

AI for Equity

Improving Translation in Obstetrics

Dr Alexander Lomas – Royal Cornwall Hospital

Dr Malcolm Broom – Glasgow Royal Infirmary/Princess Royal Maternity

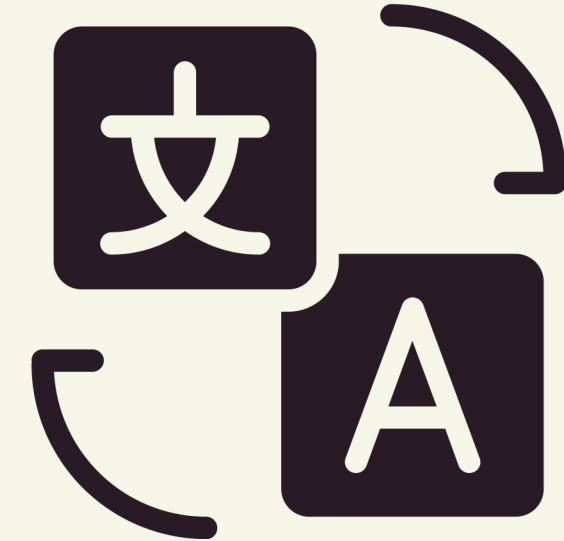


Minority ethnic groups in the UK have poorer perinatal outcomes compared to equivalent white parturient groups (1).

Language is a contributor to this.

Machine translation tools exist, but to date have been inaccurate and unsuitable for medical communication (2). Despite this, they are regularly used in healthcare.

LLMs are rapidly developing and capable of translation tasks, but can they be used for patient communication?



METHODS

GPT-4 (3) was used to translate both a patient information leaflet on epidurals and generate it's own patient information from a reference website.

Five languages selected based on worldwide prevalence: Bengali, Hindi, Portuguese, Spanish, Mandarin.

Freelance interpreters asked to grade the translations using a three-domain Likert Scale (4) and to translate the version created by GPT-4.

	1	→	5
Fluency and Readability	Poorly Understandable		Completely Understandable
Meaning	Overwhelming misleading information		No misleading information
Adequacy	0% information conveyed		100% information conveyed

Figure 1 – Simplified version of the Likert Scale sent to interpreters

LARGE LANGUAGE MODEL (LLM)

An Artificial Intelligence (AI) tool capable of generating text by learning from extensive datasets of written language

RESULTS

- Information translated by GPT-4 was generally accurate, with little to no misleading information or omissions.
- Global rating of quality was high. There were some grammatical errors which affected readability but the text remained understandable.
- Information generated by GPT-4 from a reference was overall more fluent and readable. It contained appropriate information on epidurals and risks. Reverse translating with the model was an accurate way of determining the content.



Figure 2 – Mean Likert scale values by domain for all interpreters and languages combined

INTERPRETERS' COMMENTS

"The translation is perfect. It reads smoothly and sounds natural in Hindi."

"A high level of professionalism, reflecting an understanding of medical terminology and the ability to convey complex medical procedures in a clear and accurate manner."

"Balance between technical accuracy and readability is commendable."

1: MBRRACE-UK. Maternal, Newborn and Infant Clinical Outcome Review Programme. (2021).

2: Moberly, T. Doctors are cautioned against using Google Translate in consultations. BMJ k4546 (2018) doi:10.1136/bmj.k4546.

3: OpenAI. ChatGPT. <https://chat.openai.com> (2023)

4: Khanna, R. R. et al. Performance of an online translation tool when applied to patient educational material. Journal of Hospital Medicine 6, 519-525 (2011).

Acknowledgements

This project was funded by the Eric Angel Scholarship. GPT-4 is developed by OpenAI

