Ethical Implications of Small Language Models (SLMs) in Healthcare Applications

DENECKE, Kerstin <http://orcid.org/0000-0001-6691-396X>, VAGLIANO, Iacopo <http://orcid.org/0000-0002-3066-9464>, HEWITT, Lantana <http://orcid.org/0009-0000-6459-8695>, AL TAMIMI, Abdel-Karim <http://orcid.org/0000-0003-2459-0298>, MEYSTRE, Stéphane <http://orcid.org/0000-0002-7632-9625> and TEODORO, Douglas <http://orcid.org/0000-0001-6238-4503>

Opening the Personal Gate between Technology and Health Care
M. Giacomini et al. (Eds.)

doi:10.3233/SHTI260236

Ethical Implications of Small Language Models (SLMs) in Healthcare Applications

Kerstin DENECKE[a,1], Iacopo VAGLIANO[b,c], Lantana HEWITT[d], Abdel-Karim AL-TAMIMI[d,e], Stéphane MEYSTRE[f], Douglas TEODORO[g]
[a] *Bern University of Applied Sciences, Bern, Switzerland*
[b] *Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands*
[c] *Amsterdam Public Health Research Institute, Amsterdam, the Netherlands*
[d] *Sheffield Hallam University, Sheffield, UK*
[e] *Yarmouk University, Irbid, Jordan*
[f] *MeDiTech Institute, University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Lugano, Switzerland*
[g] *University of Geneva, Geneva, Switzerland*
ORCiD ID: 0000-0001-6691-396X (KD), 0000-0002-3066-9464 (IV), 0009-0000-6459-8695 (LH), 0000-0003-2459-0298 (AA), 0000-0002-7632-9625 (SM), 0000-0001-6238-4503 (DT)

Abstract. Although Small Language Models (SLMs) show promise for healthcare applications, their use also introduces ethical risks that must be evaluated to develop appropriate mitigation strategies. Within a group activity during the SLM4Health workshop with 13 participants, we assessed ethical risks using the Digital Ethics Canvas along five lenses (welfare, autonomy, privacy, fairness, and sustainability). The participants considered two SLMs application use-cases: clinical documentation and clinical decision support. For SLMs in clinical documentation, identified risks included documentation errors from hallucinations or omissions leading to potential medical mistakes, social and data biases that can affect underrepresented groups, and a potential erosion of clinicians' skills. For clinical decision support, challenges included biases harming vulnerable groups, unequal data quality exacerbating healthcare disparities, and privacy risks due to exposure of sensitive data. In both use cases, SLMs were noted to have a smaller environmental footprint than large models, though they still require energy and resources. Participants recognized that SLMs enhance efficiency in clinical documentation and decision support but introduce ethical challenges concerning patient welfare, fairness, autonomy, and privacy.

Keywords. Ethics, small language model, large language model, ethical AI

1. Introduction

Large language models (LLMs) showed great potential for healthcare applications, but there is growing concern regarding limited resources, data privacy, and ethics [1]. Small language models (SLMs) represent an alternative which can better fit to resource-constrained healthcare settings and yet remain clinically viable. SLMs are artificial

[1] Corresponding Author: Kerstin Denecke, Institute Patient-centered Digital Health, Bern University of Applied Sciences, Quellgasse 21, 2502 Biel, Switzerland, kerstin.denecke@bfh.ch

intelligence (AI) models designed to process and generate natural language. They are similar in architecture to LLMs but with reduced complexity and computational demands. When trained on high-quality, domain-specific data, SLMs can match or outperform general-purpose LLMs. Specialized SLMs have proven competitive with models like GPT-4 in diverse tasks such as question-answering, information extraction, and de-identification of clinical documents [2–5]. While they offer clear advantages in terms of computational efficiency and decentralised deployment, their widespread clinical adoption is constrained by tangible risks. Foremost are patient and service outcomes: insufficient validation may limit an SLM's efficacy and reliability, potentially causing diagnostic errors that compromising patient safety [1]. Furthermore, their integration into clinical workflows necessitates rigorous ethical scrutiny. This includes addressing inherent biases [7] that could lead to disparities in care and ensuring data privacy [1]. It also requires safeguarding long-term environmental sustainability [6], because reduced deployment costs may incentivize excessive, widespread use. The deployment of SLMs in healthcare must be governed by principles that prioritize patient welfare, equity, and transparency to foster trust and efficacy in medical practice. This work aims to identify and evaluate the ethical risks associated with SLMs in healthcare applications. This will provide a foundation for developing strategies to mitigate these risks and ensure the safe and responsible deployment of SLMs.

2. Methods

In the workshop Small Language Models for Health (SLM4Health) held in conjunction with the AI in Medicine (AIME) 2025 conference, we conducted a group activity to reflect on the ethical implications of the use of SLMs in health applications. All participants were invited and split into 3 groups, 1 of which was online. We applied the Digital Ethics Canvas [8] to systematically identify and assess the ethical risks associated with SLMs in health applications (Figure 1).

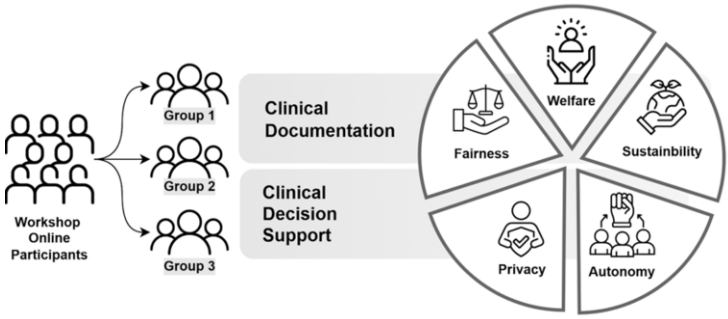

Figure 1. Methodology: Workshop participants were split into three groups, considered two applications along 5 ethical dimensions

This structured, visual framework provides guidance to examine potential ethical implications from multiple perspectives. It organises analysis around five ethical lenses: welfare, autonomy, privacy, fairness, and sustainability. All five lenses together address the potential benefits and possible harms of digital applications. *Welfare* focuses on the well-being and safety of individuals, groups, and society affected by a solution. It examines whether the technology or system promotes human welfare, minimizes harm,

and protects users and others from negative outcomes. *Autonomy* examines whether users can make informed and voluntary choices about how they interact with and are affected by the technology. *Privacy* addresses issues related to data protection and informed consent. *Fairness* considers equity of access, algorithmic bias and discrimination. *Sustainability* encompasses environmental and social impacts, such as resource use and labour conditions in digital infrastructures. The workshop participants received a brief introduction to the canvas methodology. They were asked to consider either SLMs for the task of clinical documentation or for clinical decision support and characterize the application they are considering. Then, they were asked to answer the guiding questions on the canvas and collect their answers on a Miro Board. This group reflection was scheduled for 1 hour.

3. Results

The workshop was attended by 13 participants (5 PhD students, 6 senior researchers, 2 at another career stage), divided into three groups (6 participants, 5 participants, 2 participants online). Most participants had a background in Computer Sciences (n = 10), followed by Engineering (n = 2) and Medicine (n = 1). Participants were based in Europe (n = 11), Africa (n=1) or Asia (n = 1). Two groups assessed SLMs for use in clinical documentation. The applications included generating reports, summarising information, and translating medical information into patient-friendly language. Reported benefits included improved efficiency, faster data entry, automated structured summaries and summaries adapted for patients. A third group examined SLMs that were integrated into clinical decision support tools for healthcare professionals. These systems were reported to save time, improve information retrieval and support more informed, personalised decisions using comprehensive patient records. Other advantages included access to the latest medical knowledge, enhanced clinician–patient communication and a more intuitive user experience.

Table 1 outlines how core ethical lenses (welfare, fairness, autonomy, privacy, and sustainability) manifest in the deployment of SLMs within healthcare. Compared to larger models, they are particularly effective in improving efficiency, consistency and sustainability. Their smaller size enables local deployment, thereby reducing environmental impact and enhancing data privacy. Nevertheless, both applications raise similar ethical and operational concerns when viewed through the five ethical lenses. However, the specific issues vary depending on the application.

Hallucinations, omissions and misinterpretations can undermine patient welfare by leading to errors in documentation or unsafe clinical recommendations. Overreliance on model outputs amplifies these risks. Documentation errors may propagate through subsequent care, whereas errors in decision support may directly affect diagnosis or treatment. Fairness concerns span both applications because social and data biases can affect underrepresented populations or rare conditions disproportionately. Disparities may widen further when data quality or AI literacy varies across healthcare settings.

Concerns about autonomy affect both clinicians and patients. Overreliance on SLM outputs may reduce clinicians' autonomy and diminish their skills in documentation and clinical interpretation. For patients, automated summaries or recommendations may indirectly impact autonomy by altering communication and trust dynamics. Even with local deployment, privacy risks remain, because sensitive data processed during training

and inference may still be exposed. This underscores the need for strong governance, access controls, and on-premises processing in both applications. Finally, although SLMs are generally more sustainable than larger models, they still require significant amounts of energy, storage space and human oversight. If SLM are implemented thoughtfully, with clear accountability and transparency, efficiency gains can promote responsible resource use. Ensuring that SLMs support rather than undermine ethical clinical practice requires consistent governance and bias mitigation. It also requires safeguards tailored to each use case, whether for documentation or decision support, while maintaining shared standards for fairness, autonomy and patient welfare.

Table 1. Ethical dimensions with guiding questions, associated risks, and mitigation strategies of SLM–based systems in clinical decision support and documentation.

Ethical lens	Clinical documentation	Clinical decision support
Welfare (Can the solution be used in harmful ways? What kind of impacts can errors from the solution have? What type of protection does the solution have against attacks or misuse?)	Hallucinations, omissions, or misrepresentations can lead to documentation errors, cascading into diagnostic or medical errors. Patient-facing summaries may cause distress or misunderstanding.	Biased training data may harm vulnerable or minority populations and those with rare conditions. Overemphasis on the main complaint or high recall may lead to missed issues or cause anxiety. Overreliance or replacing clinicians could compromise patient safety.
Fairness (How accessible is the solution? What kinds of biases may affect the results? Can the outcomes of the solution be different for different users or groups? Could the solution contribute to discrimination against people or groups?)	SLMs are very accessible, introduction of social biases, hallucinations can be discriminative, can lead to overreliance on automatic documentation or non-reflective use	Biases in data can disadvantage minorities or rare conditions; unequal AI literacy among clinicians affects outcomes; limited access in low-literacy or low-resource settings.
Autonomy (Can users understand how the solution works and what its limits are? Are users able to make choices (e.g. consent, settings) in their use of the solution and how? How does the solution affect user autonomy and agency?)	User control and editability can be limited; explainability has to be actively ensured; choices possible after training on responsible use; clinician–patient interaction can be impacted.	Lack of transparency limits understanding; overreliance may erode clinicians' documentation skills; dependency on tool availability; patients may rely excessively on automated summaries.
Privacy (What data does the solution collect? Is it collecting personal or sensitive data Who has access to the data? How is the data protected? Could the solution disclose / be used to disclose private information?)	On-premises deployment supports privacy; strict access control and data governance; anonymization and privacy-by-design practices can be considered.	Sensitive data exposure during training or transfer; risk of data leakage if models move across institutions; unauthorized access to patient data
Sustainability (What is the carbon footprint of the solution? What types of resources does it consume (e.g. water) -and produce (e.g. waste)? What type of human labour is involved?)	SLMs are more energy-efficient than LLMs; local deployment reduces long-term environmental cost; continuous efficiency improvements.	SLMs have a smaller footprint than LLMs; more labelled data is required for fine-tuning than for LLMs

4. Discussion and conclusion

As stated in the introduction, with this initial effort we aimed at identifying ethical risks of SLMs in healthcare applications as a first step to devise mitigation strategies for

practical applications. Various risks were identified in each category of the Digital Ethics Canvas with a focus on clinical documentation and clinical decision support. As shown in Table 1, these risks range from potentially severe harms such as hallucinations or omissions in the SLM output causing medical errors to broader concerns. These include limited transparency in some SLM architectures, which may erode trust, and the substantial energy and data storage demands of model training (even if significantly lower than for LLMs). In future efforts, the risks identified will be investigated further and mitigation strategies will be proposed. Limitations of this study included the overly broad focus of some participants expanding beyond risks, and the relatively small number of participants, even if they represented a good diversity of backgrounds and experience levels. We did not apply any predefined eligibility criteria when recruiting participants, which may have introduced selection bias. A more in-depth analysis with additional workshops, including also a person with background in ethics, may be useful to complete these initial results and to formulate also mitigation strategies.

In conclusion, the workshop highlighted the potential of SLMs to enhance clinical documentation, clinical reasoning, and patient communication. Their responsible adoption requires strong safeguards for safety, privacy, fairness, and oversight. Participants agreed that SLMs should augment, not replace, clinical expertise, regardless of whether the task is administrative or clinical. This implies human-in-the-loop validation, clear accountability, well-defined consent procedures, and secure data governance. Progress will depend on targeted pilot studies, regulatory-aligned governance structures and comprehensive clinician training. Equity and patient engagement should be integrated throughout deployment. Finally, several areas remain underexplored, including explainability, regulatory and legal considerations, and governance of synthetic data.

References

[1] Garg M, Raza S, Rayana S, Liu X, Sohn S. The Rise of Small Language Models in Healthcare: A Comprehensive Survey. arXiv; 2025. doi: 10.48550/arXiv.2504.17119

[2] Griot M, Hemptinne C, Vanderdonckt J, Yuksel D. Impact of high-quality, mixed-domain data on the performance of medical language models. J Am Med Inform Assoc 2024 Sept 1;31(9):1875–1883. doi: 10.1093/jamia/ocae120

[3] Wu C, et al. PMC-LLaMA: toward building open-source language models for medicine. J Am Med Inform Assoc 2024 Sept 1;31(9):1833–1843. doi: 10.1093/jamia/ocae045

[4] Wiest IC, et al. Deidentifying Medical Documents with Local, Privacy-Preserving Large Language Models: The LLM-Anonymizer. NEJM AI 2025 Mar 27;2(4). doi: 10.1056/AIdbp2400537

[5] Labrak Y, et al. BioMistral: A Collection of Open-Source Pretrained Large Language Models for Medical Domains. [object Object]; 2024; doi: 10.48550/ARXIV.2402.10373

[6] Luccioni AS, Strubell E, Crawford K. From Efficiency Gains to Rebound Effects: The Problem of Jevons' Paradox in AI's Polarized Environmental Debate. Proc 2025 ACM Conf Fairness Account Transpar 2025. p. 76–88. doi: 10.1145/3715275.3732007

[7] Gourabathina A, Gerych W, Pan E, Ghassemi M. The Medium is the Message: How Non-Clinical Information Shapes Clinical Decisions in LLMs. Proc 2025 ACM Conf Fairness Account Transpar Athens Greece: ACM; 2025. p. 1805–1828. doi: 10.1145/3715275.3732121

[8] Hardebolle C, Macko V, Ramachandran V, Holzer A, Jermann P. Digital Ethics Canvas: A Guide For Ethical Risk Assessment And Mitigation In The Digital Domain. European Society for Engineering Education (SEFI); 2023; doi: 10.21427/9WA5-ZY95