


Please cite the Published Version

O'Sullivan, Christopher, Gaddum, Clare and Lee, Amanda J  (2024) Use of AI to enhance written information in paediatric settings - stochastic parrot or clinical tool? Evidence-Based Nursing. ISSN 1367-6539

DOI: <https://doi.org/10.1136/ebnurs-2024-104164>

Publisher: BMJ Publishing Group

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/636189/>

Usage rights:  In Copyright

Additional Information: This article has been accepted for publication in Evidence-Based Nursing, 2024, following peer review, and the Version of Record can be accessed online at <http://dx.doi.org/10.1136/ebnurs-2024-104164>. © Authors (or their employer(s))

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Category: Nursing issues

Study type: Quantitative study - other

Author's declarative title: Use of AI to enhance written information in paediatric settings – stochastic parrot or clinical tool?

Commentary on: Brewster RCL, Gonzalez P, Khazanchi R, et al. (2024) Performance of ChatGPT and Google Translate for Pediatric Discharge Instruction Translation. *Pediatrics*. 2024;154(1)

Commentary

Implications for practice and research

- Artificial Intelligence (AI) has significant potential to impact healthcare, but further research is required to evaluate validity of translation for communication.
- Enhanced clinician communication is needed for non-native speaking patient groups, and AI translations may be useful. However, content must reflect true clinical 'meaning'.

Context

Today's world is increasingly augmented by AI and use and application of natural language process-based translation services could enhance clinical communications. They have the potential to support confidential and cost-effective communication mechanisms for non-native language patients. This study¹ evaluates quality, cultural sensitivity, terminology, context, consistency and risks associated with use of AI translators for Spanish, Portuguese and Haitian paediatric patients, illustrating shortcomings in quality, accuracy and patient preference, thus substantiating the theoretical 'stochastic parrot' where AI generated, 'pattern' processing of language can pose unprecedented risk²

Methods

A total 180 pieces of text aimed towards patient communication were drawn from medical notes and translated into Spanish, Portuguese and Creole using three methods: 1) Professional Translators, 2) Google Translate (GT), and 3) A Generative Pre-trained Transformer (ChatGPT).

Translated text was evaluated by 8 (paid) 'native speaking' medics in practice (with no translator qualifications). They completed 5-point Likert assessment rubrics to validate their subjective analysis of text adequacy (original preserved text), meaning (reflects original intent), fluency

(syntax & grammar), and severity (potential 'harm' caused through error / lack of context). Medics present their opinions of the translations offered by an interpreter, GT and Chat GPT translated texts.

Findings

The study¹ revealed GT and ChatGPT offered similar scores in adequacy, meaning, and fluency when compared with professional Spanish / Portuguese translators, yet Haitian Creole AI produced a higher rate of 'clinically significant' errors with potential to harm when compared to professional translation methods. Some clinicians preferred AI-generated translations in Spanish and Portuguese over professional translations, but not for Haitian Creole, revealing a disparity in translation validity between common and lesser spoken languages.

Commentary

The study¹ substantiates evidence that AI could provide timely and cost effective translation services, to support equitable access to clinical healthcare^{3,4}. It highlights inconsistencies in accuracy and efficacy across lesser spoken languages, potentially resulting in clinical errors and compromised patient safety^{5,6}.

Algorithmic pattern modelling is a stochastic methodology, thus, translations potentially lack true meaning². Authors of this study¹ advocate clinician caution, especially in lesser spoken languages, suggesting AI translation tools must augment rather than replace, human translators³. As clinical miscommunication can lead to adverse patient outcomes⁵, AI tools are no substitute for human expertise in clinical environments⁴.

Over-reliance on AI without adequate regulation or human oversight could worsen healthcare disparities⁶. Should AI tools be used without rigorous accuracy evaluation, clinicians may disproportionately affect marginalised 'lesser spoken language' communities. Large language models are based on pattern processing, meaning they reflect biases present across large language processors and training datasets². Thus, clinical application may reinforce existing inequalities. It is, therefore, crucial for 'human' translators' involvement to safeguard against error³. Despite its' potential, AI must be carefully evaluated³. Clinicians must establish standards for quality and accuracy of translations², especially for 'lesser spoken' languages^{1,5}.

We need robust research and regulatory frameworks for responsible, unbiased use of AI. Expanding linguistic diversity in translation datasets is key to performance across different languages, to enhance quality of clinical care³.

AI translation models must also reflect continuous evolution of language and culture. Clinical AI translations must capture all nuances of 'meaning' to avoid becoming stochastic parrots of prose². Once risks are mitigated and quality is improved, AI potentially offers a truly inclusive tool in global healthcare⁶.

Brewster's study¹ is one pioneering step to improved clinical AI translations, and a decolonized medical communication with enhanced accessibility to all^{3,6}.

References

- 1 Brewster, R. et al. (2024) 'Performance of ChatGPT and Google Translate for Paediatric Discharge Instruction Translation', *Pediatrics*, 154(1).
- 2 Bender, E. et al. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . (FAccT '21). Association for Computing Machinery, New York, 610–623.
- 3 Hasanah, L.N. (2023) 'Leveraging AI to Address Language Barriers in Healthcare: Ethical Considerations and Implementation Strategies', *Journal Humanities and Applied Science Research*, 6(4),56–64.
- 4 Li, J. et al. (2023) 'ChatGPT in Healthcare: A Taxonomy and Systematic Review', *ChatGPT in healthcare: Computer Methods and Programs in Biomedicine*, (245).
- 5 Liu, C., et al. (2023) 'Med-UniC: Unifying Cross-Lingual Medical Vision-Language Pre-Training by Diminishing Bias', *Neur IPS 2023*.
- 6 Sallam, M. (2023) ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare*. 11(6):887

Commentator details

ORDER of authorship: O'Sullivan, C., Gaddum, C., Lee, A.

Name: Dr Amanda Lee

Affiliation: Manchester Metropolitan University

Correspondence address: Birley Campus, Manchester, M15 6GR

Email: Amanda.lee@mmu.ac.uk

Competing interests

Authors have no competing interests in this commentary.

Author 1

O'Sullivan, Christopher

Manchester Metropolitan University, Faculty of Health Sciences,
Manchester, UK

Author 2


Gaddum, Clare

Manchester Metropolitan University, Faculty of Health Sciences,
Manchester, UK
Author 1
OSullivan, Christopher

Author 3

Lee, Amanda J

amanda.lee@mmu.ac.uk

 [0000-0003-1378-3123](https://orcid.org/0000-0003-1378-3123) ✓

Manchester Metropolitan University, Faculty of Health Sciences,
Manchester, M15 6BH, UK
