LETTER OF REVIEWERS

Reviewer A:
Recommendation: Revisions Required

Relevance: Very high

Novelty: High

Presentation and writing: High

Comments for authors:
The manuscript is timely and relevant. However, we provide several comments that should be considered prior to publication.
1. We suggest strengthening the “magnitude-related evidence” and avoiding potentially weak assertions. The statement “Millions of users interact with these large language models (LLMs)” may appear plausible, but it would benefit from being supported by a source that documents usage volume. Alternatively, it could be qualified (e.g., “a large and growing number of users”) to avoid a vulnerable point in the opening.

2. The sequence of the initial paragraphs is coherent; however, the progression could gain clarity by first delineating the overarching regulatory argument (governance, standards, data protection) and then presenting the subarguments as specific instances (data sensitivity, extraterritoriality and infrastructure, opacity of foundation models). This structure would make the guiding thread more explicit for non-specialist readers.

3. In the passage stating that “this law aims to serve as a counterpart to the GDPR… or the HIPAA,” we suggest a more nuanced formulation. The GDPR and HIPAA differ substantially in scope, legal nature, and applicability (general data protection versus a sector-specific healthcare framework). A reformulation would avoid an overly direct equivalence.

Interacciones seeks greater transparency in the review process and to provide credit to reviewers. If the editors decide to accept the manuscript, **would you like your name to appear as a reviewer of the article?**

No

RESPONSE LETTER

1. We suggest strengthening the "magnitude-related evidence" and avoiding potentially weak assertions. The statement "Millions of users interact with these large language models (LLMs)" may appear plausible, but it would benefit from being supported by a source that documents usage volume. Alternatively, it could be qualified (e.g., "a large and growing number of users") to avoid a vulnerable point in the opening.
Reply: We replaced the unsourced statement with a sourced usage estimate:
"These large language models (LLMs) are being used at massive scale; for example, ChatGPT alone has been reported to have hundreds of millions of weekly active users (Chatterji et al., 2025), and the users interact with these LLMs to receive guidance related to anxiety, depression, or crisis situations, marking an unprecedented shift in the digital health ecosystem (Ayers et al., 2023)."

2. The sequence of the initial paragraphs is coherent; however, the progression could gain clarity by first delineating the overarching regulatory argument (governance, standards, data protection) and then presenting the subarguments as specific instances (data sensitivity, extraterritoriality and infrastructure, opacity of foundation models). This structure would make the guiding thread more explicit for non-specialist readers.
Reply: We have modified the first four paragraphs to improve readability:
"*The accelerated adoption of generative artificial intelligence (AI) models, such as ChatGPT and Gemini, as well as other conversational agents, has transformed how people worldwide seek mental health information and support (Thirunavukarasu et al., 2023). These large language models (LLMs) are being used at massive scale; for example, ChatGPT alone has been reported to have hundreds of millions of weekly active users (Chatterji et al., 2025). Users interact with these systems to receive guidance related to anxiety, depression, or crisis situations, marking an unprecedented shift in the digital health ecosystem (Ayers et al., 2023). However, while generative AI promises to expand access to physical and mental health resources, it also introduces ethical and regulatory risks that remain insufficiently addressed (Meskó & Topol, 2023), particularly in low- and middle-income regions such as Latin America. In these settings, AI models developed in high-income countries are widely deployed without necessarily assessing the potential risks of bias that this entails (Hussain et al., 2025).*
In low- and middle-income countries (LMICs), the governance architecture for health-related generative artificial intelligence, encompassing standards, accountability, transparency, and enforceable data protection, lags behind its real-world implementation. We observe that AI-based systems are increasingly integrated into daily life without adequate standards for safety, transparency, or data protection (Morley et al., 2020). The risks arising from their therapeutic or quasi-therapeutic use in mental health therefore warrant urgent examination.
First, using LLMs for mental health support entails processing intimate and highly sensitive information, including symptoms, trauma narratives, medication histories, and crisis-related disclosures (Mandal et al., 2025; Wang et al., 2025). These interactions can also generate sensitive inferences (e.g., suicide risk, substance use, or exposure to abuse) even when users do not explicitly disclose them, increasing the potential for privacy harms if data are mishandled.
Second, the technological infrastructure that enables these services is commonly located outside the jurisdictions of LMICs, under privacy policies that permit the use, storage, and training on personal user data (Vollmer et al., 2020). This extraterritoriality complicates enforcement and redress mechanisms and weakens cross-border accountability, particularly where local regulatory agencies have limited technical capacity or unclear legal authority over foreign providers.
Third, foundational model development and data management remain opaque, including uncertainty regarding the provenance of training corpora, data governance practices, and safeguards to meet expectations of medical confidentiality (Bommasani et al., 2023). The "black box" nature of these systems also complicates auditability and post hoc investigation when harmful outputs occur, limiting effective oversight (Ethical AI governance group, 2023).

3. In the passage stating that "this law aims to serve as a counterpart to the GDPR… or the HIPAA," we suggest a more nuanced formulation. The GDPR and HIPAA differ substantially in

scope, legal nature, and applicability (general data protection versus a sector-specific healthcare framework). A reformulation would avoid an overly direct equivalence.

Reply: We modified by:

"In the Peruvian case, the Law on Personal Data Protection (Law No. 29733) is insufficient to address emerging generative AI risks because it does not encompass critical aspects such as sensitive inferences, algorithmic reuse, or re-identification risks (Smart & Montori, 2025). This gap is particularly salient because, while Peru's data protection framework shares broad intent with comprehensive regimes such as the EU's GDPR, it is not directly comparable to sector-specific U.S. frameworks such as HIPAA and does not yet address AI-specific risks (e.g., sensitive inferences, algorithmic reuse, and re-identification). As a result, users may be exposed to privacy violations with emotional, clinical, and societal consequences."