
Article:
Ahmed, H. orcid.org/0000-0001-8952-4190 (2026) Large language models for clinical trials in the Global South: opportunities and ethical challenges. AI and Ethics, 6 (1). 76. ISSN: 2730-5953

https://doi.org/10.1007/s43681-025-00943-x

AI and Ethics (2026) 6:76
https://doi.org/10.1007/s43681-025-00943-x

REVIEW

Large language models for clinical trials in the Global South: opportunities and ethical challenges

Hafiz Ahmed[1,2]

Received: 16 August 2025 / Accepted: 4 December 2025

Abstract
Large language models (LLMs) show promise for improving clinical trials in wealthy countries but remain underexplored in low- and middle-income countries (LMICs), where healthcare infrastructure is weaker and resources limited. This article explores opportunities for LLM integration and addresses ethical challenges in LMIC clinical trials through a review of recent literature (2024–2025) on LLMs in healthcare and clinical research, examining adaptation potential for LMICs using thematic analysis to identify ethical issues specific to LMIC contexts, including data control, fairness, and sustainability. LLMs can accelerate protocol development, improve multilingual patient recruitment, streamline regulatory processes, and address data gaps through synthetic records; however, implementation raises concerns about data privacy, community representation, AI transparency, and technological dependence on foreign platforms. While LLMs can enhance clinical trial efficiency and inclusivity in LMICs, successful integration requires locally-adapted models, community-centered ethical oversight, and regional partnerships, with thoughtful implementation potentially democratizing healthcare innovation benefits across Global South populations.

Keywords Large language model · Clinical trial · Public health · AI ethics · Patient recruitment · Artificial intelligence

1 Introduction

Large language models (LLMs) and artificial intelligence (AI) have recently gained attention for their potential to further improve clinical research, which is useful in tasks such as drafting trial documents and matching patients to relevant and appropriate studies. Initial tests by the researchers from Boston Consulting Group found that LLMs can quickly create clinical trial protocols and informed consent materials, therefore potentially speeding up the research process significantly [1]. Similarly, AI tools like *TrialGPT* [2] have successfully identified eligible participants with high accuracy, cutting patient screening time by 42.6% when evaluated on 183 synthetic patients involving three cohorts and over 75,000 trial. It is important to highlight here that most of these promising outcomes come from high-income countries with extensive digital health infrastructure and detailed patient data. Unlike wealthier countries (upper middle-income and above), low- and middle-income countries (LMICs) face very different challenges, including limited healthcare funding, incomplete or sparse electronic patient records, and fewer suitably qualified research professionals [3]. An analysis covering 27 clinical trials − 5 published and 22 ongoing (with 10 in China, 6 in the USA, 3 in Canada, 2 in Germany, and single studies in Italy, Denmark, Taiwan, Turkey, Saudi Arabia, plus one multinational trial spanning Germany, Italy, China, and the USA) - found a notable concentration of trials primarily in the US and China [4]. Remarkably, none of these studies took place or are currently taking place in LMIC settings. This clear imbalance raises a critical question: *How might LMICs practically harness LLMs to carry out meaningful clinical trials*, *and what unique ethical considerations and opportunities might arise within these contexts?* This article addresses the question posed above by examining the practical advantages and ethical considerations associated with the use of LLMs in clinical trials within LMICs, while also comparing them with traditional trial methods. It does so by summarizing recent scholarly literature (2024–2025) on the use of LLMs

✉ Hafiz Ahmed
hafiz.h.ahmed@ieee.org

[1] School of Electrical and Electronic Engineering, University of Sheffield, Sheffield, UK

[2] Autonex Systems Limited, London, UK

Published online: 21 December 2025

in healthcare and clinical research, highlighting how these technologies could be adapted for LMICs. Additionally, it closely examines ethical issues of the AI and LLM technologies [5, 6] uniquely relevant to LMIC contexts, including data control, fairness, transparency, and long-term sustainability, driven by insights from thematic analyses of current discussions and case examples in global health and AI ethics research. The rest of this article is organized as follows: Sect. 2 enumerates the benefits of employing LLMs for clinical trials in LMICs, while Sect. 3 outlines the associated ethical concerns and challenges. Finally, concluding remarks and future perspectives are presented in Sect. 4 .

2 Potential benefits of LLMs for clinical trials in LMICs

LLMs in clinical trials within LMICs may help bridge various challenges related to limited resources and infrastructure that often disrupt conventional trial processes. This Section summarizes these benefits. However, it is important to note that the majority of promising outcomes demonstrating LLM efficiency in clinical trials have been reported in high-income countries (HICs) with established digital health ecosystems, stable high-speed connectivity, and extensive, structured electronic patient records. When considering their application in LMICs, it is vital to acknowledge the friction points inherent in these environments, including sparse or incomplete patient data, reliance on paper records, intermittent digital infrastructure, and limited resources for advanced computational deployment. Therefore, realizing the benefits below requires a strategic approach that differentiates between HIC-demonstrated potential and LMIC-feasible adaptations that account for these resource and infrastructural constraints.

2.1 Accelerated trial design and documentation

LLMs can act as helpful partners for local researchers when planning clinical trials. They can quickly pull together drafts of protocols, case report forms, and regulatory documents by drawing on a broad base of scientific literature and international guidelines [7]. In one example, an LLM equipped with access to regulatory resources was able to produce protocol sections that aligned with FDA standards [1]. This can be especially valuable in regions where trial teams often depend on expensive outside consultants to prepare complex documentation. In LMICs, LLMs can efficiently support initial drafting and compliance checks, potentially speeding up the research process significantly. By streamlining the writing process and supporting compliance with global norms (like CONSORT or GCP), LLMs can help lower both costs and turnaround times, potentially allowing local teams to take greater ownership of their studies. According to the estimate presented in Fig. 4 in [8], regulatory compliance, including trial design and documentation, accounts for roughly 10% of the total trial cost in LMICs, indicating that potential monetary savings from using LLMs can be substantial. However, in LMIC contexts where local regulatory bodies may lack structured digital standards or have unique, non-digitized requirements, it is crucial for experienced investigators to manually verify the output of LLMs against local regulatory requirements. This verification step ensures that AI-generated documents comply with the specific ethical and legal mandates of the host country, mitigating risks associated with the LLM's training being primarily based on HIC guidelines. For a more technically detailed study on this LLM-driven clinical documentation, interested readers may consult the study by Maleki and Ghahari [9].

2.2 Improved patient recruitment and matching

Accurate patient recruitment is critical to the success and generalizability of clinical trials, yet it is often hampered in LMICs by fragmented and non-standardized record-keeping. LLMs are especially good at understanding unstructured text, making them effective at reviewing clinical notes, referral documents, or even informal reports from community healthcare workers to pinpoint suitable candidates according to detailed eligibility requirements. LLMs offer a significant opportunity here due to their proficiency in natural language processing, which allows them to extract and synthesize key clinical information from unstructured, incomplete, or handwritten patient notes. This is a strength particularly relevant to LMIC clinics, where structured electronic health records are sparse, and patient data often exists in diverse, difficult-to-analyze formats. By analyzing this heterogeneous data, which can include clinician notes, lab slips, and scanned documents, LLMs can rapidly identify patients who meet complex eligibility criteria, far outpacing manual review. A recent trial-matching system powered by an LLM successfully identified over 90% of applicable trials for individual patients, achieving accuracy comparable to medical experts and significantly reducing the time needed for patient screening [10]. However, a key distinction must be drawn: The high-accuracy patient matching reported in literature is typically based on clean, structured data from digitally mature HIC environments. In LMICs, achieving high-accuracy matching remains hypothetical unless advanced pre-processing of heterogeneous data can be reliably executed on limited infrastructure. This process may also benefit from recent advances in machine learning, such as zero-shot learning or transfer learning from related

task [11] to manage the heterogeneous, often handwritten or scanned, patient data typical of LMIC clinics. Additionally, this may reduce token usage and thus lower the cost of LLMs, a particularly useful feature in LMIC settings. Note that this work is currently an area of active research rather than a widely demonstrated capability, requiring local validation studies before deployment.

2.3 Enhanced participant communication and engagement

LLMs hold significant promise for improving communication in clinical trials by overcoming linguistic barriers. In settings with low literacy rates or multiple local languages that lack standardized written forms, LLM-based assistants can rapidly generate materials (e.g., patient-facing summaries, consent forms) in various local languages. This capacity is essential for true linguistic inclusion, potentially ensuring that more communities can access and understand trial information. Recent developments show promise for using conversational AI to educate patients through improved readability of the consent form [12] (supported by LLM's multi-lingual translation capability [13]) and even collect informed consent through interactive dialogues [4]. In LMIC settings where formal education might be limited, a friendly and supportive AI assistant could respond to participant queries around the clock, encourage participants to follow study guidelines (for example, reminders about clinic appointments or medication usage), and communicate patient concerns back to the research team [14]. By acting as accessible patient support representatives, LLMs can potentially increase participant retention rates and overall satisfaction, addressing gaps left by traditional methods where patients often struggle to receive timely answers outside scheduled visits. However, a crucial qualification must be made regarding this potential benefit. The successful deployment of LLMs for translation and communication depends significantly on the quality and representativeness of their linguistic training data. Low-resource languages with limited or non-digitized corpora present a considerable challenge. Despite recent advancements for such languages (cf. ref. [15]), the field is still not fully developed. When general-purpose models are used to translate complex medical or ethical concepts into these low-resource languages, the risk of misinterpretation increases. Over-reliance on these general models can lead to errors, simplifications, or neglect of culturally sensitive nuances. Therefore, any LLM-based translation tool must undergo rigorous local validation and be fine-tuned using data curated by local linguistic and cultural experts to ensure both accuracy and appropriateness.

2.4 Real-time data analysis and monitoring

Clinical trial sites with limited resources often struggle with effective data monitoring, causing delays in identifying safety risks or errors. LLMs have the potential to analyze trial data and reports in real time. For instance, an LLM could actively review free-text adverse event logs or clinician notes, rapidly identifying (potentially predicting as well [16]) worrying trends, like increased reports of particular side effects [17] faster than conventional monitoring methods. According to recent insights from industry professionals, LLMs might soon facilitate almost instantaneous detection of anomalies or data irregularities. This would allow researchers to immediately address any issues, significantly improving patient safety [18]. Such continuous monitoring could be particularly beneficial for trials conducted in LMIC settings where advanced data oversight is usually limited. Furthermore, LLMs might help generate interim reports or statistical summaries for Data Safety Monitoring Boards, automating tasks that typically require specialized statistical knowledge that local teams may lack.

2.5 Regulatory compliance and reporting

Carrying out clinical trials that align with international regulatory standards can be especially difficult in the Global South, primarily because of insufficient regulatory training and support. LLMs might address this by automatically checking protocols, case report forms, and consent documents to ensure that they comply with ethical and regulatory requirements. They can verify whether essential components outlined by guidelines such as CONSORT or GCP are included, helping catch mistakes that local teams may overlook. Furthermore, LLMs could simplify the creation of applications for ethics committees or regulatory authorities by generating necessary documentation and translating it into the formats expected by various agencies [18]. Acting like virtual regulatory advisors, these AI systems could guide investigators in LMIC settings through complicated compliance processes more efficiently than informal manual approaches. This would likely reduce approval delays and enhance overall trial quality, addressing typical regulatory bottlenecks encountered with traditional methods.

2.6 Knowledge transfer and capacity building

One of the most promising developments is how LLMs could open up access to medical knowledge for a wider audience [19]. In lower-resource settings, such as LMICs, researchers and healthcare providers can use LLMs to explore up-to-date evidence or trial design approaches that fit their specific needs. These tools can function like

on-demand research advisors or medical librarians, offering summaries of key literature and established practices, much of which might otherwise be hard to reach due to paywalls or language barriers. This access can help spark research questions rooted in local realities and lead to stronger, more context-specific study designs. Additionally, by adapting LLMs to reflect regional data (when available), it is possible to make them more attuned to local health conditions, languages, and cultural factors. In such cases, a custom model may offer far more relevant insights than a one-size-fits-all version [20]. Over time, using LLMs in this way could strengthen research capacity. Early-career investigators, for instance, could learn through the feedback and examples these tools provide, which may enhance overall trial quality. Ultimately, LLMs have the potential to close knowledge gaps, delivering some of the benefits typically seen in highly developed research environments to areas where resources are limited.

It is crucial to note that the above-mentioned advantages depend heavily on how well the systems are put into practice. When applied thoughtfully and under proper human supervision, LLMs have the potential to greatly improve the speed and inclusivity of clinical trials in the LMICs, making it possible to conduct studies more efficiently and affordably without compromising on quality. In contrast, traditional trial methods in many of these regions tend to be slow, rely heavily on paper records, and often need outside assistance. Leveraging LLMs could help bypass some of those hurdles and bring a more streamlined approach to trial execution.

3 Ethical concerns and challenges in LMIC settings

Although LLMs offer significant potential to enhance clinical trials, introducing them in LMICs brings with it a set of important ethical and practical concerns. Many of the ethical questions already debated in wealthier settings can become even more complex in LMICs, where long-standing inequalities and less robust oversight systems may intensify the risks [21]. As we move toward incorporating LLMs into clinical research, it is essential to carefully examine these challenges from multiple angles. These challenges are summarized below:

3.1 Data privacy and security

For LLMs to be effective, they often need access to data such as patient records and clinical trial documentation. However, many health systems in LMICs lack strong protections for personal data. Relying on external cloud-based AI platforms could put sensitive patient information at risk, potentially breaching confidentiality. Experts have cautioned that without proper oversight, using cloud-hosted LLMs could result in privacy violations and unauthorized access to data [22]. This concern is even more pressing in LMICs, where research participants may already face vulnerabilities and might not be fully aware of their data rights. This vulnerability is significantly intensified by the institutional realities in many LMICs:

- *Lack of Statutory Data Protection*: Many nations lack comprehensive national privacy legislation or a statutory data protection framework. Without robust laws defining data ownership, transfer, and breach liability, local research sites and participants are left exposed.
- *Weak Governance*: This deficit is often compounded by the frequent absence of formalized, well-resourced data governance bodies capable of auditing AI usage or enforcing international data standards.
- *Contractual Vulnerability*: This institutional gap places the burden of protection on contractual agreements. Local sites often have limited capacity and leverage to negotiate stringent contractual protections (e.g., preventing data re-use for commercial model training) when partnering with powerful external (often HIC-based) technology providers.
- *Example*: A trial site in a country without national data protection laws utilizes a free, cloud-based LLM for translating patient data. The generic Terms of Service permit the LLM provider to use uploaded texts for future model training, effectively transferring sensitive patient data control to a foreign commercial entity without adequate local legal recourse, illustrating the practical vulnerability in the absence of institutional backing.

Ethical use of these technologies requires strict adherence to privacy protocols, which may mean keeping data within national borders through on-site deployment or locally adapted models. Yet running advanced LLMs locally presents its own challenges, as they typically demand substantial computational resources. A promising workaround involves using smaller, more efficient models that are designed for environments with limited infrastructure [23]. Regardless of the approach, researchers must secure informed consent before using patient data with AI tools and ensure that strong cybersecurity protections are in place. Without these safeguards, there's a real risk that LLM deployment could compromise participant privacy and erode public trust in the research process.

3.2 Bias and generalizability

Most mainstream LLMs are trained primarily on data from wealthier countries, meaning their outputs often unintentionally reflect biases tied to those populations. Some readers may recall immediately the pulse oximeter controversy during the height of the COVID-19 pandemic [24]. Without careful consideration, these models may produce clinical trial guidelines or eligibility rules that do not align well with local genetics, lifestyles, or cultural norms.

3.2.1 Mechanisms of bias reproduction

The risk of bias in LLM-enhanced trials is magnified by three specific mechanisms inherent to the LMIC research environment:

- *Limited Local Data Availability*: If an LLM is used to screen for eligibility or predict outcomes, it will perform less reliably for specific ethnic groups or region-specific disease patterns that were underrepresented in its initial training data. This lack of local data can lead to systematic under-diagnosis or misclassification of certain LMIC populations.
- *Uneven Representation of Disease Patterns*: Models trained overwhelmingly on HIC data may fail to accurately assess conditions or co-morbidities that are prevalent in LMICs (e.g., infectious diseases, malnutrition-related conditions), leading to potentially unsafe recommendations.
- *Structural Power Imbalances*: The power imbalance between external sponsors and local investigators means that sponsors may prioritize deploying a single, pretrained global model over investing in specialized local fine-tuning. This effectively allows the structural biases embedded in the global model to be imposed upon the local research setting, reinforcing existing inequities. This situation could lead to standardized trial designs that ignore critical local biological or cultural differences, potentially excluding certain ethnic groups or overlooking region-specific disease presentations.

This situation could lead to standardized trial designs that ignore critical local biological or cultural differences, potentially excluding certain ethnic groups or overlooking region-specific disease presentations.

3.2.2 Feasibility of mitigation

To address fairness, one potential solution is to countering these biases by customizing models using local healthcare documents and involving diverse local experts to review the AI-generated outputs. Fairness and bias are ongoing ethical challenges healthcare [24]. However, the feasibility of full local adaptation of massive LLMs is constrained by realistic barriers: local institutions often lack the necessary computational resources, data governance structures, and technical expertise required to train or fine-tune billion-parameter models. A more practical mitigation strategy involves utilizing small language models (SLMs) and other efficient, specialized models designed for environments with limited infrastructure [23]. These SLMs are easier and less expensive to train and deploy locally, offering a more realistic path for local adaptation and context-specific fine-tuning. Ultimately, ethical responsibility demands that trials enhanced by AI do not reinforce existing inequities. AI-generated recommendations should always be critically assessed in terms of local relevance, fairness, and cultural sensitivity. LLMs can also be leveraged to reduce bias if thoughtfully applied, for example, by reviewing whether trial recruitment adequately includes diverse community groups [18], or by clearly translating eligibility criteria into accessible language suited to different populations, thereby promoting broader inclusivity.

3.3 Informed consent and participant autonomy: technical and reliability risks

Obtaining genuinely informed consent in clinical trials within LMICs has historically been challenging because of language barriers, technical terminology, and culturally different understandings of consent. The introduction of LLMs into this process brings both potential benefits and new ethical concerns. On the positive side, an AI-based assistant might help by clearly explaining trial information to participants in their local language, answering questions promptly, and making the consent process more understandable. Studies are already underway to evaluate how LLMs can support patient education and consent conversations [4]. Nevertheless, caution is crucial regarding the reliability and accuracy of the AI output. LLMs sometimes confidently present inaccurate information, commonly known as "hallucinations", which could seriously mislead participants about the real risks or benefits involved [25]. When these hallucinations occur in a local language translation, they can specifically introduce factual errors or misrepresentations of trial procedures or risks. In [26], it has been found that hallucinated translations can pose serious safety issues in machine translation, particularly in low-resource languages and when translating from English, revealing toxic patterns from training data. Simply scaling existing models is insufficient to address these hallucinations; instead, employing diverse fallback models can improve translation quality and mitigate harmful outcomes. Beyond simple factual error, the

risk of culturally inappropriate or simplified output in local languages is a concern, which could inadvertently alter the meaning of consent or patient rights. For instance, translating a complex concept of "risk-benefit ratio" into a language that lacks direct scientific equivalents may result in a culturally simplified phrase that minimizes the actual risks. Therefore, if participants interact with an AI chatbot for consent purposes, it must be made explicitly clear that the assistant is an AI system rather than a human doctor, to prevent undue influence or misunderstanding.

3.4 Interpersonal and cultural dimensions of consent

The involvement of an LLM in the consent process fundamentally changes the nature of the human-to-human interaction, raising ethical concerns specific to the sociocultural context of LMICs. In many communities, particularly those with a history of extractive or externally driven research [21], trust in research institutions is a complex, fragile, and interpersonally mediated process. The introduction of a sophisticated, foreign-developed AI tool may be perceived as another mechanism reinforcing a structural power imbalance between the sponsoring body (often foreign) and the local participant. This can inadvertently undermine the participant's autonomy, especially if the AI is perceived as projecting an unwarranted authority that discourages questions or dissent, or if the AI is viewed as a replacement for a human researcher. Furthermore, informed consent in many LMIC settings is not solely an individual act but involves relational decision-making. Decisions may be made in consultation with family elders, community leaders, or spouses, often involving processes of deference to collective authority. An LLM-based assistant, which is designed for one-on-one, transactional information delivery, risks bypassing or disrupting these critical cultural frameworks. The lack of a human, relational presence in the consent explanation, even if technically accurate, may fail to provide the necessary sense of safety and legitimacy required for truly informed, voluntary participation within the community's established decision-making protocols. Research protocols must therefore ensure that LLM tools serve only as supportive aids and do not replace the human-to-human interaction necessary to establish trust and respect for local relational dynamics. We emphasize the importance of distinguishing clearly between a human communication partner (the researcher) and the AI assistant to preserve the integrity of the participant-researcher relationship.

3.5 Accountability and oversight

Bringing AI into clinical trials makes assigning responsibility more complex. If an AI system generates an incorrect protocol or misjudges patient eligibility, causing harm, it becomes unclear who is at fault: the AI developers, local research teams, or the trial sponsors? This issue is especially critical in LMIC contexts, where AI-specific regulatory frameworks are limited or absent. Many LMIC regulatory bodies are still struggling to develop comprehensive guidelines for conventional trials, much less for trials using advanced AI tools.

3.5.1 Delineation of risk and responsibility

A clearer delineation of which parts of the trial workflow are most sensitive to misattribution of responsibility is necessary, as errors carry different ethical and legal implications:

- *Protocol Drafting Errors*: LLM errors in generating sections like dosing schedules or inclusion/exclusion criteria can have direct legal and patient safety implications. The local investigator is ultimately responsible for verifying and signing off on the final protocol as mandated by Good Clinical Practice.
- *Safety Monitoring Errors*: Errors introduced by an LLM during the review of safety data (e.g., failing to identify an adverse event trend or mis-summarizing a serious adverse event) pose an immediate, critical patient safety risk.
- *Recruitment/Eligibility Errors*: LLM misclassification of eligibility criteria carries primary fairness and ethical implications. If eligible patients are excluded due or if ineligible patients are included based on biased LLM suggestions, the investigator bears responsibility for the ethical breach and potential harm to the study's scientific validity.

In this context, ref. [27] offers a valuable perspective by proposing a model that combines a human-centric approach to data ethics with John Rawls' concept of the common good. Such an approach aligns with the lived realities of LMIC environments, emphasizing individual autonomy, privacy protection, and fairness, while still enabling harmonization with emerging global data-governance frameworks. Applying these ideas to AI-enabled clinical trials suggests that accountability structures should not only safeguard individuals' rights but also ensure that shared benefits are equitably distributed to local populations, preventing extractive data practices or exploitation. This perspective supports our argument for locally adapted models while also acknowledging the usefulness of complementary global frameworks

that can help standardize protections against risks such as data breaches, algorithmic bias, and opaque accountability boundaries.

Note that the regulatory gaps risk enabling "ethics dumping" [28], where external sponsors introduce sophisticated AI technologies into vulnerable communities without proper oversight or governance. Therefore, it is essential to advocate for explicit guidelines on how LLMs should be used ethically in human research. Ethics committees in LMICs need specialized training to evaluate the role and implications of AI in research proposals.

3.5.2 Implementation of concrete audit trails

Ethicists strongly recommend always maintaining human oversight [25]. Specifically, the outputs from LLMs should serve only as suggestions for human researchers, rather than being considered definitive decisions. Establishing clear audit trails, which document AI recommendations and show precisely how final decisions were made, is crucial. To implement this concretely, especially in settings where digital infrastructure may be intermittent or inconsistent, audit trails should be structured as follows:

- *Mandatory Human Sign-off Checkpoints*: Require mandatory human sign-off checkpoints for all LLM-generated critical outputs (e.g., final protocol sections, Adverse Event summaries). The researcher must confirm they have reviewed and verified the output.
- *Offline Logging Mechanism*: Audit trails should utilize local, timestamped logging to record LLM inputs, outputs, and human verification status on site-level devices. This mechanism is designed to operate even during periods of network outage.
- *Synchronization upon Connection*: The local logs must be designed to automatically synchronize with a central server when stable connectivity is restored, ensuring a complete, tamper-resistant record of the decision-making process.
- *Documentation of AI Version*: The audit trail must require the documentation of the specific LLM version, parameters, and fine-tuning data used for that particular output, ensuring transparency and reproducibility.

Researchers and investigators should ultimately verify and take responsibility for all AI-generated outputs. Without these safeguards, there is a risk that errors will remain unnoticed, or sponsors might attempt to shift responsibility onto AI systems to evade accountability, a scenario that would be ethically unacceptable. It is important to note that integrating AI effectively demands stronger (not weaker) governance mechanisms for clinical trials in LMICs.

3.6 Exploitation and equity concerns

Historically, clinical research in the Global South has often been criticized as exploitative, with richer countries benefitting disproportionately from studies conducted in poorer regions [21]. Introducing LLMs into this context could either reduce or increase such exploitation. On the positive side, LLMs might make trials quicker and cheaper, enabling local researchers in LMICs to carry out more studies that directly address local priorities, such as neglected tropical diseases. This would help ensure that local communities directly benefit from research outcomes. However, there is also a risk that if LLM technology remains dominated by large foreign corporations or sponsors, trial sites in LMICs could end up simply providing valuable data for AI training, without receiving fair benefits in return. For instance, data collected in an LMIC trial might improve a commercial AI model, yet the local healthcare system might not gain access to the resulting innovations. This situation raises critical ethical concerns around data ownership and benefit-sharing.

3.6.1 Workable governance solutions for Benefit-Sharing

Fairness and justice demand that AI-supported clinical research in LMIC settings should be directly linked to local health needs, with clear benefits flowing back to local communities. This requires transitioning from identifying the risk of exploitation to outlining workable governance solutions in contractual agreements and policy frameworks:

- *Local Co-ownership of Datasets*: Mandate local institutional co-ownership of the datasets collected during the trial, especially those used for LLM fine-tuning, ensuring that the LMIC sites have a legal voice in their downstream use.
- *Conditions on Downstream Commercial Use*: Implement clear contractual conditions that require fair compensation or licensing fees for any LLM that uses the LMIC-generated data to train a commercial product.
- *Requirements to Return Model Outputs and Capacity*: Require that any locally-trained, specialized LLMs or AI tools developed using local data be returned to the originating health system at no cost and with full documentation, thus linking LLM use directly to sustainable local capacity building.
- *Affordability Guarantees*: Mandate that successful interventions or LLM-based tools arising from the research remain affordable or free for the originating LMIC health systems and populations.

Another ethical challenge is the potential widening of existing inequalities; urban, well-funded hospitals might attract

more trials due to better access to AI technologies, leaving rural or less-equipped hospitals further behind. Addressing equity requires thoughtful deployment strategies, possibly through global partnerships that make AI tools affordable or free for all participating sites. Ultimately, researchers must consistently reflect on a fundamental question: Will using this LLM in our clinical trial strengthen local healthcare and research capacities, or will it merely take value away? Ensuring the answer prioritizes local empowerment is essential for ethically responsible research.

3.7 Reliability and quality of AI outputs

Lastly, a practical ethical issue involves ensuring the accuracy and reliability of content generated by LLMs. Traditionally, clinical trial processes rely heavily on expert judgment, but such expertise may be scarce in LMIC settings, making AI an appealing alternative. However, excessive dependence on AI for critical tasks like clinical trials carries significant risks. It is well-documented that LLMs sometimes produce incorrect or illogical outputs, which might not be obvious to a less-experienced researcher [25]. If local researchers accept AI-generated protocols without careful scrutiny, participants could face real dangers.

3.7.1 Granular examination of model failure modes

Different types of model failure and their distinct implications are examined below:

- *Errors in Clinical Interpretation/Dosing*: If an LLM incorrectly analyzes a patient's case report, misinterprets complex biological data, or suggests an unsafe medication regimen or inaccurate dosage, this poses a direct and immediate patient safety risk. This requires mandatory verification by a clinician.
- *Misclassification of Eligibility Criteria*: If the LLM overlooks or misinterprets patient data based on HIC-centric training or existing biases, it can lead to the systemic exclusion of eligible groups. This error carries significant bias and ethical exclusion risks (Sect. 3.2), undermining the trial's generalizability and fairness.
- *Overly Confident but Incorrect Summaries ("Hallucinations")*: LLMs can generate confidently asserted but false summaries [25]. In resource-limited settings where expertise is limited, there is a risk of misleading less-experienced researchers who may lack the background to spot subtle but critical inaccuracies in protocol drafts or data analysis, thereby compromising overall trial quality.

A recent Delphi study of AI in healthcare also highlighted the danger of overly trusting AI-generated recommendations [22]. This highlights the importance of training researchers to approach AI suggestions with appropriate skepticism. International sponsors should not completely delegate trial design to AI at LMIC sites simply to reduce costs, unless they also provide support for rigorous human "oversight".

3.7.2 Quality assurance and validation

In any application of LLMs, whether drafting documents, recruiting patients, or data analysis, a qualified person should always verify and refine the AI's suggestions. Furthermore, formal validation studies should be implemented, such as:

- Comparing trial protocols drafted by an LLM against those prepared by experienced investigators.
- Checking patient eligibility recommendations from AI tools against current standard methods to confirm no eligible groups are overlooked.

Requiring this kind of validation builds in quality assurance similar to software verification, integrating it directly into the ethical oversight of clinical trials. Ultimately, the goal of incorporating LLMs should be to support human expertise, not replace it, at least until there is conclusive evidence that these systems are consistently reliable across diverse global contexts.

4 Conclusion and perspective

LLMs offer promising opportunities to improve clinical trials in the Global South. They can address persistent issues, such as streamlining extensive paperwork, analyzing unstructured patient data to improve recruitment, overcoming language barriers, and providing timely analytical insights to boost trial efficiency and patient experiences. Unlike traditional methods, which are often slow and resource-intensive, strategic use of LLMs could make high-quality clinical trials more accessible to underfunded regions, empowering them to manage studies more independently. Additionally, LLMs could help adapt trials to local health priorities, enhancing their relevance and effectiveness in LMICs.

However, achieving these advantages without causing harm demands proactive ethical safeguards and a strong commitment to fairness. Implementing LLMs in clinical trials must include robust data protection protocols, active strategies to reduce bias, and clear guidelines

outlining responsibility and accountability. A recent systematic review emphasized that the potential benefits of LLMs, such as their capabilities in data analysis, information provision, and decision support, cannot be separated from serious ethical concerns about fairness, transparency, and the risks posed by convincingly wrong outputs [25]. These general ethical concerns become even more critical in LMIC settings, highlighting the importance of rigorous oversight. Global sponsors, local researchers, ethics committees, and community representatives all have roles to play in ensuring that the deployment of LLMs aligns with local values and genuinely meets community needs.

LLMs could significantly enhance clinical trials in resource-constrained regions, but they are not a simple solution. Their introduction should reinforce ethical standards, not bypass them. Through careful planning, strengthening local capacity, and vigilant regulatory oversight, LLM technology can help clinical research become more inclusive and effective in the Global South. This careful approach could accelerate medical breakthroughs benefiting communities worldwide. Conversely, careless adoption could lead to repeating historical patterns of exploitation and harm. Therefore, it is crucial for all stakeholders to responsibly guide this innovation, ensuring AI technology complements human judgment and integrity, ultimately improving health outcomes while maintaining high ethical standards in every clinical trial, wherever it occurs.

Acknowledgements We would like to thank the anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions.

Author contributions The author conceived the study, conducted the literature review, performed the thematic analysis, and wrote and revised the manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability All data that support the findings of this study are included within the article.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval This research does not require any ethical approval.

References

1. Markey,N., El-Mansouri, I., Rensonnet, G., van Langen, C., Meier, C.: From RAGs to riches: utilizing large language models to write documents for clinical trials. Clin. Trails **22**(5), 626–631 (2025). https://doi.org/10.1177/17407745251320806
2. Yang, J., Dung, N.T., Thach, P.N., Phong, N.T., Phu, V.D., Phu, K.D.: Generalizability assessment of AI models across hospitals in a low-middle and high income country. Nat. Commun. **15**(1), 8270 (2024). https://doi.org/10.1038/s41467-024-52618-6
3. Kozlakidis, Z., Muradyan, A., Sargsyan, K. (eds.): Digitalization of medicine in low- and middle-income countries. Sustainable Development Goals Series. Springer International Publishing (2024). Available from: https://doi.org/10.1007/978-3-031-62332-5
4. Omar, M., Nadkarni, G.N., Klang, E., Glicksberg, B.S.: Large language models in medicine: a review of current clinical trials across healthcare applications. PLOS Digit. Health **3**(11), e0000662 (2024). https://doi.org/10.1371/journal.pdig.0000662
5. Deng, C., Duan, Y., Jin, X., Chang, H., Tian, Y., Liu, H.: Deconstructing the ethics of large language models from long-standing issues to new-emerging dilemmas: a survey. AI Ethics (2025). https://doi.org/10.1007/s43681-025-00797-3
6. Schneider, D., Brown, E., Ward, M., Sciubba, D., Lo, S.-F.: The authorization lottery: contradictory AI prioritization patterns in healthcare resource allocation. AI Ethics (2025). https://doi.org/10.1007/s43681-025-00795-5
7. Hosseini, M., Gao, C.A., Liebovitz, D.M., Carvalho, A.M., Ahmad, F.S., Luo, Y., et al.: An exploratory survey about using ChatGPT in education, healthcare, and research. PLoS ONE. **18**(10), e0292216 (2023). https://doi.org/10.1371/journal.pone.0292216
8. Theresia, T.Y., Pham, T., Baker, L., et al.: Cost drivers and predictive modeling of clinical trial costs: analysis of 101 global health trials [version 1]. VeriXiv. **2**, 152 (2025). https://doi.org/10.12688/verixiv.1282.1
9. Maleki, M., Ghahari, S.: Clinical trials protocol authoring using LLMs. arXiv (2024). https://arxiv.org/abs/2404.05044
10. Jin,Q., Wang, Z., Floudas, C.S., Chen, F., Gong, C., Bracken-Clarke, D., et al.: Matching patients to clinical trials with large language models. Nat. Commun. **15**(1), 9074 (2024). https://doi.org/10.1038/s41467-024-53081-z
11. Wornow, M., Lozano, A., Dash, D., Jindal, J., Mahaffey, K.W., Shah, N.H.: Zero-shot clinical trial patient matching with LLMs. NEJM AI. **2**(1), AIcs2400360 (2025). https://doi.org/10.1056/aics2400360
12. Litt, H.K., GreenstreetAkman, E., Idossa, D., Florez, N., Manana, V.: Improving clinical trial consent form readability through artificial intelligence. JCO Oncol. Pract. **19**(11suppl), 109–109 (2023). https://doi.org/10.1200/op.2023.19.11_suppl.109
13. Kleidermacher, H.C., Zou, J.: Science across languages: assessing LLM multilingual translation of scientific papers. arXiv (2025). Available from: https://arxiv.org/abs/2502.17882
14. Li, Y., Wen, H., Wang, W., Li, X., Yuan, Y., Liu, G., et al.: Personal LLM agents: insights and survey about the capability, efficiency and security. arXiv (2024). Available from: https://arxiv.org/abs/2401.05459
15. Khoboko, P.W., Marivate, V., Sefara, J.: Optimizing translation for low-resource languages: efficient fine-tuning with custom

prompt engineering in large language models. Mach. Learn. Appl. **20**, 100649 (2025). https://doi.org/10.1016/j.mlwa.2025.100649
16. Kan,B., Wang, T., Zhu, H., Liang, R., Yan, E., Wang, F.L., et al.: An Event-aware dual representation model with mixture-of-experts for serious adverse events prediction in clinical trials. IEEE Trans. Consum. Electron. (2025). https://doi.org/10.1109/tce.2025.3572452
17. Chiang, C., Luo, M., Dumkrieger, G., Trivedi, S., Chen, Y., Chao, C., et al.: A large language model–based generative natural language processing framework fine-tuned on clinical notes accurately extracts headache frequency from electronic health records. Headache **64**(4), 400–409 (2024). https://doi.org/10.1111/head.14702
18. Kondru, R.: Delivering better outcomes for patients with LLMs. Applied Clinical Trials Online (2024). Available at: https://www.appliedclinicaltrialsonline.com/view/delivering-better-outcomes-patients-llms
19. Chen, A., Liu, L., Zhu, T.: Advancing the democratization of generative artificial intelligence in healthcare: a narrative review. J. Hosp. Manag. Health Policy **8**:12 (2024). https://doi.org/10.21037/jhmhp-24-54
20. Fong,A., Adams, K.T., Boxley, C., Revoir, J.A., Krevat, S., Ratwani, R.M.: Does one size fit all? Developing an evaluation strategy to assess large language models for patient safety event report analysis. JAMIA Open **7**(4), ooae128 (2024). https://doi.org/10.1093/jamiaopen/ooae128
21. Bittker, B.M.: The ethical implications of clinical trials in low- and middle-income countries. Human Rights Magazine – American Bar Association (2021). Available at: https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/the-truth-about-science/the-ethical-implications-of-clinical-trials/
22. Denecke, K., May, R., Rivera Romero, O.: Potential of large language models in health care: Delphi Study. J. Med. Internet Res. **26**, e52399 (2024). https://doi.org/10.2196/52399
23. Garg, M., Raza, S., Rayana, S., Liu, X., Sohn, S.: The rise of small language models in healthcare: a comprehensive survey. arXiv (2025). https://arxiv.org/abs/2504.17119
24. Valbuena, V.S.M., Merchant, R.M., Hough, C.L., American Medical Association (AMA): Racial and ethnic bias in pulse oximetry and clinical outcomes. JAMA Intern. Med. **182**(7), 699 (2022). https://doi.org/10.1001/jamainternmed.2022.1903
25. Haltaufderheide, J., Ranisch, R.: The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs). npj Digit. Med. **7**(1), 183 (2024). https://doi.org/10.1038/s41746-024-01157-x
26. Guerreiro, N.M., et al.: Hallucinations in large multilingual translation models. Trans. Assoc. Comput. Linguist. **11**, 1500–1517 (2023). https://doi.org/10.1162/tacl_a_00615
27. Puthur, L., Karakunnel, M., Miranda, D., Arora, T.: Building a Data ethical future: data policies and frameworks. J. Dharma **48**(2), 231–248 (2023)
28. Schroeder, D., Chatfield, K., Muthuswamy, V., Kumar, N.K.: Ethics dumping – how not to do research in resource-poor settings. J. Acad. Stand. Against Poverty **1**, 32–55 (2021). https://doi.org/10.5281/zenodo.8089799