Evaluating Algorithm Bias in AI-Driven Clinical Decision Tools: Implications for Nursing

Judgment and Patient Safety in Low-Resource Hospitals

Student Name

Institution

Instructor

Course Number

Date

Evaluating Algorithm Bias in AI-Driven Clinical Decision Tools: Implications for Nursing Judgment and Patient Safety in Low-Resource Hospitals

The incorporation of artificial intelligence (AI) into the medical field has been of potential interest in the last few years, and this has been in the form of AI-based clinical decision support systems (AI-CDSS) which have been touted to enhance diagnostic accuracy, workflow optimization, and staffing issues. However, even with these potential advantages, it has been demonstrated that the bias of algorithms installed in these tools can compromise the judgment of nurses and put the lives of patients in danger. The issue is especially intense in the hospitals with limited resources because structural inequities, insufficient personnel, and resource constraints increase risks.

Prejudice during AI-CDSS may be demonstrated via data collection, feature selection, model training, and deployment and may result in disparate results between patient subgroups. The patients represented in training data insufficiently (such as ethnic minorities or low-income groups) may be either underdiagnosed or misidentified, which can raise the likelihood of receiving delayed or inadequate treatment (Zink et al., 2024). These differences are especially acute in low-resource hospitals where nurses can turn to the AI advice significantly because specialists are inaccessible.

Nursing care plans generated by AI can also have an algorithmic bias. Recent simulation studies have discovered that care plans generated by a large language AI model depend on patient demographics and clinical quality, where care plans generated by socially advantaged groups receive lower safety ratings than care plans generated by disadvantaged groups (Baig et al., 2024). This illustrates the fact that AI tools can unknowingly introduce a few stereotypes or assumptions, and this may compromise patient safety.

The use of AI tools that are biased by nurses is likely to undermine clinical judgment. In places where there is a high-level of staffing shortage, excessive use of AI results can reduce critical thinking, clinical intuition, and discerning subtle elements that are absent in data, such as psychosocial or social determinant considerations (Baig & Yadegaridehkordi, 2024). Structural power discrepancies inherent to healthcare systems make the problem more severe, and nurses alone cannot resolve it, but the system needs comprehensive changes to protect equity and safety.

Qualitative studies provide evidence to these issues. The ethical issues of AI-based resource allocation among hospital staff members, such as nurses, have been raised, especially in facilities with limited resources. Among the respondents, the systematic downsides of unfair AI models were noted, which may result in further inequities by disadvantaging some categories of patients (d'Elia et al., 2022). Such bias may result in the wrong triage decision and the safety of patients may be compromised in low-resource hospitals.

However, AI-CDSS can be made a valuable contribution in case of bias reduction. Such best-practice principles as the application of varied and representative training data, the performance of models in demographic subgroups, involvement of clinicians in the design of models, and transparency and explainability of AI outputs can be mentioned (Markus, Kors, & Rijnbeek, 2021). These measures can help AI to support, and not to substitute nursing judgment without the risk of serious injuries being limited.

In conclusion, AI-based clinical decision-making tools are a major threat to nursing decision-making and patient safety, particularly in low-resource hospitals due to the algorithmic bias thereof. Such bias, in its unmitigated form, may increase health disparities, lower the quality of care, and cause mistrust in AI-assisted clinical decisions. Nevertheless, AI-CDSS can empower nurses to provide safe and fair care with the appropriate precautions such as

representative datasets, on-going observation, the participation of clinicians, and clear reporting. Finally, AI must support, rather than substitute, human judgment so that technological innovation can be useful to every patient irrespective of their background and context.

References

Baig, M. I., & Yadegaridehkordi, E. (2024). *ChatGPT in higher education: A systematic literature review and research challenges*. International Journal of Educational Research, 127, 102411. https://doi.org/10.1016/j.ijer.2024.102411

Baig, M. M., Hobson, C., GholamHosseini, H., Ullah, E., & Afifi, S. (2024). *Generative AI in improving personalized patient care plans: Opportunities and barriers towards its wider adoption*. Applied Sciences, 14(23), 10899. https://doi.org/10.3390/app142310899

d'Elia, A., Gabbay, M., Rodgers, S., Kierans, C., Jones, E., Durrani, I., … & Frith, L. (2022). *Artificial intelligence and health inequities in primary care: A systematic scoping review and framework*. Family Medicine and Community Health, 10(Suppl 1), e001670. https://doi.org/10.1136/fmch-2022-001670

Markus, A. F., Kors, J. A., & Rijnbeek, P. R. (2021). *The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies*. Journal of Biomedical Informatics, 113, 103655. https://doi.org/10.1016/j.jbi.2020.103655