Human-Centered AI in Healthcare: Balancing Patient Autonomy and Physician Judgment

Anne Gerdes
The University of Southern Denmark, Department of Design, Media and Educational Science, Kolding, Denmark

gerdes@sdu.dk

Abstract: This article outlines ethical issues related to integrating artificial intelligence (AI) into shared decision making (SDM), focusing on how to meet: (1) the need for explainability in enacting autonomy, (2) the need for respecting patients' values and preferences in treatment decisions, and (3) the impact of AI on physician expertise. First, it is argued that the kind of explainability required to support patient and physician autonomy can be met through rigorous model validation combined with context-sensitive post hoc explanations. Next, turning to a patient perspective, the article argues against the assumption that having AI pre-rank treatment recommendations undermines patient autonomy and therefore ought to be avoided. Instead, the article recognizes AI's potential to reduce cognitive overload and emphasizes balancing AI-guided decision-making properly. Subsequently, the physician's perspective is considered, analyzing how AI impacts physician expertise, particularly in light of automation bias, deskilling, and the erosion of practice-based judgment. The article warns against a shift toward actuarial decision-making driven by algorithmic risk stratification, which may compromise core ethical principles. The article concludes by promoting human-centered AI integration to enhance human agency—empowering patients to make informed choices and allowing physicians to exercise sound clinical judgment.

Keywords: AI healthcare, Shared decision making, Explainability, Autonomy, Practice-Based judgment

1. Introduction

AI in healthcare is a powerful tool that relies on data-driven machine learning to build predictive models. Thus, AI tools can offer diagnostic support, risk assessment, prediction of treatment responses, or rank treatment recommendations, potentially enhancing medical decision-making. However, integrating AI into clinical practice also raises fundamental challenges for shared decision-making (SDM), patient autonomy, and physician expertise. In this context, the article discusses the patient-physician-centered relational aspects of SDM, emphasizing challenges to patient and physician autonomy in AI-supported healthcare while also highlighting AI's potential benefits in enhancing SDM. Furthermore, the article analyzes AI's impact on physician expertise, warns against an actuarial turn, and discusses deskilling in the wake of AI-supported diagnostics.

The article is structured as follows: Section 2 analyzes how AI influences SDM, explicitly addressing the impact of transparency challenges on patient and physician autonomy. It is argued that post hoc explainability combined with thorough validation of AI models preserves the parties' opportunities to enact autonomy. Furthermore, besides concerns about transparency and addressing a patient perspective, the article argues against AI pre-ranked treatment recommendations being rejected because the pre-ranking threatens patient autonomy. Instead, the article points to AI's potential to reduce cognitive overload, which means we ought to focus on how best to turn information overload into manageable information while attending to the patient's values and preferences. Subsequently, drawing on examples from radiology, section 3 considers the physician's perspective, focusing on automation bias, deskilling, and the potential erosion of practice-based judgment. The article warns against a shift toward actuarial decision-making driven by algorithmic risk stratification, which may compromise core ethical principles. The article concludes by advocating for human-centered AI integration designed to augment human agency—empowering patients to make informed decisions and preserving physicians' capacity for sound clinical judgment.

2. The Impact of AI on Shared Decision-Making

Patient-centered care implies that the patient ought to be treated as an equal in SDM concerning medical or treatment decisions. While sharing information is a precondition for SDM, it also involves encountering the patient in a unique context, aiding the patient in eliciting their treatment preferences, which requires the "creation of an informed decision space in which a patient's specific values and preferences are represented and respected" (Bjerring and Busch 2021). However, respecting the patient's autonomy in informed decision-making should not place the entire burden of complicated treatment decisions on the shoulders of the patient. Thus, in SDM, the physician must balance participating without dominating the process (Charles et al. 1997).

Against this setting, the success of AI depends on transparency in ensuring trustworthy AI-supported decision-making. Consequently, making an informed decision requires that shared information be adequately explained,

which implies that it is understandable how an AI model arrives at an output. The quest for transparency is echoed in empirical studies too, as people demand explainable AI in healthcare settings, including human oversight and accountability (Holm and Ploug 2023; Ploug et al. 2021). Moreover, disregarding explainability in favor of performance accuracy could lead to a kind of double paternalism, first between the physician and the AI tool, which delivers accurate yet non-understandable recommendations that the physician follows, whereafter information is delivered to the patient, who follows the physician (Lorenzini et al. 2023).

Clearly, a machine learning system based on linear regression is straightforward and understandable as its model reflects a simple relationship between input and output. The model is interpretable because we can scrutinize its inner workings and clarify how it reached a decision. However, many models of high-performance complex neural networks are not interpretable and require explainer models to create approximate post hoc explanations. Therefore, transparency necessitates explainable AI (XAI) at a level that can provide a post hoc approximate explanation of what goes on inside noninterpretable AI models (Goebel et al. 2018; Linardatos et al. 2021; Ribeiro et al. 2016). Although it is impossible to reach a complete explanation with explainers, the application of black-box AI models can be considered epistemically and ethically justified if the system offers a pragmatically useful, albeit non-exhaustive, explanation of its outputs and is accompanied by rigorous model validation after standards comparable to the randomized controlled trials standard (RCT) (Gerdes 2024). In SDM contexts, the combination of retrospective assessment, prospective validation, and post hoc explanations may sufficiently meet the demand for explainability. Nevertheless, ensuring that AI-supported SDM preserves the autonomy of both patient and physician takes more than transparency. Consequently, in what follows, the article addresses patient autonomy by discussing the pros and cons of AI-generated treatment recommendations, followed by a discussion of AI challenges to professional autonomy (sec. 3).

AI-driven rankings might challenge patient autonomy as they potentially structure patient choices in a way that subtly nudges them toward a particular treatment. However, one may argue that this kind of libertarian paternalism (Thaler et al. 2010; Thaler and Sunstein 2003) might be considered justifiable, as the system's ranking reconstructs complex knowledge, providing a form of choice architecture that reflects the most optimal therapeutic options in alignment with what a rational agent would consider to be in their best interest. The persuasive element is epistemically sound, and the structure facilitates patients in reflecting on their values and preferences. As an illustrative case, Shapiro et al. (2023) describe XDECIDE, a clinical decision support system deployed in the context of precision oncology. This system provides patients with personalized treatment recommendations by drawing on real-world evidence from a cancer outcomes registry, findings from medical and clinical research, and the expertise of "molecular pharmacologists and oncologists (…) to balance $_{X}$DECIDE AI computational advantages with human intuition and experience" (Shapiro et al. 2023). As such, human experts discuss AI-ranked therapeutic options. A finalized report is delivered via an interactive web platform to the patient and oncologist, allowing each to suggest their preferred treatment.

One could argue that this approach undermines shared decision-making as the patient's autonomy is threatened because they are presented with a pre-ranked list, implying that their personal preferences and values do not drive the ranking (Aurelia 2023; Debrabander and Mertes 2022; McDougall 2019). However, in the $_{X}$DECIDE scenario, information overload is limited, which might otherwise have hampered the patient's ability to exercise their autonomy in the first place. Suppose the patient is presented with a scrambled list to avoid influencing their choice. In that case, they might prefer treatment B over treatment C and turn to the physician for expert clarification on this comparison. If B proves superior from an expert perspective, the patient will likely continue by comparing B with treatment A, then potentially with D, and so forth—working through pairwise comparisons to understand the relative merits of each option. It is reasonable to assume the patient becomes overwhelmed and cognitively overloaded in such circumstances, particularly when multiple treatment options exist. Providing a pre-ranked list makes information manageable, enabling the patient, in collaboration with the oncologist, to comprehensively assess and anticipate the consequences of different treatment options without exhausting their cognitive resources. One should, of course, not underestimate that in a clinical crisis, a patient can get overwhelmed by even a single option. Still, this approach facilitates patients incorporating their values and preferences in ranking the pre-ranked options, thereby sustaining their autonomy. To facilitate SDM, it is essential to carefully consider how to strike a proper balance, turning information overload into manageable information while attending to the patient's values and preferences. Likewise, the discussion of AI's impact on autonomy includes attention to how AI tools might support or challenge the physician's expertise.

3. Challenges to Physician Expertise - Deskilling and the Risk of an Actuarial Turn

The following discussion will explore how AI tools can either hinder or improve physicians' expertise. It will depart from radiology, where significant advancements in computer vision technology have greatly contributed to the development and integration of AI applications. As such, radiology can be viewed as a leader in adopting AI. Consequently, traditional workflows are changing within radiology, and the radiologist's role has been elevated in many ways, as routine tasks can now be distributed to AI tools (Najjar 2023).

In medical imaging for breast cancer detection, AI tools can support radiologists during screenings to reduce work pressure. For instance, in double-read mammography screening, an AI tool may function as a second reader, and studies demonstrate improved detection accuracy within AI-supported scenarios (Lång et al. 2023). Likewise, a retrospective population-wide mammography screening accuracy study concludes that AI makes it feasible to replace one or partially both readers in double-read mammography screening while also emphasizing the need for a strong quality and assessment setup before deploying AI-integrated screening (Elhakim 2024).

Moreover, in various medical domains, studies demonstrate that hybrid decision-making, teaming humans and AI, leads to better performance than either (Hekler et al. 2019; Šķilters et al. 2024). For instance, referring to "AI-augmented" radiologists, Cheikh et al. (2022) find evidence that AI-assisted decision-making facilitates radiologists and increases diagnostic confidence "through the high sensitivity and NPV [negative predictive values] of AI" (Cheikh et al. 2022). This observation aligns with Reverberi et al. (2022), who conclude: "In hybrid decision-making, the individual strengths of humans and AI come together to optimize the joint outcome."

However, as AI takes over routine tasks—such as reading images and filtering out "simple" cases— it might challenge opportunities for developing and maintaining reading expertise. If radiologists focus solely on complicated cases with potential pathology, their pattern recognition skills could weaken due to a lack of exposure to routine variations. Thus, potential deskilling raises fundamental questions about the nature of expertise and professional judgment, which have traditionally been cultivated through theoretical knowledge and practice-based experience. In healthcare, as in other professional domains, the development of skills and the ability to exercise professional judgment is reflected in the Aristotelian notion of practical wisdom, i.e., *phronesis*, which enables professionals to make sound judgments while attending to morally relevant aspects of concrete situations (Aristotle 1934; Eikeland 2008). This perspective on enacting expertise in practice emphasizes an interconnectedness of theoretical knowledge and practice-based expertise. Against this backdrop, expertise develops from repeated engagement with the world, a process also recognized in medical training and quality assurance frameworks. For instance, in the field of radiology, according to *The European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis* (European Commission et al. 2013), breast radiologists must read at least 1000 mammograms annually to maintain diagnostic skills, and those participating in screening programs must read at least 5000 mammograms. Additionally, studies have shown that reading performance decreases for radiologists who undertake fewer examinations (Théberge et al. 2014), while those with a higher volume of mammogram readings have lower false-positive rates (Wong et al. 2023). Moreover, Wong et al. (2023) caution against using single reader characteristics to measure experience in mammography interpretation as their studies demonstrate that "A combination of reader characteristics such as feedback, lifetime mammograms read, number of CME credit, practice type has been shown to provide a better measure of experience, and reader performance" (Wong et al. 2023). These observations highlight the limitations of reducing expertise to simplistic quantitative metrics. The development of expertise is characterized by the expert having a deep situational understanding of their domain and relying on tacit experience-based knowledge when making judgments (Dreyfus 1986). Just as *phronesis* develops through iterative experience, radiological expertise is reinforced through repeated practice, ensuring excellence in professional judgment. In addition, although AI support may enable inexperienced radiologists to perform at the same level as more experienced ones, it is crucial to ensure that novices still acquire the necessary skills in the first place.

Furthermore, another source of deskilling is the human inclination to trust automated decisions without critical thinking. This tendency toward automation bias is well-documented in human factors engineering and human-computer interaction (e.g., Sarter 2001) as well as in medicine (e.g., Kim et al. 2025; Tsai et al. 2003). Automation bias refers to the phenomenon where individuals uncritically accept incorrect system outputs, disregarding contradictory evidence, even when such evidence is accurate or aligns with their professional expertise. Dratsch et al. (2023) demonstrate that novices, moderately experienced radiologists, and very experienced radiologists perform equally well when reading mammograms, regardless of whether AI predictions are correct, and automation bias occurs across all levels of expertise. Nevertheless, very experienced readers overrule incorrect

AI predictions more often than the other groups. Similarly, Kim et al. (2025) find that inexperienced readers "recommended significantly more intense follow-up examinations when presented with false-positive AI findings."

Dratsch et al. (2023) argue that automation bias can be reduced by fostering critical reflection, e.g., providing users with confidence levels for system outputs and incorporating explainable interfaces that offer insight into how the AI tool makes its decisions. They also emphasize that automation bias might be reduced by ensuring users take responsibility for their own decisions. However, these arguments assume that future radiologists will have adequate opportunities to cultivate the expertise necessary for sound professional judgment. But, if physicians lose opportunities to build and maintain skills while AI's accuracy increases, the benefits of hybrid human-AI team decision-making collapse, undermining physicians' ability to exercise professional judgment and potentially leading to further deskilling and responsibility gaps.

However, although human oversight remains an ethical and legal requirement at present, the prospect of AI outperforming physicians could shift discussions toward fully autonomous AI decision-making. Moreover, the present public preferences regarding explainability and human accountability (Ploug et al., 2021) may evolve, potentially leading to patients preferring AI accuracy over traditional physician-patient trust relationships. On the face of it, it might seem tempting to envision a future scenario in which AI outperforms humans. But such a scenario would compromise the foundational principles of SDM, and it would presumably not be realizable in the near future.

First, AI decision-making relies on predictive models that generate probability-based assessments rather than holistic clinical judgments. Thus, applying fully autonomous AI tools risks prioritizing statistical optimization over patient-centered care, implying that medical decisions become driven by algorithmic risk stratification rather than holistic clinical reasoning and SDM. Such an actuarial turn in healthcare would threaten principles of medical ethics, particularly regarding autonomy, beneficence, and the role of physician-patient trust (Beauchamp and Childress 2019).

Second, even in promising AI application areas such as medical imaging, systematic reviews show that many studies on AI tools outperforming clinicians lack proper validation (Freeman et al. 2021; Nagendran et al. 2020; Wynants et al. 2020). Although progress has been made, as demonstrated by Lång et al. (2023), significant challenges remain (Lekadir et al. 2022). One such challenge is data poverty in healthcare, which may result in models failing to serve the needs of underrepresented groups (Gao et al. 2023). Similarly, predictive accuracy depends on carefully selecting proxies to avoid inappropriate choices inadvertently reinforcing health inequalities (Obermeyer et al. 2019). The lack of prospective and randomized controlled trials, standardized reporting, and the inclusion of underrepresented groups present challenges that highlight the importance of ethics in AI development within healthcare (Plana et al., 2022), while also underscoring the difficulties in realizing fully autonomous predictive AI for complex cases.

Third, it is positive that numerous approaches exist for proactively addressing ethical issues in the development and deployment of AI systems. Guidelines and various Ethics by Design frameworks have been established at the EU level, grounded in fundamental rights and ethical principles (Brey and Dainow 2024; Dainow and Brey 2021). These guidelines define key requirements for trustworthy AI, including human oversight, robustness and safety, privacy and data governance, transparency, diversity, non-discrimination, fairness, and accountability (EU Commission 8 April 2019). The AI Act promotes the ethical development and use of AI tools (EU Commission 2021). Within healthcare, the WHO guideline highlights several value-based design methods (World Health Organization 2021). Likewise, standards for documenting clinical studies have incorporated AI-specific guidelines, such as The STARD-AI Protocol (Sounderajah et al. 2021) and The CONSORT-AI Extension (Liu et al. 2020). These standards enhance transparency and reliability by clarifying quality requirements for data handling, model training, and validation. Additionally, the website FUTURE AI (2022) presents a comprehensive guideline dedicated to AI in healthcare.

Despite such promising frameworks, value-based approaches are not a panacea. For instance, value sensitive design, one of the most acknowledged design approaches, endorsed by WHO World Health Organization (2021), has had a limited impact on actual system development (Gerdes and Frandsen 2023). Developers often find such frameworks too abstract and prefer purely technical methods for handling bias and ensuring privacy (FATML 2024). Thus, the greatest challenge for value-based design approaches lies in orchestrating interdisciplinary collaboration among technicians, domain experts, and related stakeholders. Notably, it is essential to ensure the active involvement of clinical domain experts (Gerdes 2022).

In sum, AI's role in healthcare should not be defined by wishful thinking, anticipating future scenarios in which fully autonomous AI trumps human performance, especially not since such a scenario might lead to an actuarial turn in healthcare. Instead, we should concentrate on keeping humans in the loop and ensuring AI supports physician expertise rather than replacing it. While many have emphasized the need for clinicians to develop AI literacy (Aslam and Hoyle 2022; Misra et al. 2024), it is equally important to consider how core clinical skills, such as pattern recognition in radiology, can be preserved in the wake of AI-assisted clinical image interpretation and diagnostics.

4. Conclusion

This article has analyzed the challenges related to the impact of AI-assisted decision-making, focusing on patient autonomy in SDM and physician expertise. From these perspectives, opaque AI performance should not compromise justifiable decision-making and undermine patient and physician autonomy. Consequently, transparency is required, which can be partly achieved through post hoc explanations of AI decision inputs, whose non-exhaustiveness can be compensated for by demanding rigorous retrospective and prospective validation of AI models, following the RCT standard. Furthermore, within SDM, a frequently voiced claim is that patients' values and preferences should drive the ranking of treatment options, implying that AI-pre-ranked treatment recommendations threaten patient autonomy. However, this argument overlooks the fact that patient autonomy is also threatened by cognitive overload, which may hinder opportunities for informed choice. Rather than rejecting AI-supported ranking outright, it is more important to consider how to balance AI's potential to turn information overload into manageable information while respecting the patient's values and preferences. Moreover, focusing on preserving the physician's autonomy requires avoiding AI-driven deskilling and ensuring that physicians in the future will have practice-based training opportunities to cultivate the skills necessary for developing expertise. Ultimately, the objective should be to integrate AI in ways that enhance human agency, strengthening patients' ability to make informed choices and physicians' capacity to exercise sound clinical judgment.

Ethics declaration: Ethical clearance was not required for the research.

AI declaration: The author has used generative AI (ChatGPT, OpenAI) as a support tool for improving grammar and stylistic clarity.

References

Aristotle. (1934) Nicomachean Ethics. Rackham, H. (translated), Harvard University Press, William Heinemann Ltd, Cambridge, MA. London.

Aslam, Tariq M. and Hoyle, David C. (2022) "Translating the Machine: Skills that Human Clinicians Must Develop in the Era of Artificial Intelligence." Ophthalmology and Therapy, vol. 11, no. 1, pp. 69-80. doi:10.1007/s40123-021-00430-6.

Aurelia, Sauerbrei. (2023) "The Impact of Artificial Intelligence on the Person-Centred, Doctor-Patient Relationship: Some Problems and Solutions." BMC Medical Informatics and Decision Making, doi:10.1186/s12911-023-02162-y.

Beauchamp, T. L. and Childress, J. F. (2019) Principles of biomedical ethics (Eighth edition.), Oxford University Press., Oxford.

Bjerring, Jens Christian and Busch, Jacob. (2021) "Artificial Intelligence and Patient-Centered Decision-Making." Philosophy & Technology, vol. 34, no. 2, pp. 349-371. doi:10.1007/s13347-019-00391-6.

Brey, Philip and Dainow, Brandt. (2024) "Ethics by design for artificial intelligence." AI and Ethics, vol. 4, no. 4, pp. 1265-1277. doi:10.1007/s43681-023-00330-4.

Charles, C. et al. (1997) "Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango)." Soc Sci Med, vol. 44, no. 5, pp. 681-692. doi:10.1016/s0277-9536(96)00221-3.

Cheikh, Alexandre Ben et al. (2022) "How artificial intelligence improves radiological interpretation in suspected pulmonary embolism." European Radiology, vol. 32, no. 9, pp. 5831-5842. doi:10.1007/s00330-022-08645-2.

Dainow, Brandt and Brey, Philip. (2021) "Ethics By Design and Ethics of Use Approaches for Artificial Intelligence." general editor, European Commission. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf.

Debrabander, J. and Mertes, H. (2022) "Watson, autonomy and value flexibility: revisiting the debate." J Med Ethics, vol. 48, no. 12, pp. 1043-1047. doi:10.1136/medethics-2021-107513.

Dratsch, Thomas et al. (2023) "Automation Bias in Mammography: The Impact of Artificial Intelligence BI-RADS Suggestions on Reader Performance." Radiology, vol. 307, no. 4, p. e222176. doi:10.1148/radiol.222176.

Dreyfus, H. L., & Dreyfus, S. E. (1986). Mind over machine: The power of human intuition and expertise in the era of the computer, Free Press.,

Eikeland, Olav. (2008) The Ways of Aristotle – Aristotelian Phrónesis, Aristotelian Philosophy of Dialogue, and Action Research, Peter Lang.

Elhakim, Mohammad Talal. (2024) "Large-scale validation of artificial intelligence for breast cancer detection in Danish mammography screening." Syddansk Universitet.

EU Commission. (8 April 2019) "High-Level Expert Group on AI. Ethics guidelines for trustworthy AI - Shaping Europe's digital future." https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

EU Commission. (2021) "Proposal for a regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. EUR-Lex - 52021PC0206." https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELLAR:e0649735-a372-11eb-9585-01aa75ed71a1.

European Commission et al. (2013) European guidelines for quality assurance in breast cancer screening and diagnosis – Fourth edition, supplements, L. v Karsa et al., Publications Office,

FATML. "FATML -Fairness, Accountability, and Transparency in Machine Learning." https://www.fatml.org/. Accessed 0709 2024.

Freeman, Karoline et al. (2021) "Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy." Bmj, vol. 374, p. n1872. doi:10.1136/bmj.n1872.

FUTURE AI. "FUTURE-AI: Best practices for trustworthy AI in medicine." https://future-ai.eu/.

Gao, Y. et al. (2023) "Addressing the Challenge of Biomedical Data Inequality: An Artificial Intelligence Perspective." Annu Rev Biomed Data Sci, vol. 6, pp. 153-171. doi:10.1146/annurev-biodatasci-020722-020704.

Gerdes, Anne. (2022) "A participatory data-centric approach to AI Ethics by Design." Applied artificial intelligence, vol. 36, no. 1, doi:10.1080/08839514.2021.2009222.

Gerdes, Anne. (2024) "The role of explainability in AI-supported medical decision-making." Discov Artif Intell, vol. 4, no. 1, pp. 29-27. doi:10.1007/s44163-024-00119-2.

Gerdes, Anne and Frandsen, Tove Faber. (2023) "A systematic review of almost three decades of value sensitive design (VSD): what happened to the technical investigations?" Ethics and Information Technology, vol. 25, no. 2, p. 26. doi:10.1007/s10676-023-09700-2.

Goebel, Randy et al. (2018) "Explainable AI: The New 42?" Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), edited by A. Holzinger et al., vol. 11015, Springer International Publishing, pp. 295-303.

Hekler, Achim et al. (2019) "Superior skin cancer classification by the combination of human and artificial intelligence." European Journal of Cancer, vol. 120, pp. 114-121. doi:https://doi.org/10.1016/j.ejca.2019.07.019.

Holm, Søren and Ploug, Thomas. (2023) "Population preferences for AI system features across eight different decision-making contexts." PLOS ONE, vol. 18, no. 12, p. e0295277. doi:10.1371/journal.pone.0295277.

Kim, S. H. et al. (2025) "Automation bias in AI-assisted detection of cerebral aneurysms on time-of-flight MR angiography." Radiol Med, doi:10.1007/s11547-025-01964-6.

Lekadir, Karim et al. (2022) "Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts." Panel for the Future of Science and Technology EPRS, European Parliamentary Research Service. doi:10.2861/568473.

Linardatos, Pantelis et al. (2021) "Explainable AI: A Review of Machine Learning Interpretability Methods." Entropy, vol. 23, no. 1, p. 18. https://www.mdpi.com/1099-4300/23/1/18.

Liu, Xiaoxuan et al. (2020) "Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension." The Lancet Digital Health, vol. 2, no. 10, pp. e537-e548. doi:10.1016/S2589-7500(20)30218-1.

Lorenzini, G. et al. (2023) "Artificial intelligence and the doctor-patient relationship expanding the paradigm of shared decision making." Bioethics, vol. 37, no. 5, pp. 424-429. doi:10.1111/bioe.13158.

Lång, Kristina et al. (2023) "Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study." The Lancet Oncology, vol. 24, no. 8, pp. 936-944. doi:10.1016/S1470-2045(23)00298-X.

McDougall, Rosalind J. (2019) "Computer knows best? The need for value-flexibility in medical AI." Journal of Medical Ethics, vol. 45, no. 3, pp. 156-160. https://www-jstor-org.proxy1-bib.sdu.dk/stable/26884777.

Misra, Rohan et al. (2024) "How should we train clinicians for artificial intelligence in healthcare?" Future Healthcare Journal, vol. 11, no. 3, p. 100162. doi:https://doi.org/10.1016/j.fhj.2024.100162.

Nagendran, Myura et al. (2020) "Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies." BMJ (Clinical research ed.), vol. 368, p. 1. doi:http://dx.doi.org/10.1136/bmj.m689.

Najjar, R. (2023) "Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging." Diagnostics (Basel), vol. 13, no. 17, doi:10.3390/diagnostics13172760.

Obermeyer, Ziad et al. (2019) "Dissecting racial bias in an algorithm used to manage the health of populations." Science, vol. 366, no. 6464, pp. 447-453. doi:doi:10.1126/science.aax2342.

Ploug, T. et al. (2021) "Population Preferences for Performance and Explainability of Artificial Intelligence in Health Care: Choice-Based Conjoint Survey." J Med Internet Res, vol. 23, no. 12, p. e26611. doi:10.2196/26611.

Reverberi, Carlo et al. (2022) "Experimental evidence of effective human-AI collaboration in medical decision-making." Scientific Reports, vol. 12, no. 1, p. 14952. doi:https://doi.org/10.1038/s41598-022-18751-2.

Ribeiro, Marco Tulio et al. (2016) ""Why Should I Trust You?": Explaining the Predictions of Any Classifier." KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, pp. 1135–1144. doi:10.1145/2939672.2939778.

Sarter, Nadine B. (2001) "Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing." Human Factors, vol. 43, no. 4, p. 573. doi:https://doi.org/10.1518/001872001775870403.

Shapiro, Mark A. et al. (2023) "AI-Augmented Clinical Decision Support in a Patient-Centric Precision Oncology Registry." AI in Precision Oncology, vol. 1, no. 1, pp. 58-68. doi:10.1089/aipo.2023.0001.

Šķilters, Jurģis et al. (2024) "Towards A Human-AI Hybrid Medicine: Future Medicine — A Hybrid System Where AI Complements Instead of Replaces Humans." Proceedings of the Latvian Academy of Sciences, vol. 78, no. 4, pp. 233-238. doi:https://doi.org/10.2478/prolas-2024-0032.

Sounderajah, Viknesh et al. (2021) "Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol." BMJ Open, vol. 11, no. 6, p. e047709. doi:10.1136/bmjopen-2020-047709.

Thaler, Richard H. et al. (2010) "Choice Architecture ", https://ssrn.com/abstract=1583509 or http://dx.doi.org/10.2139/ssrn.1583509.

Thaler, Richard H. and Sunstein, Cass R. (2003) "Libertarian Paternalism." The American Economic Review, vol. 93, no. 2, pp. 175-179. http://www.jstor.org.proxy1-bib.sdu.dk:2048/stable/3132220.

Théberge, Isabelle et al. (2014) "Radiologist Interpretive Volume and Breast Cancer Screening Accuracy in a Canadian Organized Screening Program." JNCI: Journal of the National Cancer Institute, vol. 106, no. 3, doi:10.1093/jnci/djt461.

Tsai, Theodore L. et al. (2003) "Computer Decision Support as a Source of Interpretation Error: The Case of Electrocardiograms." Journal of the American Medical Informatics Association, vol. 10, no. 5, pp. 478-483. doi:10.1197/jamia.M1279.

Wong, D. J. et al. (2023) "Do Reader Characteristics Affect Diagnostic Efficacy in Screening Mammography? A Systematic Review." Clin Breast Cancer, vol. 23, no. 3, pp. e56-e67. doi:10.1016/j.clbc.2023.01.009.

World Health Organization. "Ethics and Governance of Artificial Intelligence for Health: WHO guidance." https://www.who.int/publications/i/item/9789240029200. Accessed 11-17-2022.

Wynants, L. et al. (2020) "Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal." Bmj, vol. 369, p. m1328. doi:10.1136/bmj.m1328.