focuses on retrospective responsibility, or accountability. Potential bearers of responsibility when AI systems are involved in healthcare include physicians using AI, developers who created the AI system, vendors supplying it, healthcare organisations adopting it, and regulators authorising its use [20, 21].

2.2.1 The Preconditions of Responsibility

According to the current philosophical debate, responsibility requires three main preconditions: (1) causality, a causal connection between the agent's action and the outcome; (2) knowledge, the ability to understand potential consequences; and (3) freedom, the capacity to act without coercion [22, 19]. If a person either had no control over the flow of events, could not understand the results of her actions, or was violently forced to take a certain action, we normally do not consider such a person responsible.

Historically, these preconditions provided a framework for clearly linking an agent, its action, and the outcome. However, in the current sociotechnical society with complex organisational structures and advanced technology, attributing responsibility has become increasingly difficult [23, 24].

One widely discussed complication to the causality precondition is the so-called 'problem of many hands' [25], describing the difficulty of attributing responsibility when many agents are involved. Regarding the knowledge condition, the opaqueness of neural network-based AI systems poses a problem to the epistemic position, given the difficulty of understanding the inner workings of such systems. Regarding the condition of freedom, some argue that it depends on the metaphysical existence of free will. However, for the purpose of this practical analysis of responsibility, we adopt Strawson's approach and focus on observable actions and their consequences, bracketing the underlying metaphysical question [26]. If one adopts the common sense view that free will exists, autonomy is often seen as a matter of degree, with children, for example, regarded as less autonomous than adults.

2.2.2 AI-systems as Moral Agents

Until recently, responsibility has been uniquely attributed to moral agents, and moral agenthood has generally been reserved for humans [27, 19]. Some argue that sufficiently advanced computer systems could be held morally responsible, if they exhibit intentionality and the capacity to reason [28, 29]. Others dismiss this notion, citing insufficient intentionality and a lack of mental states, emotions, and empathy [30, 31].

An alternative model suggests AI-agents may be held accountable without the attribution of moral responsibility [32]. The originators use a dog analogy: if a dog misbehaves, we do not hold the animal morally responsible, but accountable in the sense that it may be punished, retrained, or taken out of service. This model has, however, been criticised, since it tends to focus on the technology, distracting from the human agents developing or using the system, and obscuring responsibility attribution [30].

2.2.3 The Responsibility Gap

A human using a simple tool, such as a hammer, is generally considered responsible for the handling of the tool. If the hammerhead falls off unexpectedly and causes injury, the manufacturer could instead be considered responsible. With autonomous and opaque AI-systems, the attribution of responsibility is less clear. According to Matthias [33], AI systems that optimise performance in real-time can make human oversight impossible, leading to outcomes that are difficult to control or predict. If we also presuppose that AI-systems cannot be moral agents, the question of responsibility remains unresolved, creating a *responsibility gap*.

Thus, responsibility gaps arise from the borderless interconnection between human agency and the agency-like properties of AI systems. It has been suggested that responsibility gaps can only be understood within the broader sociotechnical system in which the technology is embedded, a system including the end-user, the AI, and the broader context and organisation [34]. In this view, resolving responsibility gaps is crucial for the widespread adoption of AI in healthcare [35].

A recent qualitative study examined clinicians' views on the responsibility gap when AI is used in clinical care, and errors occur. While all respondents assigned responsibility to the physician, some acknowledged the existence of a responsibility gap, while others denied it or deemed it irrelevant [36].

2.2.4 Responsibility, trust and transparency

Trust is a fundamental element of the health care system, helping to ensure that individuals seek medical advice, adhere to treatment, and participate in research [37]. In the era of medical AI, the concept of trust is especially important, given the abstract and opaque nature of these systems [38]. A recent study demonstrated that the division of responsibility between AI and healthcare professionals has a fundamental impact on trust [39]. Transparency and informed consent are considered central factors to maintain trust during integration of AI into clinical care [40]. A survey of 2,000 UK patients revealed significant scepticism toward healthcare AI, with 47% expressing negative sentiments [3]. Alongside data indicating a strong desire among patients for disclosure regarding AI use in their clinical care [41], this underscores the importance of transparency to maintain trust.

2.3 Knowledge gaps and study purpose

Significant gaps exist regarding stakeholder perspectives on responsibility allocation and transparency when AI is used for medical decision-making. Specifically, there is

a lack of quantitative studies, and data comparing physician and public views remain limited.

The purpose of this study was to explore, and compare, the opinions of physicians and the general public regarding accountability and transparency, with the following primary research question (RQ):

RQ1 Do physicians and the general population differ significantly in attributing responsibility to the physician, when AI is used for medical decisions with negative outcomes?

Secondary aims were the following:

RQ2 How is responsibility distributed among remaining actors when the patient interacts directly with an AI agent, removing the physician from the decision process?

RQ3 Do physicians and the general population differ in their perception of a patient's right to know when AI is used by a physician?

RQ4 What is the general population's view on transparency when they interact with a chatbot, regarding the right to know if the answers come from a human or an AI?

RQ5 What are the thoughts and opinions of the participants about the distribution of responsibility when AI is acting in healthcare?

3 Method

3.1 Study Design, Population and Data Collection

A nationwide, cross-sectional survey was conducted in Sweden between February and May 2025. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist [42] was used to improve quality and transparency of reporting. Participants were divided into two groups: 1) physicians under 70 years with a valid social security number and a registered Swedish address ($n = 500$), sampled from the National Board of Health and Welfare register that includes all licensed practising physicians in Sweden; and 2) participants aged 18 years or older ($n = 500$), sampled from the Swedish population registry, covering approximately 99% of residents. Individuals were selected per group using simple random sampling without stratification.

Survey invitations with study information were distributed by postal mail. Respondents could respond by post or online. Two reminders were sent to non-respondents, the second offering a lottery-ticket incentive. Responses were anonymous to mitigate potential privacy concerns. A personal code was required for online completion, but was not linked to responses. No demographic data were collected directly; however, respondents could voluntarily register their participation. Demographic analyses were restricted to this registered subgroup.

3.2 Survey Design

During the early design phase, experts in digitalisation, clinical AI and medical ethics were consulted. Questionnaire items were piloted and refined through cognitive interviews with physicians and the general population to ensure clarity [43, 44]. Results of one section of the survey were previously reported [45] and covered questions on AI accuracy, use, and trust. The remaining section focused on accountability and transparency regarding AI use in medical decisions, the focus of this study.

An abbreviated and translated overview of the survey questions is provided in Figures 1 and 2. The questionnaires are also provided as supplementary files including English-translated versions. The questionnaire was largely identical for both groups, with minor wording adjustments to simplify the language for the general public to some extent. Regarding accountability, respondents were asked to identify the primarily responsible actor(s) in two fictitious cases. For the first case, an optional free-text justification was also requested. Regarding transparency, the first question assessed the extent to which respondents believed patients should be informed if a physician uses AI in medical decisions. A second transparency question, administered only to the general population group, asked the degree to which patients should be informed whether they chat with an AI or a real person.

Scenario Summary:
Region Management procures an AI diagnostic tool from the company "Healthcare-AI." Clinic management instructs the doctors to use it. Patient Peter, experiencing abdominal pain, consults with Dr. Maria who inputs symptoms and findings into the AI. The AI advises against emergency referral, recommending self-care advice instead. Dr. Maria thinks this seems reasonable and follows the AI's advice. Peter's condition later worsens; he is diagnosed at the Emergency room (ER) with appendicitis. The ER doctor estimates that the delay prolonged his hospital stay by four days.

Question 1:
The patient should have been referred to the ER immediately. Who do you think are **most** responsible for him going home instead?

- ☐ Regional Management
- ☐ Healthcare-AI (The AI provider)
- ☐ The AI System
- ☐ Clinic Management
- ☐ Dr. Maria (The Physician)
- ☐ Peter (The Patient)
- ☐ Other: _____
- ☐ No one

Adjusted Scenario:
Consider a similar scenario where the patient Peter instead interacts directly with the AI from home, via a chat interface, without any doctor involved. Peter is fully aware he is communicating with an AI. Based on the symptoms described by Peter, the AI advises him to stay home, with the same result and consequences.

Question 2:
The patient should have been advised to go to the ER immediately. Who do you think are **most** responsible in this case, for Peter not going directly to the ER?

- ☐ Regional Management
- ☐ Healthcare-AI (The AI provider)
- ☐ The AI System
- ☐ Clinic Management
- ☐ Peter (The Patient)
- ☐ Other: _____
- ☐ No one

Fig. 1 Survey questions on accountability (abbreviated).

Question 3:
Do you think doctors must inform their patients if they consult an AI? We are not referring to factual questions, but rather questions where the AI assists the doctor with diagnosis or management suggestions on real patient cases. Select one alternative:

- ☐ Yes, the patient always has the right to know.
- ☐ In most cases, the patient has the right to know, but sometimes it doesn't matter.
- ☐ In some cases, the patient has the right to know.
- ☐ In a few specific cases, the patient has the right to know, but usually, it doesn't matter
- ☐ No, it never matters.

Question 4:
Imagine that you contact healthcare services, and the person who answers could be either a nurse or an AI that is **just as good** as a nurse.
Do you think you should always be informed whether you are speaking with a nurse or an AI? Select one alternative:

- ☐ Yes, the patient always has the right to know.
- ☐ In most cases, the patient has the right to know, but sometimes it doesn't matter.
- ☐ In some cases, the patient has the right to know.
- ☐ In a few specific cases, the patient has the right to know, but usually, it doesn't matter
- ☐ No, it never matters.

Fig. 2 Survey questions on transparency (abbreviated).

3.3 Sample Size Calculation

Building on a dataset established in a previous study [45], sample size was determined by prior power calculations. Per STROBE recommendations, the primary consideration is whether the analysis will yield results with sufficient statistical precision; consequently, sample size considerations are informal [46]. Following Lakens [47], a sensitivity power analysis was performed using G*Power (version 3.1.9.7) [48] on the data for Question 1 (RQ1). With a total sample size of $N = 378$ ($n_1 = 155$, $n_2 = 223$), a two-tailed alpha level of $\alpha \leq .05$, a baseline proportion (p_1) of 0.74, and a power of 0.80, the study was sufficiently powered to detect a second proportion (p_2) of 0.85 or higher.

3.4 Analysis

We employed a mixed-methods approach, combining quantitative analysis of multiple-choice survey responses with reflexive Thematic Analysis (TA) [49] of accompanying free-text responses. Using survey data for reflexive TA is well established [50]. The mixed methods approach enabled a more nuanced understanding of participants' views on accountability when AI is used in medical decisions than quantitative methods alone.

3.4.1 Quantitative Analytical Methods

Statistical analysis was performed using R (version 4.5.3) [51]. Upset plots [52] were generated using the ComplexHeatmap R package [53], and Sankey diagrams using the NetworkD3 package [54]. Primary and secondary outcomes and associated analyses were pre-specified. Analyses for secondary outcomes should nevertheless be considered exploratory, and p-values for secondary outcomes should be interpreted with caution.

Multiple-response items were reported descriptively using absolute counts (n) and relative frequencies. For questions allowing multiple selected alternatives, relative frequencies may sum to more than 100%. Responses marked 'other' were redistributed based on the accompanying text, as detailed in Supplemental Appendix A.1.

To facilitate the analysis of the attribution of responsibility to physicians, responses to the first scenario were categorised into three subgroups: 'physician responsible', 'physician and others responsible', or 'only others responsible'. For the second scenario, where the physician was not an option, three similar subgroups were compiled around the patient as responsible.

Relative frequencies of the physician option in the first scenario, and the size of the subgroups in both scenarios, were compared between the two respondent groups using chi-square tests.

To illustrate how respondents shifted their selections between scenarios, respondents selecting the physician in the first scenario were tracked to their corresponding choices in the second scenario. This was visualised using two Sankey diagrams, serving as descriptive visualisations rather than bases for generalisable conclusions.

Ordinal variables were reported as percentages, medians and interquartile ranges (IQRs). The Mann-Whitney U test was used to assess significant differences between groups, and the Wilcoxon signed-rank test was used to compare differences between

question variations within the same group. Odds ratios (ORs) with 95% confidence intervals (CIs) were reported where applicable.

3.4.2 Reflexive Thematic Analysis

Aligning with Braun and Clarke's recommendation on methodology transparency [55], our theoretical stance is grounded in a critical realist paradigm. This ontology assumes an external reality exists independently of perception but is known imperfectly and probabilistically. Our stance aligns with post-positivism, acknowledging that while meaningful insights into participants' experiences are possible, all knowledge is approximative. Participants use language to express perspectives, which we interpret as meaningful approximations of their views.

We adopted an inductive, semantic approach to coding, developing themes bottom-up from the data. We recognise that our assumptions, backgrounds, and interpretive lenses shape meaning construction; thus, reflexivity is integral to the analytical process. Using Finlay's terminology, we primarily engaged in introspective reflexivity, where researchers' experiences guide interpretation [56]. Throughout the process, we strove to remain attentive to how our interpretations shaped theme development. Given reflexivity is influenced by personal factors, Mauthner and Doucet advocate for greater transparency regarding the assumptions embedded in research methods [57]. We address this through a positionality statement [58] (Supplemental Appendix A.2), and by adhering to the Reflexive TA Reporting Guidelines (RTARG) [55]. The resulting themes are treated not as objective truths, but as well-supported, interpretative insights into participants' perspectives.

To enhance understanding of the free-text responses, free-text answers were combined with preceding choices to form the text for analysis. After data familiarisation, two authors independently generated codes, using rich coding to preserve meaning. Code value was not associated with prevalence. The coding process involved two or more complete readings of the material by both authors. A refined list of codes was produced through a joint discussion and reflexive dialogue. Codes were arranged into preliminary themes, which were examined and grouped. These preliminary themes were mapped to original quotes and reviewed for coherence and distinction. A third author coded a subset of responses, verifying coherence and examining theme mapping and possible reassignment. The final set of three themes was subsequently discussed among the majority of the authors.

4 Results

The response rate was 45% for physicians and 31% for the general population. Demographic analysis suggested a representative sample, detailed in a previous publication based on another part of the same survey [45]. A table summarising the demographic characteristics of the respondents, reproduced from that study, is available in Supplemental Table A1.

One physician respondent and three from the general population did not answer the transparency questions, and were excluded from that analysis.

4.1 Responsibility

4.1.1 Question 1: The physician uses AI

In the scenario where a physician uses AI with a negative outcome, physician respondents were significantly more likely than respondents from the general population to assign responsibility to the physician: 92% [95% CI: 88.3–94.5%] vs. 74% [95% CI: 67.7–78.6%], $\chi^2(1) = 22.08$, $p < 0.001$, OR = 4.08 [95% CI: 2.18–7.92]. The observed difference of 18% exceeded the calculated threshold of 12% in the sensitivity power analysis, confirming that the available sample size was adequate to detect the observed difference. As shown in Figure 3, the most frequently selected responsible actor by both groups was the physician alone (43% vs 38%), followed by combinations of the physician and other human actors.

When categorising responses into 'only physician', 'physician and others', or 'only others', illustrated in figure 4, no statistically significant difference in assigning sole responsibility to the physician was found (43% vs 38%, $p = 0.34$). However, physicians were significantly more likely to assign shared responsibility between the physician and others (48% vs 35%, $p = 0.016$), whereas the general population was more likely to assign responsibility to others only, compared to the physicians (26% vs 8%, $p < 0.001$).

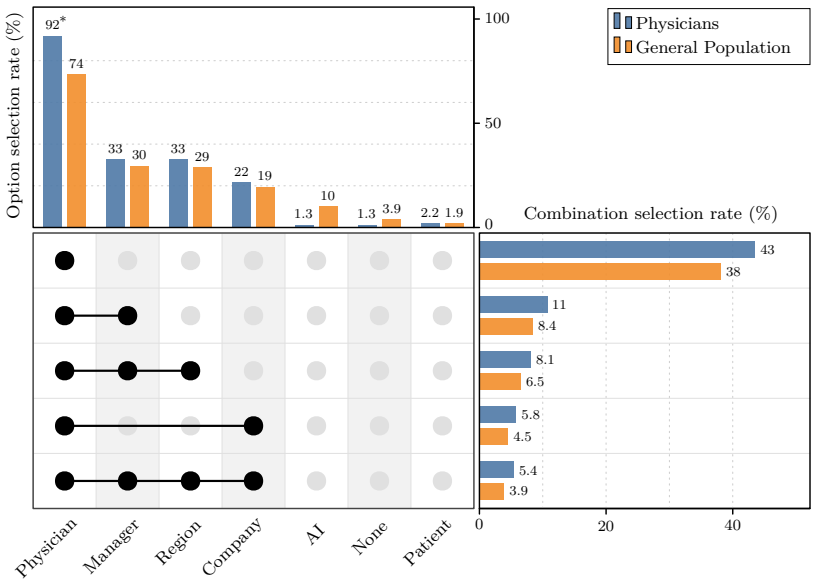


Fig. 3 Who is responsible for errors when the physician uses AI? (*$p < 0.001$). Selection rates for individual options, indicated by the matrix columns, are shown in the top-left panel. Selection rate for the five most frequent combinations, indicated by the matrix rows, are shown in the bottom-right panel. See Figure 1 and 2 for scenario summary.

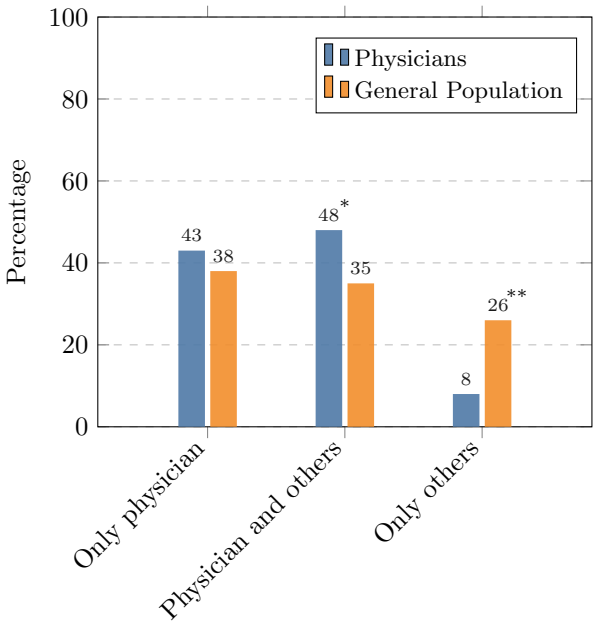


Fig. 4 Who is responsible for errors when the physician uses AI. Grouped by selection of the physician option (*$p = 0.016$, **$p < 0.001$). See Figure 1 and 2 for scenario summary.

4.1.2 Question 2: The Patient Uses AI

As Figure 5 shows, in the second scenario, where the patient interacts directly with the AI for advice without physician involvement, the general population was more likely to assign accountability to the patient (46%) compared to physicians (36%). While the patient alone was clearly the most frequently selected combination among the general population, the physicians' selections were more evenly distributed. The general population also assigned sole responsibility to the patient significantly more frequently than the physicians (24% vs 9.9%, $p < 0.001$) as shown in Figure 6.

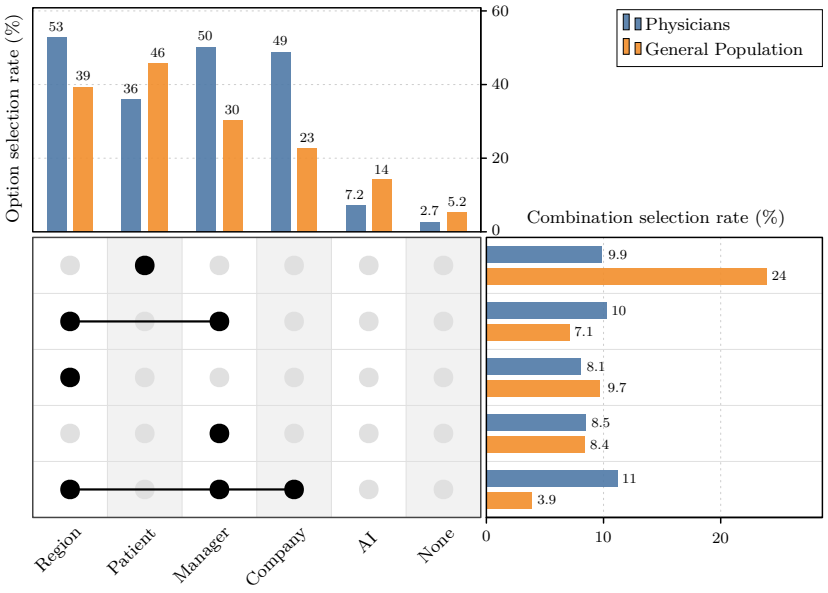


Fig. 5 Who are responsible for errors when the patient uses AI? Selection rates for individual options, indicated by the matrix columns, are shown in the top-left panel. Selection rate for the five most frequent combinations, indicated by the matrix rows, are shown in the bottom-right panel. See Figure 1 and 2 for scenario summary.

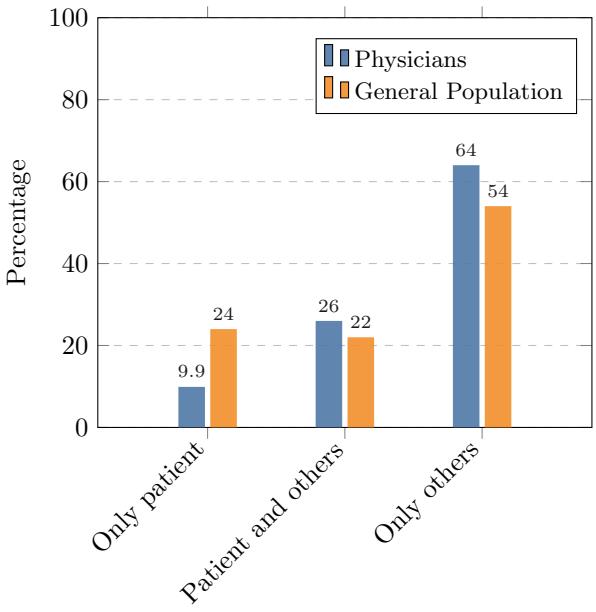


Fig. 6 Who are responsible for errors when the patient uses AI. Grouped by selection of the patient option. See Figure 1 and 2 for scenario summary.

4.1.3 Comparison of Responsibility Distribution Between Scenario 1 and 2

Supplemental Figures A1 and A2 illustrate how respondents who selected the physician as responsible in Scenario 1 distributed their selections in Scenario 2, where the physician was not an option.

For physician respondents, Scenario 2 selections were similar regardless of whether they had selected sole or shared physician responsibility in Scenario 1. A lower proportion of responsibility was also assigned to the patient compared with the other human actors.

In contrast, general-population respondents who assigned sole physician responsibility selected the patient alone more frequently than those who assigned shared responsibility in Scenario 1.

4.2 Transparency

4.2.1 Question 3: Physician AI Use Transparency

A significantly higher proportion of the general population affirmed the patient's right to always know if a physician uses AI compared to the physician group (68% vs. 23%, $p < 0.001$) (Figure 7). The odds of physicians affirming the 'always right to know' position were significantly lower than those of the general population (OR = 0.14 [95% CI: 0.08–0.22]). The physician group exhibited a polarised distribution, with respondents leaning towards the extremes and selecting the middle option less frequently.

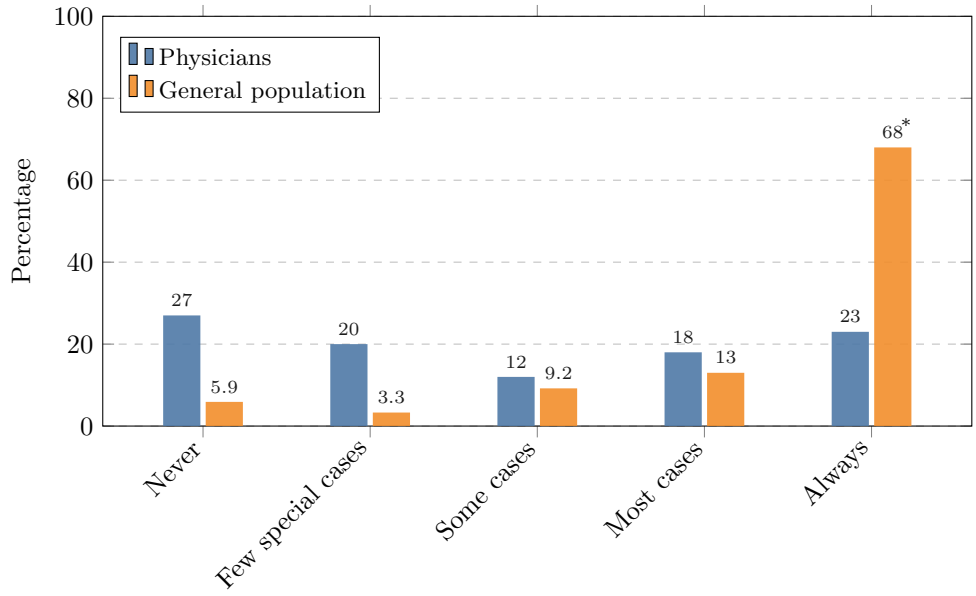


Fig. 7 Patient's right to know if physician uses AI (*$p < 0.001$).

4.2.2 Question 4: AI Chat Transparency

Figure 8 illustrates the general populations' view on the patient's right to know about physician AI use compared to the right to know if a chat is with an AI or a human counterpart. The results indicate that respondents were significantly more likely to consider transparency always necessary in the latter case (82% vs. 68%, $p = 0.0083$).

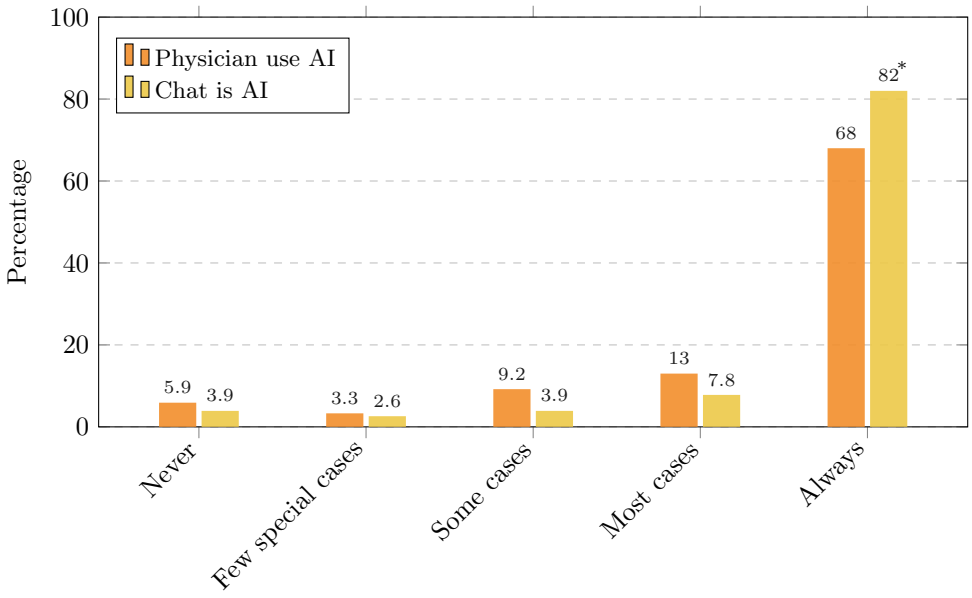


Fig. 8 General population's view on right to know of physician AI use vs chat is AI ($^*p = 0.0083$).

4.3 Thematic Analysis Results

The reflexive TA resulted in three themes: 1) *responsibility for AI use in healthcare is shared between different actors*, 2) *trust in the healthcare system risks being undermined*, and 3) *scepticism towards technical maturity*. These are presented in the following sections. Physician participants are designated PHY1 – PHY108, and general population participants as POP1 – POP73. All quotes have been translated from Swedish to English by the authors.

4.3.1 Responsibility for AI Use in Healthcare Is Shared Between Different Actors

The physician bears the greatest share: Participants consistently describe the physician as a moral and professional agent whose responsibility cannot be offloaded to a 'non-human' entity. Participants framed that *'the AI can only be an aid/knowledge support'* (PHY10) rather than an autonomous decision-maker. Analogies to traditional clinical tools suggest that while AI is an additional tool in the medical toolbox, liability has not transferred to the tool.

> *'Each physician is responsible for their assessment and how they use the tools in their clinical routine. In this way, I believe that AI works in the same way as test results... If a physician misses a serious infection due to a low CRP, it is the physician's responsibility, not the head of the health centre who ensured that we can analyse CRP, and it is not the test result's fault either'.* — PHY1

The physician's responsibility is tied to their *'clinical experience and judgment that should question automated responses'* (PHY23). Even if *'everyone has a shared responsibility'* (POP15), in the event that the AI makes an error, PHY4 states that *'the doctor still has to make her own assessment'* and continue to argue that even if the doctor would make the same assessment *'and did not blindly trust the AI, then she has also made an incorrect assessment'* (PHY4). In this view, the doctor must always make her own, independent judgment for which she is ultimately responsible. The clinician must thus synthesise the AI's output and *'make a summary assessment based on the patient's history and clinical findings'* (PHY10) to reach a final decision.

The organisational and institutional context: Several respondents highlight potential issues regarding the 'systemic pressures' exerted by the healthcare organisation and political institutions.

> *'the employer encourages the use of AI, which may be interpreted as a must requirement. The operations manager is the one who has the overall responsibility for ensuring patient safety'.* — PHY6

The regional management and operations managers are seen as responsible for the environment in which AI is deployed. Thus, depending on how communication and routines around the use of AI are implemented, the individual physician's degree of responsibility may be perceived as mitigated.

'If the directive from either the regional management or the manager was that the physicians should base their assessment on the AI rather than their own, then they are responsible. If the directive was that the physician should use AI as a decision support, then the physician is responsible'. — POP2

The results indicate that liability shifts towards organisational players when the *'region has purchased an insufficiently secure system'* (PHY23) and that managers are *'responsible for implementing assistive devices correctly'* (PHY64). It is noted that *'The region is responsible for clear procedures regarding the use of AI as a tool'* (PHY63), and if these are flawed, the clinician is placed in an untenable position.

Technical integrity and manufacturer liability: The manufacturer of the AI system is positioned as the guarantor of the tool's clinical validity. Participants argue that a company *'should not sell an unsafe product'* (PHY33) and must be held accountable if the system *'fails to provide accurate recommendations'* (POP1). There is a clear distinction between the clinical application of the tool (physician's responsibility) and its technical reliability (manufacturer's responsibility).

'It is the manufacturer's responsibility to ensure that the AI is designed for the tasks it is intended to perform'. — POP27

'Just as a plumber is responsible for the tools used to fix a leak, a doctor is responsible for the tools used in their work.. — POP17

When the outcome is bad, blame is often partitioned in the answers. The manufacturer is described as responsible for having *'developed a flawed product'* (PHY15), while the physician is still responsible for having *'relied entirely on the AI and has not included/made any of her own judgment'* (PHY15).

Synthesis of shared responsibility: Ultimately, the data reflect a 'shared responsibility' model where each actor's failure compounds the other's.

'Everyone has a shared responsibility.' — POP15

As one participant noted, the physician acts as the executor of a decision chain that begins with the creation and procurement of an AI product and ends with clinical application:

'The physician herself bears the highest responsibility as she should be able to assess the patient. [...] The healthcare company should not sell an unsafe product and the region should not buy a program that is unsafe'. — PHY33

While the AI is viewed as a *'decision support, not a definitive answer'* (PHY51), participants suggest that for the usage as a whole to be patient-safe, *'the region should not buy a program that is unsafe'* (PHY33), the manufacturer must ensure it does not *'lack sufficient clinical validity'* (POP1), and the physician must maintain the professional distance necessary to *'decide whether the AI's prescription should be followed'* (POP39).

The responsibility for the usage of AI in clinical practice is thus not perceived as an isolated burden. Accountability is rather viewed to be distributed between several actors involving the physician, healthcare administrators, and the manufacturer of the AI-technology. Despite this view of overall shared liability, there is a clear recurring pattern: the individual physician currently still remains the 'final safeguard', regardless of the sophistication of the tool.

4.3.2 Undermining of Trust in the Healthcare System

The role of the physical encounter: Participants repeatedly emphasised that *'clinical examination and a proper medical history are as important as before'* (PHY26), positioning the physical body as a source of truth that AI cannot access. The act of 'looking', 'touching', and 'listening' is constructed not just as a diagnostic necessity, but as a prerequisite for professional responsibility.

> *'Most of the time, it is enough to LOOK at the patient to understand that something is seriously wrong! LOOK at the patient'.* — POP4

The data suggest that if a physician only relies on AI and neglects the physical exam, ignoring to *'palpate the stomach'* (PHY33) or fails to *'touch and listen to the patient and see the patient's emotions and expressions'* (POP21)—they are not just making a technical error, but are also breaking an informal contract. There is an interpretive sense that *'a human contact cannot be compared to a computer'* (POP49), and to replace the former with the latter risks rendering the medical profession redundant. As one participant noted, if a physician's assessment is merely a reflection of an algorithm, *'the patient could just as easily use AI themselves instead of turning to a physician'* (POP36).

Unlearning of expert knowledge: An expressed concern is that the widespread integration of AI may lead to a degradation of clinical competence. Proponents argue that relying extensively on AI for decision-making could result in deskilling and hinder the acquisition of foundational expert knowledge.

> *'However, I believe that there is a long-term risk that widespread AI use could eventually lead to impaired decision-making and judgment among doctors.'* — PHY1

The conviction that years of clinical experience yield a unique quality of judgment that cannot be replicated is put forward by some as a defence against this notion. PHY50 argues that *'I have a hard time seeing any wise colleague abandon their own overall assessment in favour of something else'*, but continues with reflecting that if physicians lean more and more towards AI for their decision making *'what good then is their experience?'*.

The role of transparency in institutional trust: Trust is further complicated by a potential lack of transparency regarding AI's role in the consultation. Participants argue that a 'zero vision' regarding medical errors is unrealistic—*'mistakes will always*

occur' (PHY34)—but that trust is maintained through the assumption of human oversight. If that oversight is absent, participants felt the patient has a right to know.

> *'It should have been clear to the patient that the assessment was not made by the physician, if the physician cannot take responsibility for the outcome of the assessment'.* — POP9

The results indicate a trust gap between the healthcare institution and the private corporations providing the technology. While a patient may *'trust [their] hospital'* (POP11) and *'trust the doctor's word'* (POP70), they do not explicitly express trust in companies behind the infrastructure. This scepticism is exacerbated by a cynical view of institutional procurement, with suggestions that *'regions are so bad at procurement that they will almost certainly have bought the cheapest/worst variant'* (PHY3).

Individual vs systemic blame: Integration of AI is perceived as threatening to decouple the clinical encounter from its traditional foundations—interpersonal trust and physical examination. Participants expressed a clear concern that if AI displaces the 'human' element of medicine, the legitimacy of the entire healthcare system, rather than just the individual clinician, is brought into question. Several respondents suggest that when AI-assisted errors occur, the healthcare system as a whole risks bearing the reputational brunt.

> *'Even the most experienced of physicians can miss a diagnosis that is atypical; it is easy to be wise in the emergency room with hindsight.* — PHY19

> *'It is the healthcare system that bears the responsibility, not various companies around it. As a person, I trust my hospital, but I do not know which other companies are behind the hospital and therefore I have no trust in them'.* — POP11

Ultimately, the results indicate that for trust to be preserved, the AI should remain an assistive device that *'cannot replace clinical assessment'* (PHY101). The participants' insistence that *'the physician should touch and listen to the patient'* (POP21) can be seen as a defence mechanism against the perceived dehumanisation of care that AI can entail.

4.3.3 Scepticism Towards Technical Maturity and Motivations

Physicians were sceptical about overreliance on AI and envisioned that responsibility would fall on the regional management when hallucinations or misdiagnosis would happen, as described by one participant:

> *'Then the region has a more overall responsibility for letting these notorious liars into everyday clinical practice long before they are ready.'* — PHY11

The responsibility of the region was also echoed by the general population responses, for example, by POP10: *'The regional management and the head of the health centre must be clear (in writing) that physicians should only use AI as support. All physicians must know that they are the ones who make the final decision'.* The responsibility for the region is primarily expressed as ensuring that the procured system is up to standard; *'The employer must ensure that it does not procure substandard systems and lull employees into a false sense of security.'* (PHY14); but also

to make sure that not too many support systems are mixed *'The management/region is responsible for mixing in lots of products and services that complicate or simplify the work.'* (PHY8). It was further expressed that an AI chatbot system might not be enough as support for diagnosis on its own:

> *'It is essential to eliminate any manual data transfer into the AI to assess the severity of the symptoms appropriately. Additionally, the AI must have full access to laboratory results and the ability to analyse this data in conjunction with its predictive algorithms to provide accurate diagnoses.'* — POP1

In this view, an AI system that requires manual input by the physician is not mature enough, given that this input introduces an undefined error margin. Individual data is also essential for the system to become valuable.

Another indication of scepticism towards the maturity of the technology to act on its own was the highlighting by several participants to keep the human in-the-loop: *'AI can be used as guidance but certainly not as the final answer'* (PHY93). In the public responses, this was also repeated: *'AI is far from perfect, which is why the physician, with her experience and multidimensional perspective, can use AI as support but always make the decision based on her own knowledge/feeling/experience'* (POP20). Some participants from the general population also expressed a fear of severe mistakes being made when using AI: *'Triage with AI = deadly! Because then people stop thinking!'* (POP4). PHY3 put it more bluntly, advising the medical profession not to *'trust some decision support crap'* and POP3 gives further advice that *'the specialist doctor should think a little himself and not just rely on the AI which is obviously guessing'*. PHY47 finally, echoes this sentiment more directly: *'I DO NOT think AI is suitable for treatment decisions!'*.

Some physicians asked for other resources instead of AI: *'I think that more time and better education provide better care. The risk with AI use is rather less time, less contact and worse care.'* (PHY46). There was also a mistrust towards the reasons for introducing an AI support system, with some expressing that management *'pressures the doctors to use the product, often to "further develop the business"'* (PHY15). Fear was expressed that this development mostly aims to increase production and, with this, the workload:

> *'One can question the purpose behind why management wants doctors to use AI. For what purpose? To make it easier for doctors, or so that doctors can see more patients?'* — PHY13

Respondents in the general population in turn expressed scepticism towards the underlying motivation of *'the AI companies that market to and manipulate ignorant regional management'* (POP4):

> *'Whoever sells the product just wants their money without guarantees.'* — (POP19)

5 Discussion

5.1 Key Findings

The primary objective of this study was to determine whether significant differences exist in how physicians and the general population attribute responsibility to the physician for negative outcomes when AI is utilised in clinical decision-making. The results indicated a clear divergence between the two groups: physicians were significantly more likely to assign responsibility to the physician (92% vs 74%) and assigned shared responsibility between the physician and other actors to a greater extent (48% vs 35%). Conversely, compared to the physicians group, the general population attributed responsibility more frequently to combinations of actors not including the physician (8% vs 26%). In the second scenario, where the physician was not an option, the general population assigned more responsibility to the patient than the physicians did. The AI itself was not regarded as a potential moral agent that bears responsibility for medical outcomes by either group (1.3% vs 10%).

These quantitative findings were supported by the reflexive thematic analysis, which identified a shared responsibility model where the physician remains the final safeguard. The data confirmed that while responsibility is also distributed among manufacturers, healthcare organisations and the patient, the individual physician continues to bear the largest share of the responsibility burden.

The notion of shared responsibility is a common conception, and was selected by a majority in both respondent groups. However, shared responsibility can be problematic as it tends to get diluted, as has been previously discussed in relation to "the problem of many hands" [25]. Based on qualitative data, respondents stressed the responsibility of healthcare providers and operational managers to implement routines safeguarding against systemic failures. Further, concerns regarding procurement competence and quality assurance in the organisation were also recurring. This may indicate that respondents do not consider the various agents equally responsible for all parts of potential AI failures, but rather intelligently attribute responsibility in line with a particular agent's area of influence.

Our findings differ compared to a previous US survey by Khullar et al. from 2021 where the public was more likely than clinicians to hold physicians responsible for AI-related medical errors (66% vs. 57%). Conversely, US physicians were more inclined than the public to assign liability to algorithm manufacturers (44% vs. 33%) and purchasing healthcare organisations (29% vs. 23%) [59]. This discrepancy may reflect differences between the US and Swedish healthcare models, as well as an evolution in public and professional sentiment toward AI over time. Interestingly, our findings align more closely with UK data from Blease et al. from 2025, which reported that general practitioners feel personally responsible for AI-generated content even during tool failures [12]. However, the magnitude of this perceived personal responsibility appears even more pronounced in our Swedish sample.

Views on transparency diverged between the groups to an even greater extent, where a majority of the general population (68%) affirmed the patient's right to always know if AI was used, in contrast to only 23% of physicians. The results revealed a

polarised physician group, comprising both transparency advocates and those viewing transparency as not necessary. Potential underlying drivers and causes of this polarisation are topics for future studies.

5.2 The Nature of Responsibility: AI as a Tool or an Agent

Although several actors share responsibility, respondents identified the individual physician as ultimately responsible. Physicians are considered to be responsible not only for their own judgment, but also for their use of AI itself, assuming accountability when things go wrong. This aligns with previous work by Lang et al., where all 20 clinicians (cardiologists and nurse coordinators) expressed that if anyone was responsible for the decision of AI in healthcare, it would be the physicians [36].

The finding that the physician group attributed significantly more responsibility to physicians than the general population may reflect a deep-seated professional norm persisting despite AI integration. This norm aligns with the Swedish legal framework, where the Health and Medical Services Act stipulates that each licensed physician bears independent personal medical responsibility for their actions [14]. The thematic analysis underscores this norm through the recurring framing of AI as a tool rather than an autonomous agent, emphasising the physician's role in making independent judgement. Participants frequently invoked analogies to traditional clinical instruments, such as CRP tests, X-ray examinations, or other diagnostic tools.

But are diagnostic AI systems truly tools comparable to traditional diagnostic instruments, or do they rather constitute a different type of entity? Direct biophysical measurements and physical instruments (such as CRP values or stethoscopes) differ in nature from AI systems, as their output correlates directly with a measurable physical or biological state. Traditional statistical tools, such as FRAX, SCORE, or the Wells' score, are more similar to AI models, being based on statistical probabilities derived from large population datasets [60].

However, traditional statistical tools rely on transparent, linear algorithms with finite, well-defined variables, where the calculations can be audited, verified, and understood [61]. In contrast, modern AI architectures process data through complex, non-linear layers and scaling multi-agent networks — introducing a level of behavioural emergence that makes tracing specific conclusions practically impossible [62]. Furthermore, when processing open-ended clinical inputs like text or imaging, AI models exhibit functional non-determinism, where minor input variations can trigger unpredictable output shifts [63]. Compared to traditional diagnostic tools, advanced AI systems thus operate with a degree of structural opacity that defies standard clinical validation and oversight. This distinction constitutes a crucial factor for responsibility attribution, as a refutation of the tool analogy could strengthen the argument for possible intrinsic AI agency.

5.3 Filling the Responsibility Gap

Current legal frameworks governing healthcare responsibility predate autonomous AI. Jassar et al. note that while AI holds potential to improve health services, reliance on unverifiable 'black box' algorithms creates legal uncertainty [64]. As outlined in the

Research Handbook on Health, AI and the Law, malpractice laws typically incentivise adherence to the standard of care, regardless of AI recommendations. Consequently, following a non-standard AI recommendation resulting in injury increases the likelihood of physician liability [65], creating a dilemma for the physician if tasked to choose between following AI advice or the usual standard of care.

Choudhury and Chaudhry [66] note that clinicians are generally viewed as the final decision-makers, bearing the primary moral and legal responsibility. They discuss shared liability with AI providers as a developing area, where liability hinges on training data accuracy and the adequacy of user support. A recent review confirms that legal liability remains unresolved, and as the authors suggest, uncertainty about liability can lead to 'foundational distrust' and hinder the implementation of AI [67].

The responsibility gap concept suggests that the use of AI systems may create situations where human oversight is impossible, and tool failures cannot be traced to human error or manufacturer negligence, resulting in an unresolved responsibility attribution [33]. The response groups attributed negligible responsibility to the AI itself, maintaining that accountability remains firmly human-centric. Our findings thus indicate that professional norms, rather than technological or regulatory solutions, currently bridge this gap in the absence of AI agency.

This has significant implications for AI governance framework development. If physicians continue to accept primary responsibility despite the introduction of advanced AI systems, near-term policy interventions should focus on clarifying the boundaries of this responsibility. Lang suggests the concept of 'shared responsibilisation', where stakeholders actively strive to assume responsibility beyond strict legal requirements to legitimise AI use [35]. If actors outside of the direct medical application pre-emptively assume responsibility to a greater extent, such a framework of responsibilisation could help protect the physician and possibly increase AI adoption and acceptance.

While traditional regulations emphasise individual physician responsibility, a more complex model of shared responsibility could be needed for advanced AI systems. In our data, the physician was the most frequently selected actor, but a majority of respondents did not assign sole responsibility, instead favouring a shared model of accountability. Existing legal interpretations may inadvertently force physicians to accept sole accountability for outcomes driven by systemic factors or algorithmic opacity beyond their direct control. Policy guidance could clarify that even if a physician retains the final decision-making authority, the legal burden of AI-mandated decisions also rests with the health care provider and their procuring department or management. This distinction would help prevent the responsibility gap from becoming a liability gap discouraging the adoption of beneficial technologies.

5.4 The Moral Crumple Zone

The reliance on physicians as ultimate safeguards also aligns with the concept of the *moral crumple zone* [68]. In safety engineering, crumple zones absorb impact to protect passengers; in AI ethics, human operators often absorb blame when autonomous systems fail. The finding that a clear majority of physicians accept personal responsibility suggests that they perceive themselves as part of the moral crumple zone.

Given the opacity of AI systems, physicians' ability to verify AI's reasoning behind their recommendations remains limited, yet they retain responsibility for outcomes. This creates a tension between a theoretical shared responsibility model and practical reality, where human actors are the only accountable agents and physicians bear the largest share.

In a contrasting view, Tigard frames the responsibility gap as a critical crossroads, arguing against maintaining the 'human in the loop' at all costs. If we cannot locate responsibility for autonomous systems, society may be forced to slow AI deployment to ensure human moral oversight. To fully leverage the potential of healthcare AI, we must instead update our moral and legal norms, shifting from a property-based view (requiring human-like consciousness) to a process-based view grounded in social practices of holding agents accountable. By adopting a pluralistic responsibility framework, society can impose consequences on autonomous AI systems such as rewriting problematic code or restricting behaviours, to influence a system's future behaviour [69].

Similarly, Floridi argues AI agents can be held accountable purely by virtue of their interactivity, autonomy, and adaptability [32]. In both accounts, the capacity to act and adapt replaces traditional moral agency as an accountability prerequisite, a decoupling that aligns well with how medicine structures accountability. Clinical responsibility is inherently hierarchical and dynamic, where junior physicians routinely shift primary accountability to specialists and subspecialists. This institutionalised redistribution rests on the principle that responsibility follows demonstrated expertise and knowledge.

Leveraging medicine's existing expertise-tracking hierarchy could offer a practical foundation for a proficiency-based accountability framework for autonomous AI systems. If accountability for clinical decisions is grounded in demonstrated decision-making capability, a rigorously validated AI system outperforming human specialists could bear accountability within its validated clinical domain, even if not qualifying as a fully fledged moral agent. The AI system would assume the structural role of a high-level consultant, bearing accountability proportional to its validated capability. Accountability for errors would then partially shift to the system, relieving physicians of sole accountability when deferring to a demonstrably superior agent.

5.5 Competence, Trust and Transparency

A recurring concern in the data is the potential for deskilling or unlearning of clinical skills due to AI reliance. This mirrors historical anxieties and scepticism regarding new technologies, such as the introduction of the stethoscope [70] or the widespread adoption of evidence-based clinical guidelines in the 1990s, where fears were expressed about the risk of erosion of the physician's intuitive diagnostic capabilities [71].

The relationship between AI and clinical outcomes is likely more complex and nuanced. While adherence to AI recommendations could lead to a deskilling of clinical judgement [72], there is also evidence suggesting that AI may enhance competence in specific sub-domains [73]. AI models have, for instance, demonstrated performance exceeding that of human experts on specific diagnostic tasks such as medical image analysis and skin lesion classification [74, 75].

Several respondents expressed that AI use may lead to a regression in clinical rigour, particularly regarding interpersonal judgment and physical assessment. The physical encounter is viewed as an important factor for maintaining trust, and should AI reliance threaten the physician's ability to perform physical examinations, or explain reasoning face to face, it risks eroding part of the trust in the Healthcare system. A scenario where trust shifts from the medical institution to the corporate provider of the technology could also be problematic, given that public trust in corporate entities is generally lower than in the public healthcare system [76]. A shift also carries a potential ethical vulnerability: private tech providers may exploit pre-existing institutional trust without necessarily inheriting corresponding ethical safeguards, as neither corporate entities,nor their software engineers, are bound by professional ethical frameworks inherent in clinical practice [38].

Within healthcare ethics, transparency is regarded as an important prerequisite for trust and informed consent [37]. The finding that 68% of the general population demanded the right to always know about AI usage contrasts sharply with the physicians' more guarded stance. This divergence suggests a misalignment between professional norms and public expectations; the public perceives AI use as an important factor warranting disclosure, whereas physicians may view it as a background element, similar to other clinical tools requiring no specific disclosure. If AI use evolves in a direction where the physician remains the official decision-maker while the primary decision tool is invisible, patients may still believe they receive care from a physician, even if the decision-making in reality is dominated by a system of which they are unaware. Such a transparency gap could risk unwanted negative effects, should an AI-related adverse event occur. If patients are unaware of AI involvement and the system fails, the erosion of trust could well extend beyond the individual clinician to the healthcare system as a whole.

The public's greater demand for transparency regarding AI chatbots compared to physician AI use also suggests a hierarchy of trust. Patients may be more sceptical of direct AI interaction, regarding it as less reliable than physician-mediated AI. This reinforces the physician as a necessary validator in the eyes of the public, where trust is placed explicitly in physicians as gatekeepers of the healthcare system, as opposed to the more diffuse trust in technology providers, management or organisations providing health care.

5.6 Policy Implications

As previously noted, the findings imply a need for governance frameworks regarding AI in healthcare on several levels. Clear directives on AI usage are needed, distinguishing support and decision-making roles. To mitigate the risk of physicians' moral crumple zones, boundaries of human oversight need clarification. Implementing governance strategies balancing technical validation, ethics, and safety checks prior to deployment is vital to mitigate procurement failures. Regulations could benefit from procurement clauses regarding accountability, audit trails, and the option for patients to opt out of AI-driven decisions. Guidelines could furthermore include criteria for AI safety and transparency, moving beyond general quality assurance, directly addressing AI decision-making challenges.

5.7 Limitations

While the achieved sample size was sufficient, based on the described power analysis performed, the findings are rooted in a Swedish healthcare context, potentially limiting generalisability to other systems with differing regulatory frameworks and traditions. Response rates (45% for physicians and 31% for the general population), although in line with recent healthcare survey research, indicate a potential for non-response bias. While basic sample demographics were representative, respondents may systematically differ in other factors, such as socioeconomic status or country of birth, and may also hold stronger opinions on AI than non-respondents.

Furthermore, as with all survey-based research, the data rely on self-reported responsibility attribution, which may not fully reflect actual decisions made in clinical practice. Although scenarios were designed to reflect realistic clinical situations, they may not fully capture the cognitive load and high-stakes pressure of real-world decision-making, which could influence how responsibility is attributed. Future work should ideally verify the findings with observational studies of AI deployment in real-world settings.

Finally, using a single scenario with two variants limits the analysis of factors influencing responsibility attribution in greater depth. Future research could build on these findings by including multiple scenarios with subtle variations to examine how different conditions affect the respondents' judgments.

6 Conclusion

Although the outcomes of integrating complex AI tools into healthcare depend on several actors, physicians hold themselves responsible for negative outcomes of AI-supported decisions to a high degree. Based on our qualitative and quantitative findings, both physicians and the general population view the individual physician as the critical human-in-the-loop, bearing the ultimate professional accountability of AI use in healthcare. Future governance should consider the balance between the need for innovation and the protection of the physician from undue liability and the preservation of public confidence in the safety of AI-assisted care.

There is a substantial discrepancy between physicians and the general population regarding the perceived need for transparent disclosure of AI use in healthcare. The cause of this discrepancy warrants further investigation. The public's insistence on disclosure highlights a need to define the boundaries of informed consent in the age of AI, where the tool used may be as significant as the decision made.

Looking forward, there is a need for further studies involving a broader perspective of cases in which AI is used in healthcare, to capture a broader range of perspectives and improve generalisability. This could support the design of clearer guidelines on AI accountability that distinguish between clinical decision-making errors by physicians and systemic or technical failures. Further research is required to investigate how AI integration affects clinical decision-making over time and whether such changes shift how physicians and the public attribute responsibility for AI-supported decisions. Addressing these areas will be important in the ongoing work towards safe and transparent deployment of AI as a beneficial technology in healthcare.

Declarations

- Abbreviations:
  AI Artificial Intelligence
  GP General Practitioner / General Practitioners
  TA Thematic Analysis
- Ethics approval and consent to participate: This study was performed in accordance with relevant guidelines and regulations, including the ethical standards of the Declaration of Helsinki and its later amendments. Ethical approval was granted by the Swedish Ethical Review Authority (Etikprövningsmyndigheten), reference number 2024-02513-01. Informed consent to participate was obtained from all of the participants in the study.
- Consent for publication: Not applicable.
- Availability of data and materials: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.
- Competing interests: DA works professionally with software development, including development of AI solutions. No customers are involved in the healthcare segment. RA is a minor shareholder (1 of >1800 individual owners) in Praktikertjänst AB, a private healthcare and dental care cooperative in Sweden. The cooperative does not develop AI. DA and RA have personal savings and occupational pension invested in several standard index-tracking diversified savings funds. Some of the global index funds include large technology companies. DS has had previous unpaid, pro bono activity for the advisory board of the medtech company AI Medical Technology, based in Stockholm, Sweden. RKKG has received financial support from the Swedish Medical Association for general practitioners to attend a conference in 2024. The remaining authors have no competing interests to declare.
- Funding: The research was funded by the Regional Research Council in the Västra Götaland region (VGFOUREG-1033380, project funding – regional collaborative projects). The funding source had no role in the study design, data collection, analysis, decision to publish or preparation of the manuscript.
- Authors' contributions: DA: Conceptualisation, Methodology, Quantitative analysis, Main qualitative analysis, Writing — original draft, Validation, Visualisation. ST: Conceptualisation, Methodology, Main qualitative analysis, Writing — original draft, Supervision. DS: Conceptualisation, Funding acquisition, Data curation, Methodology, Investigation, Joint qualitative discussion, Writing — original draft, Project administration, Supervision. RA: Conceptualisation, Funding acquisition, Data curation, Methodology, Investigation, Validation, Software, Project administration, Writing — review and editing. JW: Conceptualisation, Data curation, Methodology, Investigation, Validation, Software, Writing — review and editing. LA-N: Conceptualisation, Methodology, Investigation, Writing — review and editing. RKKG: Conceptualisation, Funding acquisition, Joint qualitative discussion, Writing — review and editing. PN: Conceptualisation, Funding acquisition, Complementary qualitative analysis, Writing — review and editing. CRB: Conceptualisation, Writing — review and editing. AM: Conceptualisation, Funding acquisition, Joint qualitative discussion, Writing — review

and editing. CW: Conceptualisation, Funding acquisition, Data curation, Writing — review and editing. P-DS: Conceptualisation, Funding acquisition, Joint qualitative discussion, Writing — review and editing. DA is responsible for the overall content as guarantor. During the preparation of this work, the authors used AI tools (ChatGPT and Gemini) to refine language and code (eg, for the survey submission system and statistical analysis in R). After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

- Acknowledgements: Not applicable.
- Clinical trial number: Not applicable.

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section.

Editorial Policies for:

Springer journals and proceedings: https://www.springer.com/gp/editorial-policies

Nature Portfolio journals: https://www.nature.com/nature-research/editorial-policies

Scientific Reports: https://www.nature.com/srep/journal-policies/editorial-policies

BMC journals: https://www.biomedcentral.com/getpublished/editorial-policies

Appendix A Supplementary material

A.1 Review of 'other' option

Responses selecting 'other' were reviewed for assignment of responsibility. Six respondents stated that Dr Maria was still partly responsible in the second scenario, where no physician was directly involved, and the 'other' option was therefore excluded for those respondents. Five physician respondents attributed partial responsibility to the Medical Responsible Physician; this was interpreted as the Operational Manager, the formal level of responsibility in the Swedish context. One respondent attributed responsibility to the Medical Responsible Advisor for procurement and quality, interpreted as the Municipality/Regional level. Five answers corresponded to existing options and were transferred accordingly, while one comment unrelated to the question was excluded.

A.2 Positionality statement

In line with the concept of situated knowledges [77], we advocate for the plurality of perspectives that may otherwise be overlooked. Seven authors identify as men and three as women; all authors are white. Nine authors speak the local language and were born in the country of the study, a democratic state with a strong socialist tradition and universal healthcare. One author is from an English-speaking country in Europe.

Daniel Andersson is a general practitioner with a professional background in IT and is concurrently engaged in consultancy work within programming and systems development.

Rasmus Arvidsson is a general practitioner and researcher, and has previously worked with programming and development of patient facing questionnaires in the medical technology industry.

Jonathan Widén has an M.Sc. in Computer Engineering and, besides working in the industry, researches AI in primary care. He has contributed with technical expertise and programming skills to other AI-related projects and is committed to ensuring AI is being implemented in a way that augments and enhances healthcare staff.

Anna Moberg is a general practitioner, researcher, and senior lecturer. She leads a research group focusing on digital auscultation and digital health technologies. Her research interests include infectious diseases and digital health. As a clinician and researcher working with digital technologies in healthcare, she is familiar with both the opportunities and challenges associated with implementing artificial intelligence and other emerging technologies in clinical practice.

Peter Nymberg is a district nurse, researcher, and associate professor in general practice research with long-standing clinical and academic experience in Swedish primary care. This background shapes his interest in ethically and practically relevant questions in primary care, while requiring reflexive awareness of how professional experience may influence the interpretation of participants' accounts and healthcare practices.

Ronny Gunnarsson is a general practitioner and professor in General practice / family medicine at the University of Gothenburg, Sweden, as well as an adjunct professor at the College of Medicine and Dentistry at James Cook University, Australia. His main research focus has been antibiotic stewardship in primary care, but he has also done research on the doctor-patient consultation, the relationship between PhD students and their supervisors and perceptions of AI. He has experience of both empiric-atomistic (quantitative) and empiric-holistic (qualitative) research.

Pär-Daniel Sundvall is a professor of general practice and Head of the School of Public Health and Community Medicine at the University of Gothenburg, as well as a practising general practitioner. His research encompasses infectious diseases, antimicrobial stewardship, AI and other digital health technologies, and medical education.

Carl Wikberg is a researcher, PhD, and district nurse with a background in primary care and public health research. He is involved in several research projects focusing on primary care, infectious diseases, antibiotic stewardship, adolescent mental health, and the implementation of new models of care. He is familiar with both the opportunities and challenges associated with introducing artificial intelligence and digital tools into clinical practice.

Charlotte Blease, PhD, is a health informaticist with a training in philosophy and ethics. She has written extensively about AI and ethics, particularly from a cognitive science, psychological, and human limitations perspective.

David Sundemo is a postdoctoral fellow, leading a research group with a focus on emergent technology in primary care. He is also a resident physician, specialising in primary care.

Sofia Thunberg is a postdoctoral fellow in human-computer interaction with a background in cognitive science. She has previous experience of AI and robotic healthcare projects through a posthumanistic lens [78, 79].

A.3 Study demographics

Table A1 Respondent characteristics compared with the sampling frame

	Physicians		**General population**	
Answer	Sample	Sample frame	Sample	Sample frame
Total respondents	223	—	155	—
Respondents with known demographics	171	—	93	—
Mean age (±SD)	47±11	48±11	53±17	51±20
Median age (IQR)	45 (37-55)	46 (38-57)	53 (38-65)*	49 (34-66)
Female proportion (%)	46%	51%	48%	50%

*$p < 0.05$ *compared with the sampling frame.*

A.4 Responsibility Distribution in Scenario 1 and 2

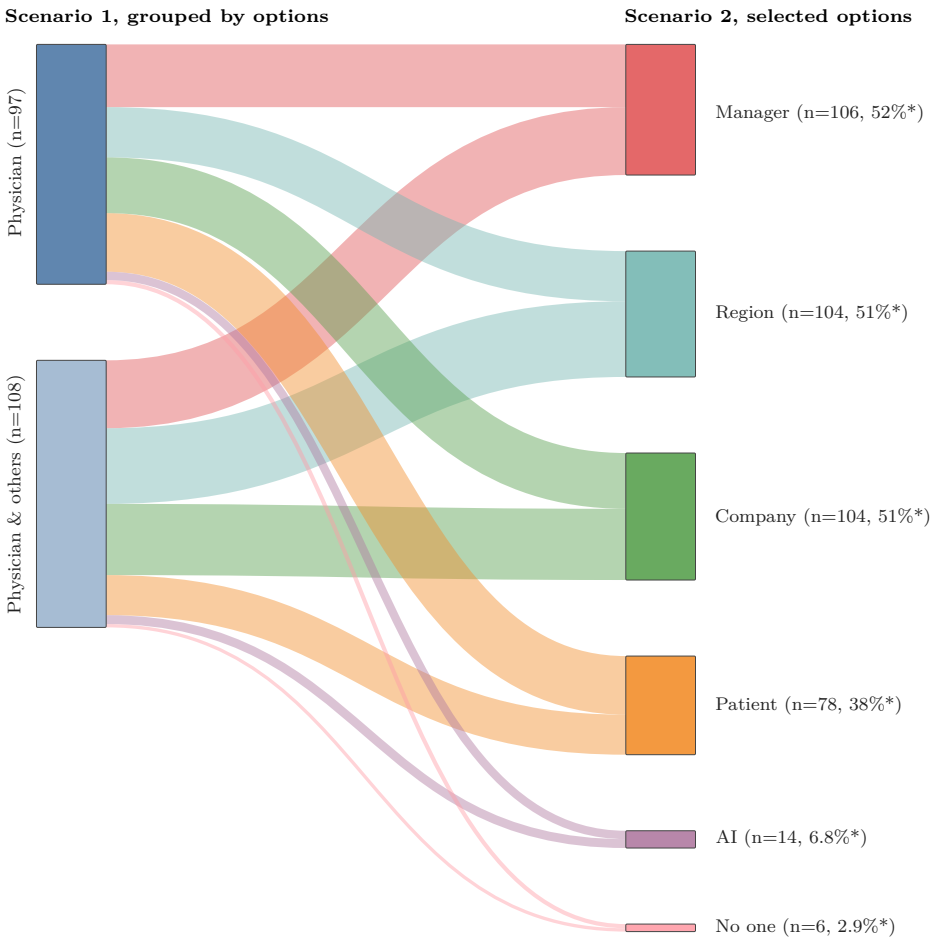


Fig. A1 Scenario 2 selections among physicians who selected the physician as responsible in Scenario 1 (n = 205). Multiple responses were permitted in both scenarios. For Scenario 1, respondents were grouped according to whether they selected the physician alone (n = 97) or together with one or more other actors (n = 108). The right-hand nodes show all options selected by these respondents in Scenario 2. Because each respondent could select multiple options, respondents may contribute to more than one flow and the percentages therefore sum to more than 100%. The 18 physician respondents who did not select the physician in Scenario 1 are not included. *Percentages use n = 205 as the denominator.

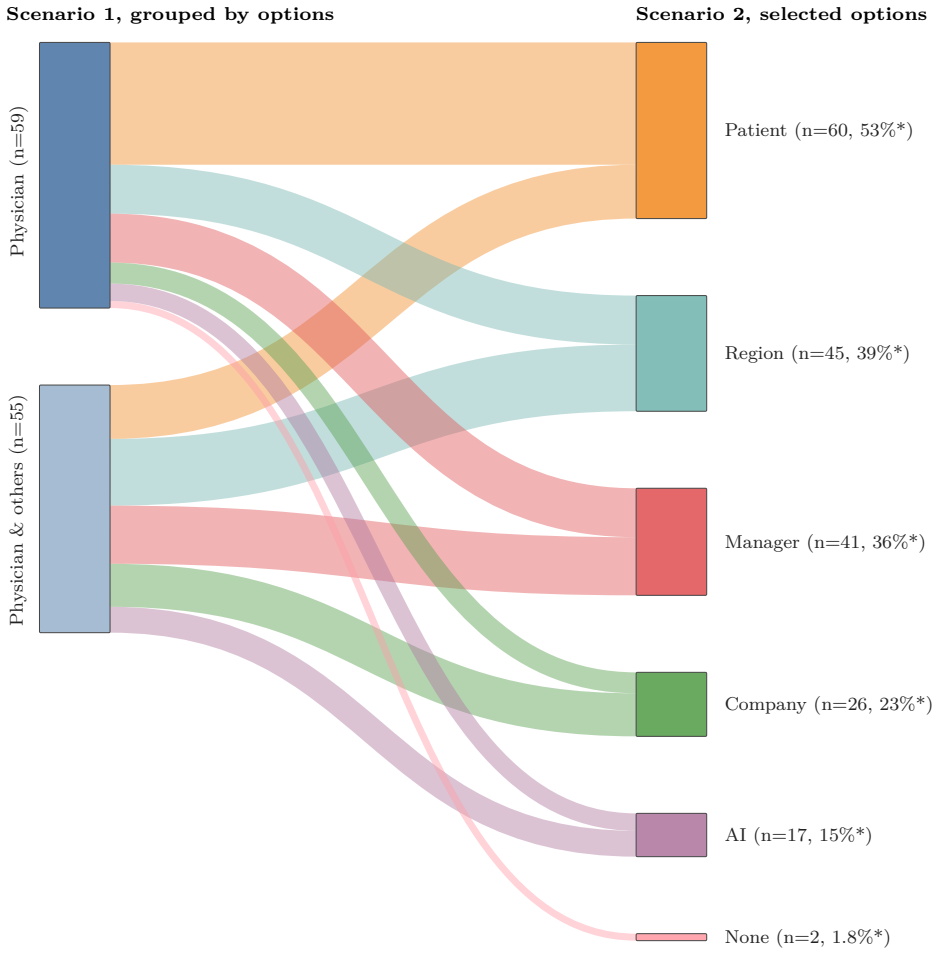


Fig. A2 Scenario 2 selections among the general population who selected the physician as responsible in Scenario 1 (n = 114). Multiple responses were permitted in both scenarios. For Scenario 1, respondents were grouped according to whether they selected the physician alone (n = 59) or together with one or more other actors (n = 55). The right-hand nodes show all options selected by these respondents in Scenario 2. Because each respondent could select multiple options, respondents may contribute to more than one flow and the percentages therefore sum to more than 100%. The 41 respondents who did not select the physician in Scenario 1 are not included. *Percentages use n = 114 as the denominator.

A.5 Supplementary files

The survey documents are provided as supplementary files in their original Swedish versions (SurveyPhysicians.pdf and SurveyGeneralPopulation.pdf) and in English translation (SurveyPhysiciansEn.pdf and SurveyGeneralPopulationEn.pdf).

References

[1] Eric J Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, 25(1):44–56, 2019.

[2] Bailey Richards. Bill Gates says AI will replace doctors, teachers and more in next 10 years, making humans unnecessary "for most things". *People*, 2025.

[3] Stephanie Kumpunen, Cyril Lobont, Lauretta Garrard, Danielle Fisher, Rosa Lau, et al. *How are GPs using AI? Insights from the front line.* Nuffield Trust, 2025.

[4] Charlotte Blease, Carolina Garcia Sanchez, Jens Gaab, Joseph E Alderman, Luke N Allen, et al. Generative artificial intelligence adoption and the future of the primary care profession: An online survey of UK general practitioners' experiences and opinions. *preprint*, 6 2026.

[5] Junaid Bajwa, Usman Munir, Aditya Nori, and Bryan Williams. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthcare Journal*, 8(2):e188–e194, 2021.

[6] Karim Lekadir et al. *Artificial intelligence in healthcare: applications, risks, and ethical and societal impacts.* European Parliamentary Research Service, 2022.

[7] Rameez Ahmad, Khawaja Hisham-Ul-Hassan, Muhammad Ammar, Masood Cheema, Mehreen Naz, et al. Integration of ai tools in healthcare, employee perceptions, productivity and workload management. *Migration Letters*, 21:1457–1469, 2024.

[8] Nishita Mehta, Anil Pandit, and Sharvari Shukla. Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study. *Journal of Biomedical Informatics*, 100:103311, 2019.

[9] Mamoon Rashid and Madhuri Sharma. AI-assisted diagnosis and treatment planning — a discussion of how AI can assist healthcare professionals in making more accurate diagnoses and treatment plans for diseases. In Rajesh Singh, Anita Gehlot, Navjot Rathour, and Shaik Vaseem Akram, editors, *AI in Disease Detection*, chapter 14, pages 313–336. John Wiley & Sons, Ltd, 2025.

[10] Malvika Sharma, Carl Savage, Monika Nair, Ingrid Larsson, Petra Svedberg, and Jens M Nygren. Artificial intelligence applications in health care practice: Scoping review. *J Med Internet Res*, 24(10):e40238, Oct 2022.

[11] M. I. Ahmed, B. Spooner, J. Isherwood, M. Lane, E. Orrock, et al. A systematic review of the barriers to the implementation of artificial intelligence in healthcare. *Cureus*, 15(10):e46454, 2023.

[12] Charlotte Blease, Carolina Garcia Sanchez, Cosima Locher, Brian McMillan, Jens Gaab, et al. Generative artificial intelligence in primary care: Qualitative study of UK general practitioners' views. *Journal of Medical Internet Research*, 27, 2025.

[13] Sveriges läkarförbund. Swedish healthcare system. https://slf.se/in-english/swedish-health-care-system. Accessed: 2026-05-19.
[14] Sveriges riksdag. Hälso- och sjukvårdslag (2017:30). https://rkrattsbaser.gov.se/sfst?bet=2017:30, 2017. Swedish law. Accessed: 2026-05-19.
[15] Sveriges riksdag. Patientlag (2014:821). https://rkrattsbaser.gov.se/sfsr?bet=2014:821, 2014. Swedish law. Accessed: 2026-05-19.
[16] Sveriges riksdag. Patientsäkerhetslag (2010:659). https://rkrattsbaser.gov.se/sfst?bet=2010:659, 2010. Swedish law. Accessed: 2026-05-19.
[17] Government Offices. For sustainable digital transformation in Sweden - a digital stategy, 2017.
[18] Regeringskansliet. Sveriges digitaliseringsstrategi 2025–2030, 2025.
[19] Merel Noorman. Computing and Moral Responsibility. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2023 edition, 2023.
[20] C. W.L. Ho, D. Soon, K. Caals, and J. Kapur. Governance of automated image analysis and artificial intelligence analytics in healthcare. *Clinical Radiology*, 74:329–337, 5 2019.
[21] Kathryn P. Linthicum, Katherine Musacchio Schafer, and Jessica D. Ribeiro. Machine learning in suicide science: Applications and ethics. *Behavioral Sciences & the Law*, 37:214–222, 5 2019.
[22] Hans Jonas. *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. University of Chicago Press, Chicago, 1984.
[23] Ron Westrum. The social construction of technological systems. *Social Studies of Science*, 19(1):189–191, 1989.
[24] Ulrike Felt, Rayvon Fouché, Clark Miller, and Laurel Smith-Doerr. *The Handbook of Science and Technology Studies, Fourth Edition*. The MIT Press, 01 2017.
[25] Dennis F. Thompson. Moral responsibility of public officials: The problem of many hands. *American Political Science Review*, 74(4):905–916, 1980.
[26] Peter Strawson. Freedom and resentment. *Proceedings of the British Academy*, 48:187–211, 1963.
[27] Neelke Doorn and Ibo van de Poel. Editors' overview: Moral responsibility in technology and engineering. *Science and Engineering Ethics*, 18(1):1–11, March 2012.
[28] Daniel C. Dennett. When HAL kills, who's to blame? computer ethics. In David G. Stork, editor, *HAL's Legacy: 2001's Computer as Dream and Reality*, pages 351–365. MIT Press, Cambridge, MA, 1997.
[29] John P. Sullins III. When is a robot a moral agent? *The International Review of Information Ethics*, 6:23–30, December 2006.
[30] Deborah G. Johnson. Computer systems: Moral entities but not moral agents. *Ethics and Information Technology*, 8(4):195–204, November 2006.
[31] Sven Nyholm. Attributing agency to automated systems: Reflections on human-robot collaborations and responsibility-loci. *Science and Engineering Ethics*, 24, 08 2018.
[32] Luciano Floridi and J. W. Sanders. On the morality of artificial agents. *Minds and Machines*, 14(3):349–379, August 2004.

[33] Andreas Matthias. The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3):175–183, September 2004.

[34] Filippo Santoni de Sio and Giulio Mecacci. Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology*, 34(4):1057–1084, December 2021.

[35] Benjamin H. Lang, Sven Nyholm, and Jennifer Blumenthal-Barby. Responsibility gaps and black box healthcare AI: Shared responsibilization as a solution. *Digital Society*, 2(3):52, 2023.

[36] Ben H. Lang, Kristin Kostick-Quenet, Jared N. Smith, Meghan Hurley, Rita Dexter, et al. Should physicians take the rap? normative analysis of clinician perspectives on responsible use of 'black box' AI tools. *AJOB Empirical Bioethics*, 16:201–212, 2025.

[37] Nir Eyal. Using informed consent to save trust. *Journal of Medical Ethics*, 40(7):437–444, 2014.

[38] Michael Guckert, Kristina Milanovic, Jennifer Hannig, David Simon, Tamara Wettengl, et al. The disruption of trust in the digital transformation leading to health 4.0. *Frontiers in Digital Health*, 4, 3 2022.

[39] Thomas Ploug, Amanda Jørgensen, and Søren Holm. Drivers of trust in AI across the domains of finance, law, and healthcare: a conjoint study. *AI & Society*, 2026.

[40] M. Chau, M. G. Rahman, and T. Debnath. From black box to clarity: Strategies for effective AI informed consent in healthcare. *Artificial Intelligence in Medicine*, 167, 9 2025.

[41] Sally Moy, Mona Irannejad, Stephanie Jeanneret Manning, Mehrdad Farahani, Yomna Ahmed, et al. Patient perspectives on the use of artificial intelligence in health care: A scoping review. *Journal of Patient-Centered Research and Reviews*, 11:51–62, 4 2024.

[42] Erik von Elm, Douglas G Altman, Matthias Egger, Stuart J Pocock, Peter C Gøtzsche, Jan P Vandenbroucke, and for the STROBE Initiative. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *PLOS Medicine*, 4(10):1–5, 10 2007.

[43] Robert M. Groves, Floyd J. Fowler, Mick. Couper, James M.. Lepkowski, Eleanor. Singer, et al. *Survey methodology*. Wiley, 2009.

[44] Floyd J.. Fowler. *Survey research methods*. SAGE, 2014.

[45] Rasmus Arvidsson, Jonathan Widén, Lina Al-Naasan, Ronny Gunnarsson, Peter Nymberg, et al. Acceptable accuracy for medical AI: a survey of physicians and the general population in Sweden. *BMJ Health & Care Informatics*, 33(1), April 2026.

[46] Jan P. Vandenbroucke, Erik von Elm, Douglas G. Altman, Peter C. Gøtzsche, Cynthia D. Mulrow, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): Explanation and elaboration. *International Journal of Surgery*, 12:1500–1524, 12 2014.

[47] Daniël Lakens. Sample size justification. *Collabra: Psychology*, 8, 3 2022.

[48] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. G*Power

3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(1):175–191, 2007.

[49] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3:77–101, 2006.

[50] Virginia Braun, Victoria Clarke, Elicia Boulton, Louise Davey, and Charlotte McEvoy. The online survey as a qualitative research tool. *International Journal of Social Research Methodology*, 24:641–654, 2021.

[51] R Core Team. R: A language and environment for statistical computing, 2026.

[52] Alexander Lex, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. Upset: Visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1983–1992, 2014.

[53] Zuguang Gu. Complex heatmap visualization. *iMeta*, 1(3):e43, 2022.

[54] J.J. Allaire, Christopher Gandrud, Kenton Russell, and CJ Yetman. networkd3: D3 javascript network graphs from r, 2025. R package version 0.4.1.

[55] Virginia Braun and Victoria Clarke. Supporting best practice in reflexive thematic analysis reporting in palliative medicine: A review of published research and introduction to the reflexive thematic analysis reporting guidelines (RTARG). *Palliative Medicine*, 38:608–616, 6 2024.

[56] Linda Finlay. Negotiating the swamp: The opportunity and challenge of reflexivity in research practice. *Qualitative Research*, pages 209–230, 2002.

[57] Natasha S Mauthner and Andrea Doucet. Reflexive accounts and accounts of reflexivity in qualitative data analysis. *Sociology*, 37:413–431, 2003.

[58] Tejaswini Joshi, Heidi Biggs, Jeffrey Bardzell, and Shaowen Bardzell. Who is "I"?: Subjectivity and ethnography in HCI. In *Conference on Human Factors in Computing Systems (CHI)*. Association for Computing Machinery, 5 2024.

[59] Dhruv Khullar, Lawrence P. Casalino, Yuting Qian, Yuan Lu, Enoch Chang, et al. Public vs physician views of liability for artificial intelligence in health care. *Journal of the American Medical Informatics Association*, 28:1574–1577, 7 2021.

[60] J. A. Kanis, O. Johnell, A. Oden, H. Johansson, and E. McCloskey. FRAX and the assessment of fracture probability in men and women from the UK. *Osteoporosis International*, 19:385–397, 2008.

[61] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1:206–215, 5 2019.

[62] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, et al. Emergent abilities of large language models. *Transactions on Machine Learning Research*, 10 2022.

[63] Alex John London. Artificial intelligence and black-box medical decisions: Accuracy versus explainability. *Hastings Center Report*, 49:15–21, 1 2019.

[64] Sunam Jassar, Scott J. Adams, Amy Zarzeczny, and Brent E. Burbridge. The future of artificial intelligence in medicine: Medical-legal considerations for health leaders. *Healthcare Management Forum*, 35:185–189, 5 2022.

[65] Sara Gerke, I. Glenn Cohen, and W. Nicholson Price II. Liability for use of artificial intelligence in medicine. In Solaiman B and Cohen IG, editors, *Research Handbook on Health, AI and the Law*, chapter 9. Edward Elgar Publishing, 2024.

[66] Avishek Choudhury and Zaira Chaudhry. Large language models and user trust: Consequence of self-referential learning loop and the deskilling of health care professionals. *Journal of Medical Internet Research*, 26, 1 2024.

[67] Tina Tung, Shah Md Nehal Hasnaeen, and Xiaopeng Zhao. Ethical and practical challenges of generative AI in healthcare and proposed solutions: a survey. *Frontiers in Digital Health*, 7, 2025.

[68] Madeleine Clare Elish. Moral crumple zones: Cautionary tales in human-robot interaction. *Engaging Science, Technology, and Society*, 5:40–60, 3 2019.

[69] Daniel W. Tigard. Artificial moral responsibility: How we can and cannot hold machines responsible. *Cambridge Quarterly of Healthcare Ethics*, 30:435–447, 7 2021.

[70] René Figueroa and Joseph Varon. Auscultation transformed: the ingenious invention of the stethoscope by rené théophile hyacinthe laënnec. *History and Philosophy of Medicine*, 6:9, 2024.

[71] Hallam Stevens. Evidence-based medicine from a social science perspective. *Australian Journal of General Practice*, 47(12):889–892, December 2018.

[72] Christina Nilsson. The artificial intelligence (AI) competence paradox: how AI reshapes clinical expertise. *Transforming Government: People, Process and Policy*, 4 2025.

[73] Eyad A Qunaibi, Ayman M Al-Qaaneh, Baraa F Ismail, Hussam I Muhidat, Farhia S Rageh, Najwa A Musallam, and Alaa K Fawzy. Effectiveness of informed AI use on clinical competence of general practitioners and internists: Pre-post intervention study. *JMIR Med Educ*, 12:e75534, Feb 2026.

[74] Titus J. Brinker, Achim Hekler, Alexander H. Enk, Joachim Klode, Axel Hauschild, et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *European Journal of Cancer*, 113:47–54, 2019.

[75] Scott Mayer McKinney, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, et al. International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788):89–94, January 2020.

[76] Björn Binzer, Jennifer Kendziorra, Anne Katrin Witte, and Till J. Winkler. Trust in public and private providers of health apps and usage intentions: A sectoral privacy calculus and control perspective. *Business and Information Systems Engineering*, 66:273–297, 6 2024.

[77] Donna Haraway. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14(3):575, 1988.

[78] Maria Puig de la Bellacasa. Matters of care in technoscience: Assembling neglected things. *Social Studies of Science*, 41(1):85–106, feb 2011.

[79] Amelia DeFalco. Towards a theory of posthuman care: Real humans and caring robots. *Body and Society*, 26:31–60, 2020.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SurveyPhysicians.pdf
- SurveyGeneralPopulation.pdf