Letter to the Editor

# Who Reviews the Rewrite? Separating Patient-Initiated and Institutionally Curated Large Language Model Simplification

Aakash Dave, BS

Department of Psychology, Center for Bioethics and Social Justice, Institute for Quantitative Health Science and Engineering, Michigan State University, East Lansing, MI, United States

**Corresponding Author:**

Aakash Dave, BS
Department of Psychology, Center for Bioethics and Social Justice
Institute for Quantitative Health Science and Engineering, Michigan State University
775 Woodlot Dr
East Lansing, MI
United States
Phone: 1 5863565111
Email: daveaaka@msu.edu

**Related Articles:**
Comment on: https://ai.jmir.org/2026/1/e77149
Comment in: https://ai.jmir.org/2026/1/e107588

***JMIR AI 2026;5:e107277;*** doi: 10.2196/107277

**Keywords:** health information; patient education material; readability; large language models; LLMs; AI

Miftaroski and colleagues [1] demonstrate that several large language models (LLMs) can reduce the linguistic complexity of German patient education materials; the interpretation of these findings is, however, complicated by whether the inquiry was patient initiated or institutionally curated.

The study approximates lay use by entering medical text into various publicly accessible LLMs with zero-shot prompts and no model parameter tuning. However, they conclude that rephrased material requires expert review to preserve medical accuracy and completeness. Thus, there are two distinct workflows that are not interchangeable. Expert review is a plausible safeguard when an institution revises material before publication, but it cannot be assumed when patients themselves seek an immediate explanation of information they do not understand.

This distinction is further complicated by the reliability analyses, in which 3 reviewers with medical informatics backgrounds achieved a Fleiss $\kappa$ of 0.264% and an agreement of 54.6% when identifying false or decontextualized statements [1]. The authors attribute this low agreement, at least in part, to a lack of deep domain expertise, with the source being incompletely characterized. It seems that the panel had heterogeneous academic backgrounds, and the study fails to report important variables: prespecified rating rubric, reviewer calibration, domain-specific clinical expertise, reviewer-level ratings, or an adjudication process. The authors also probe accuracy, clarity, and plausibility, but it is unclear whether these were evaluated as distinct constructs or incorporated into a single judgment. I argue that the observed $\kappa$ value cannot distinguish ambiguity within the outputs from heterogeneity in reviewer knowledge or decision thresholds, given that the Guidelines for Reporting Reliability and Agreement Studies highlight the importance of rater selection, training, and the measurement process [2].

These limitations are consequential, because readability and clinical adequacy are not the same. For example, Flesch Reading Ease and Wiener Sachtextformel scores quantify features of words and sentences [1], but they do not establish whether the patient is actually informed by or can act on the resultant information. Moreover, the Patient Education Materials Assessment Tool separately evaluates understandability and actionability [3], while direct testing can determine whether readers correctly interpret and apply health information [4]. Thus, improved readability scores do not directly support the inference that unreviewed LLM simplification is directly linked to improvement in patient understanding [3,4].

In future studies within this realm, the authors should prespecify the intended workflow. Materials that are intended for patients ideally should be assessed for all, if not some of the following: comprehension, error recognition, intended actions, trust calibration, and performance across health literacy levels. Next, institutionally generated patient education materials should also ascertain if domain experts can reliably identify clinically meaningful omissions or

distortions and whether the additional review process actually provides a measurable advantage over conventional human authorship.

Given the aforementioned concerns, what we can conclude from this paper is that LLMs can alter linguistic complexity, but not that this translates to the production of clinically safe patient communication. In brief, medical text may become easier to read without becoming safer to understand per se.

## Acknowledgments

The author used ChatGPT (OpenAI; GPT 5.5) to assist with brainstorming, relevant literature parsing, and syntactical formatting. The author independently verified all sources, revised the manuscript, and takes full responsibility for its content.

## Funding

None declared.

## Conflicts of Interest

None declared.

## References

1. Miftaroski A, Zowalla R, Wiesner M, Pobiruchin M. Leveraging large language models to improve the readability of German online medical texts: evaluation study. JMIR AI. Jan 23, 2026;5:e77149. [doi: 10.2196/77149] [Medline: 41575871]
2. Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. Jan 2011;64(1):96-106. [doi: 10.1016/j.jclinepi.2010.03.002] [Medline: 21130355]
3. Shoemaker SJ, Wolf MS, Brach C. Development of the Patient Education Materials Assessment Tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. Patient Educ Couns. Sep 2014;96(3):395-403. [doi: 10.1016/j.pec.2014.05.027] [Medline: 24973195]
4. Szabó P, Bíró É, Kósa K. Readability and comprehension of printed patient education materials. Front Public Health. 2021;9:725840. [doi: 10.3389/fpubh.2021.725840] [Medline: 34917569]

## Abbreviations

**LLM:** large language model

*Edited by Ivan Steenstra; This is a non–peer-reviewed article; submitted 16.Jul.2026; accepted 29.Jul.2026; published 31.Aug.2026*

*Please cite as:*
*Dave A*
*Who Reviews the Rewrite? Separating Patient-Initiated and Institutionally Curated Large Language Model Simplification*
*JMIR AI 2026;5:e107277*
*URL: https://ai.jmir.org/2026/1/e107277*
*doi: 10.2196/107277*