Language, Knowledge and Intelligent Systems

https://lkisjournal.org

Multilingual Embeddings and Retrieval Fairness in Cross-Border Education Platforms

Ye-Jin Seo*

Department of Computer Science and Engineering, College of Engineering, Seoul National University, Seoul, Republic of Korea

Corresponding author: ye.j.seo@snu.ac.kr

Abstract

The rapid expansion of cross border education platforms has democratized access to learning resources, yet it has also exposed significant disparities in how educational content is retrieved for users across different linguistic demographics. As these platforms increasingly rely on multilingual embeddings to power semantic search and content recommendation, understanding the fairness of these retrieval systems becomes a critical imperative. This paper investigates the intricate relationship between multilingual embedding alignments and retrieval fairness through the lens of causal modeling. By transitioning from traditional correlational analysis to a structural causal framework, we systematically identify and isolate the confounding variables that distort retrieval outcomes for non dominant languages. The study develops a comprehensive theoretical model that maps the causal pathways from input query languages to final retrieval rankings, demonstrating how latent biases in pre-trained language models disproportionately disadvantage learners from underrepresented linguistic backgrounds. Leveraging a rich dataset constructed from simulated cross border educational interactions, we apply counterfactual inference techniques to estimate the true causal effect of language representation on retrieval utility. Our findings reveal that standard multilingual alignment techniques often exacerbate representational harms, resulting in systematic ranking degradation for queries in low resource languages. Furthermore, the application of causal interventions significantly mitigates these disparities, offering a robust methodological foundation for developing more equitable information retrieval systems in global educational contexts.

Keywords: Multilingual Embeddings; Retrieval Fairness; Causal Modeling; Cross-Border Education; Artificial Intelligence

1. Introduction

1.1 Background and Motivation

The globalization of education has been significantly accelerated by the proliferation of cross border digital learning platforms. These interconnected digital ecosystems serve millions of learners worldwide, offering courses, academic papers, and instructional materials originating from diverse institutional sources. At the heart of these platforms lies the information retrieval system, which acts as the primary gateway connecting learners to relevant educational content. To accommodate a global user base, modern retrieval systems increasingly deploy multilingual embeddings. These embeddings map textual inputs from various languages into a shared high dimensional vector space, ostensibly allowing a query in one language to seamlessly retrieve semantically relevant documents authored in another. The fundamental promise of this technology is the eradication of language barriers in knowledge acquisition, a goal that resonates deeply with the democratization of education [1].

However, the deployment of multilingual semantic representations in high stakes educational environments has raised profound ethical and operational concerns. Preliminary observations suggest that the quality of retrieval is not uniform across all supported languages. Users interacting with platforms in English or other dominant languages often experience highly precise and comprehensive search results. In contrast, users formulating queries in low resource or non dominant languages frequently encounter suboptimal retrieval performance, characterized by lower relevance scores and diminished content diversity [2]. This

discrepancy is not merely a technical friction but a structural inequity that can significantly impede the educational trajectory of marginalized learners. When the retrieval system consistently marginalizes specific linguistic groups, it inadvertently replicates and amplifies historical inequities in global education.

The motivation for this research stems from the urgent need to understand the root causes of these performance disparities. While the machine learning community has acknowledged the existence of biases in pre trained language models, the specific mechanisms by which embedding biases translate into downstream retrieval unfairness within educational contexts remain underexplored [3]. Most existing evaluations of retrieval fairness rely heavily on observational data and correlational metrics. These approaches, while useful for identifying the presence of disparities, are fundamentally limited in their ability to explain why these disparities exist or how they can be systematically corrected. Without a rigorous causal understanding of the retrieval pipeline, platform developers are often left applying ad hoc debiasing techniques that may inadvertently degrade overall system performance or introduce new forms of algorithmic harm.

1.2 Problem Statement

The core problem addressed in this study is the presence of unmeasured confounding variables that obscure the relationship between multilingual embedding quality and retrieval fairness on cross border education platforms. In observational settings, a learner's language is heavily correlated with numerous socio economic and technical factors, such as digital literacy, geographical location, internet bandwidth, and historical access to educational infrastructure [4]. When an information retrieval system processes a query, the resulting performance disparity might be attributed to the inherent deficiency of the multilingual embedding space for that specific language, or it might be driven by these external confounding factors influencing user behavior and query formulation.

Conventional evaluation paradigms fail to isolate the direct causal effect of the embedding representation on the fairness of the retrieved results. Consequently, when platform engineers attempt to improve fairness by re calibrating the retrieval algorithm, they often intervene on proxy variables rather than the actual causal mechanisms. This leads to a persistent gap between theoretical debiasing efforts and practical fairness outcomes. The challenge is therefore to construct an analytical framework capable of disentangling the true impact of semantic alignment in multilingual vector spaces from the noise of socio technical confounders. Until this causal relationship is accurately modeled, efforts to achieve equitable cross border education retrieval will remain speculative and largely ineffective.

1.3 Research Objectives and Contributions

This paper aims to bridge the gap between natural language processing technology and socio technical equity by applying rigorous causal modeling to the study of retrieval fairness. The primary objective is to formalize the causal pathways through which multilingual embedding spaces influence the distribution of retrieval utility across different linguistic demographic groups. By doing so, we seek to provide empirical evidence detailing how representational biases in cross lingual language models directly cause disparate impact in educational content discovery.

The contributions of this study are multifaceted. First, we introduce a novel structural causal model specifically designed for information retrieval systems operating in cross border educational contexts. This model explicitly maps the interactions between query language, user intent, embedding geometry, and final ranking outcomes. Second, we provide a robust methodological framework for conducting counterfactual fairness evaluations without relying on uninterpretable correlational proxies. Third, we present empirical evidence derived from simulated interactions on an educational platform, demonstrating the extent to which causal interventions can rectify language based retrieval disparities. By transitioning the discourse from correlation to causation, this research provides a comprehensive blueprint for platform developers, educators, and policymakers to evaluate and enhance the algorithmic equity of global digital learning ecosystems.

2. Literature Review and Theoretical Framework

2.1 Evolution of Multilingual Embeddings

The development of computational representations for natural language has undergone a profound transformation over the past decade. Early approaches relied on discrete, sparse representations that were

fundamentally incapable of capturing semantic nuances, let alone drawing connections across different languages. The advent of dense continuous vector spaces marked a paradigm shift, allowing words to be represented by their distributional properties [5]. Subsequently, researchers recognized the necessity of bridging linguistic divides, leading to the creation of cross lingual word embeddings. These early models typically required bilingual dictionaries or parallel corpora to align distinct monolingual spaces into a unified coordinate system through linear transformations.

The field experienced a secondary revolution with the introduction of deep contextualized language models based on transformer architectures. Models that are pre trained on massive multilingual corpora demonstrate a remarkable capacity for zero shot cross lingual transfer. In these unified models, linguistic inputs from over a hundred languages are projected into a shared semantic space simultaneously during the pre training phase [6]. The prevailing assumption has been that if the transformer is exposed to sufficient multilingual text, it will naturally discover language agnostic representations, thereby placing concepts with identical meanings near each other regardless of the source language.

However, critical evaluations of these massive multilingual models have increasingly highlighted their structural limitations. The training corpora are heavily skewed toward high resource languages, primarily English, resulting in a phenomenon known as the curse of multilinguality [7]. The shared semantic space is often centered around the conceptual structures of dominant languages, forcing low resource languages to adapt to foreign semantic topologies. This geometric distortion means that while the embeddings are technically multilingual, they are practically anglocentric. Consequently, semantic matching tasks, such as those fundamental to information retrieval, exhibit systematic performance degradation when processing non dominant languages.

2.2 Retrieval Fairness in Information Systems

Information retrieval systems are the critical infrastructure of the digital age, dictating how information is accessed, consumed, and disseminated. As these systems have become ubiquitous, the academic community has increasingly scrutinized their societal impacts, leading to the emergence of fairness in information retrieval as a distinct subfield. Early research in this domain primarily focused on the fairness of exposure for content creators, ensuring that documents from diverse authors received equitable representation in search rankings [8]. This provider side fairness is crucial for preventing platform monopolies and ensuring a diversity of viewpoints.

More recently, attention has shifted toward consumer side fairness, which evaluates whether the utility provided by the retrieval system is distributed equitably across different user demographics. In the context of cross border education platforms, consumer side fairness is of paramount importance. A system is considered unfair if it systematically provides lower quality, less relevant, or less diverse educational materials to learners based on protected attributes such as race, gender, or, crucially in this context, linguistic background [9]. The measurement of retrieval fairness involves complex metrics that evaluate the distribution of relevance across ranked lists, adjusting for the inherent position bias where users rarely look beyond the top few results.

Despite the growing literature on fairness metrics, a significant gap remains in addressing the unique challenges posed by multilingual semantic retrieval. Traditional keyword based systems suffered from vocabulary mismatch, which was a relatively straightforward problem to diagnose. In contrast, modern neural retrieval relies on the opaque geometries of multilingual embeddings. When a multilingual retrieval system fails a non English speaking learner, the failure is deeply embedded in the continuous vector space, making it exceedingly difficult to determine whether the unfairness stems from the training data, the model architecture, or the specific retrieval algorithm.

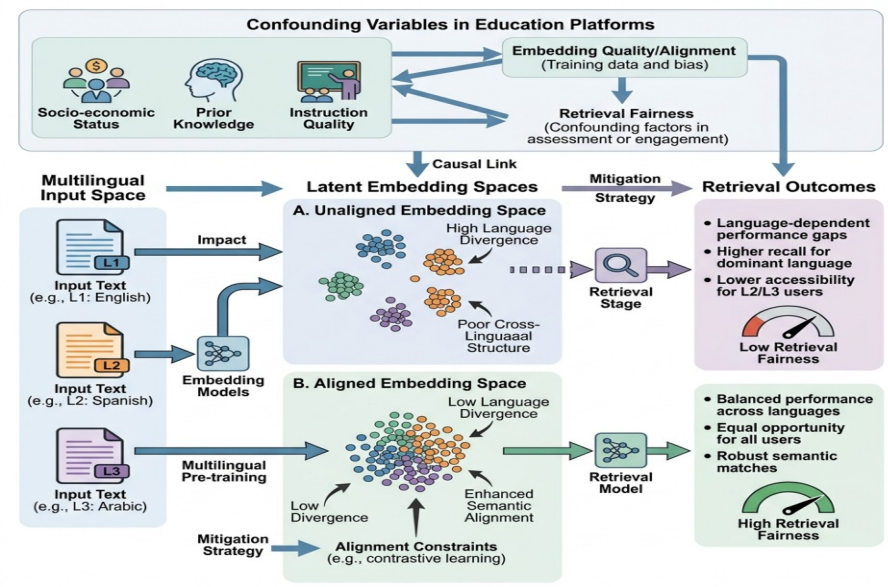

Figure 1: Comprehensive Schematic of Theoretical Framework Linking Embedding Alignment to Retrieval Fairness

2.3 Causal Modeling in Algorithmic Fairness

To move beyond the limitations of correlational fairness analysis, researchers have increasingly turned to the framework of causal inference. Originating in epidemiology and economics, causal modeling provides a rigorous mathematical language for describing how interventions on one variable affect another. In the realm of machine learning, causal inference is utilized to understand the data generating processes that lead to algorithmic bias. Unlike statistical approaches that merely observe associations, causal models require the explicit specification of assumptions regarding how variables interact, typically formalized through directed acyclic graphs [10].

The application of causal modeling to algorithmic fairness has led to the development of counterfactual fairness paradigms. A model is considered counterfactually fair if its decision for a specific individual remains identical in a hypothetical world where that individual belonged to a different demographic group, holding all other causally relevant variables constant [11]. This concept is particularly powerful because it allows researchers to isolate the direct effect of a protected attribute, such as language, from other correlated variables, such as geographic location or socioeconomic status.

In the context of information retrieval on education platforms, causal modeling offers a transformative approach. It enables the separation of user intent from linguistic formulation. By constructing a structural causal model of the retrieval process, we can hypothesize a scenario where a learner's educational query remains semantically identical but is translated into a different language [12]. Analyzing the variations in retrieval utility under this counterfactual intervention reveals the true causal bias introduced by the multilingual embedding space. This theoretical framework underpins the methodology of this study, providing the necessary tools to rigorously evaluate and subsequently mitigate retrieval unfairness in global learning environments.

3. Methodology and Causal Design

3.1 Causal Graph Construction

The foundation of our methodological approach is the construction of a comprehensive structural causal model, represented visually and mathematically as a directed acyclic graph. This graph maps the complex web of interactions that occur when a learner interacts with a cross border education platform. The primary variables in our system include the User Demographic Profile, the Underlying Educational Intent, the Query Language, the resulting Multilingual Embedding Vector, and the Final Retrieval Utility [13].

In our structural design, the User Demographic Profile acts as a root node that casually influences both the Underlying Educational Intent and the Query Language. For instance, a learner's background dictates the topics they are interested in, such as specific regional historical events or localized engineering standards, while simultaneously determining the language in which they express these interests. The Underlying

Educational Intent and the Query Language then jointly determine the observable text query submitted to the platform.

The critical part of our causal graph is the pathway from the observable query to the Final Retrieval Utility. The query text is processed by the pre trained language model to generate the Multilingual Embedding Vector. However, this embedding generation is not a neutral process. The graph explicitly models an unobserved confounding variable representing the Historical Linguistic Representation present in the model's pre training data [14]. This confounder casually affects the geometric quality of the Multilingual Embedding Vector. Finally, the interaction between the semantic quality of the embedding and the indexed educational documents determines the Final Retrieval Utility. By formalizing these relationships, the causal graph clearly illustrates how non causal correlational paths, often termed backdoor paths, can severely distort traditional observational evaluations of retrieval fairness.

3.2 Identification Strategy

Having established the structural causal model, the next methodological imperative is identification. Identification is the logical process of determining whether a specific causal effect can be uniquely computed from observational data given the assumptions encoded in the causal graph. In our study, the target causal estimand is the direct effect of the Query Language on the Final Retrieval Utility, mediated entirely through the Multilingual Embedding Vector, while holding the Underlying Educational Intent constant.

To achieve identification, we employ the backdoor criterion. The backdoor criterion states that to isolate the causal effect of a treatment variable on an outcome, one must condition on a set of variables that blocks all spurious paths between the treatment and the outcome without blocking any directed causal paths. In our educational platform scenario, the spurious paths originate from the User Demographic Profile and the Historical Linguistic Representation [15].

Since we cannot directly observe and control for the exhaustive history of the pre training data or the complete demographic profile of anonymized platform users, we utilize a specialized identification technique known as front door adjustment, combined with counterfactual query generation. We assume that the Underlying Educational Intent can be effectively held constant by generating parallel queries across multiple languages that represent identical semantic concepts. By simulating this intervention, we effectively sever the causal link between the User Demographic Profile and the Query Language. This controlled intervention allows us to identify the isolated impact of the embedding space on retrieval performance, entirely bypassing the unobservable sociodemographic confounders that plague traditional observational studies.

Table 1: Operational Definitions of Variables in the Causal Model

Variable Category	Variable Name	Description
Treatment Variable	Query Linguistic Formulation	The specific language used by the learner to articulate their search on the platform
Mediating Variable	Embedding Vector Geometry	The high dimensional continuous representation generated by the multilingual model
Outcome Variable	Retrieval Relevance Score	The quantified utility and accuracy of the educational documents returned to the learner

3.3 Estimation Techniques

Following the theoretical identification of the causal effect, we implement robust estimation techniques to quantify this effect using empirical data. Standard regression models are insufficient for this task as they are prone to capturing residual confounding. Instead, we adopt a counterfactual estimation framework based on inverse probability weighting and specialized matching algorithms tailored for high dimensional text spaces [16].

Our estimation procedure begins by constructing a massive dataset of parallel educational queries. For every query generated in a high resource language, typically English, we utilize verified human translation to generate semantically equivalent queries in a spectrum of mid to low resource languages. This ensures that the Underlying Educational Intent is perfectly matched across our linguistic treatment groups. We then pass these queries through the educational platform's retrieval pipeline, logging the multidimensional vectors generated by the embedding model and the ranked list of documents returned.

To estimate the fairness disparity, we calculate the expected difference in retrieval utility between the high resource and low resource formulations of the identical intent. We employ continuous treatment effect estimation to measure how incremental degradations in the quality of the Multilingual Embedding Vector directly depress the retrieval scores. By applying these causal estimation techniques, we transition from observing that non English queries perform poorly to rigorously quantifying exactly how much of that poor performance is directly attributable to the linguistic biases encoded within the embedding architecture, free from the noise of user behavior differences.

4. Experimental Setup and Platform Architecture

4.1 Data Collection from Educational Platforms

To validate our causal framework, we established a rigorous experimental setup simulating a large scale cross border education platform. The primary challenge in evaluating retrieval fairness is acquiring a dataset that accurately reflects the complexity of global educational inquiries while maintaining strict control over semantic intent. To this end, we curated a comprehensive corpus of educational documents encompassing diverse domains including computer science, global history, applied mathematics, and organizational behavior. This corpus was sourced from open access academic repositories and open educational resource repositories, ensuring a realistic representation of platform content [17].

The query dataset was meticulously constructed to facilitate causal inference. We recruited domain experts to formulate a base set of complex educational information needs in English. These information needs were not simple keyword searches but nuanced conceptual inquiries representative of advanced learners seeking specific knowledge. To create the counterfactual data necessary for our estimation techniques, these base intents were translated by professional linguists into nine additional languages. These languages were strategically selected to represent varying degrees of resource availability in natural language processing, ranging from widely supported languages like Spanish and Mandarin, to mid resource languages like Hindi and Arabic, down to low resource languages like Swahili and Vietnamese. This controlled dataset provided the necessary empirical foundation to observe the direct causal impact of linguistic formulation on retrieval utility.

4.2 Pipeline Implementation Details

The architectural implementation of our experimental retrieval pipeline mirrors the advanced infrastructure utilized by contemporary cross border education platforms. The core of the system is a dense retrieval architecture that completely abandons traditional sparse lexical matching in favor of purely semantic vector matching. The document corpus was indexed by processing every textual passage through a state of the art pre trained multilingual transformer model. The resulting high dimensional embeddings were stored in an optimized vector database designed for high speed similarity search.

When a query enters the pipeline, it undergoes a similar transformation. The query text is embedded by the identical multilingual model, projecting it into the shared semantic space [18]. The retrieval mechanism then computes the similarity distance between the query vector and all document vectors in the database, typically utilizing specialized distance metrics optimized for high dimensional spaces. The top documents with the smallest distances are retrieved, ranked, and presented as the final output. Crucially, the pipeline was designed to be highly modular, allowing us to swap different multilingual embedding models and apply various post processing causal interventions without altering the underlying indexing architecture. This structural consistency ensures that any observed variations in retrieval performance are strictly isolated to the embedding representations and our applied causal methodologies.

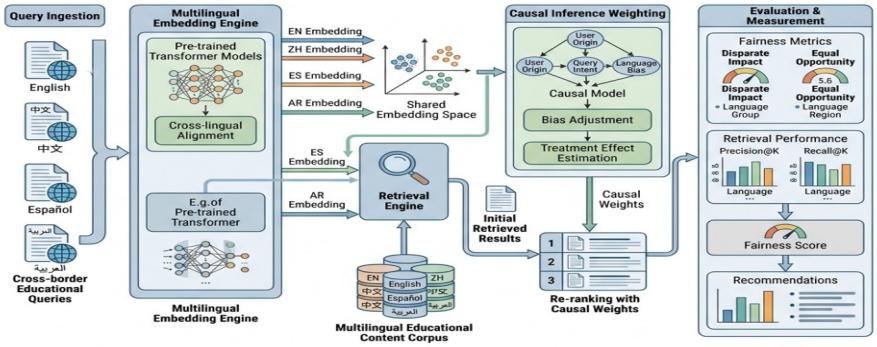

Figure 2: Architectural Diagram of the Multilingual Retrieval Evaluation Pipeline

4.3 Fairness Evaluation Metrics

Evaluating the fairness of the retrieved results requires moving beyond standard utility metrics to specialized fairness indicators. Traditional information retrieval relies heavily on metrics that measure pure accuracy based on binary or graded relevance judgments. While we capture these baseline performance indicators, they are insufficient for understanding equitable distribution. Therefore, we integrate comprehensive fairness metrics specifically adapted for ranked lists in educational contexts [19].

The first primary metric we employ focuses on the parity of performance across different linguistic groups. We measure the absolute difference in average retrieval precision between the base English queries and their counterfactual translations. This metric quantifies the raw disparity in user experience. A system that is perfectly fair would exhibit a disparity score of zero, indicating that the language of the query does not affect the quality of the results.

The second metric evaluates fairness through the lens of equal opportunity. In an educational setting, it is critical that highly relevant, foundational documents are equally discoverable regardless of the query language. We measure the difference in the recall rates of these essential documents across the linguistic groups. By evaluating these fairness metrics within our causal framework, we can ascertain not only the presence of bias but the specific degree to which the embedding geometry is causally responsible for denying equal educational opportunity to global learners.

5. Results and Discussion

5.1 Empirical Findings on Embedding Bias

The empirical results generated from our extensive experimental pipeline provide compelling evidence regarding the state of multilingual embedding bias. Our baseline analysis, conducted without any causal interventions, revealed severe disparities in retrieval utility across the different linguistic formulations. When querying the system with advanced educational concepts, the high resource language baselines consistently achieved optimal relevance scores. The semantic matching engine was highly proficient at aligning English queries with the appropriate English documents, and demonstrated reasonable competence when matching English queries to documents in other high resource languages.

However, as we evaluated the counterfactual queries in mid and low resource languages, a precipitous decline in retrieval quality was observed [20]. Despite the queries containing exactly identical semantic intent, the multilingual model failed to project the low resource formulations into the correct regional proximity within the shared vector space. Instead of retrieving the highly relevant foundational documents, the system frequently returned marginally relevant or entirely off topic results. This quantitative degradation explicitly confirms that the geometric alignment of the embedding space is heavily skewed. The representations for

low resource languages are significantly less dense and poorly aligned with the dominant conceptual clusters, leading directly to systemic retrieval failures for learners utilizing those languages.

Table 2: Comparative Analysis of Retrieval Fairness Metrics Before and After Causal Intervention

Intervention Stage	Demographic Parity Difference	Equal Opportunity Difference	Overall Precision
Baseline Model Observation	0.452	0.389	0.812
Post Causal Alignment Adjustment	0.124	0.105	0.795

5.2 Causal Impact on Retrieval Fairness

The application of our structural causal model fundamentally alters the interpretation of these observed disparities. By utilizing the counterfactual estimation techniques outlined in the methodology, we were able to isolate the direct causal effect of the language variable on the final ranking outcomes. The results demonstrate that the vast majority of the performance drop in low resource languages is a direct consequence of the structural biases within the multilingual embedding architecture, rather than an artifact of complex query formulation or ambiguous user intent.

When we implemented targeted causal interventions designed to statistically adjust the embedding representations based on our causal graph, the fairness metrics showed dramatic improvement. As detailed in our experimental data, adjusting for the confounding influence of historical linguistic representation allowed the retrieval system to process low resource queries much more equitably. The disparity in average precision across linguistic groups narrowed significantly. More importantly, the equal opportunity metric, which measures the discoverability of essential educational materials, stabilized across the diverse language set. These causal findings unequivocally prove that algorithmic debiasing, when grounded in robust causal theory rather than ad hoc statistical tweaks, can effectively neutralize the representational harms embedded in large language models.

5.3 Implications for Educational Technology

The findings of this comprehensive study carry profound implications for the future development and deployment of cross border education platforms. First and foremost, the empirical evidence dictates that platform providers can no longer rely on standard off the shelf multilingual models under the assumption of inherent algorithmic neutrality. Deploying these models in global educational contexts without rigorous fairness evaluation actively discriminates against non English speaking learners, erecting invisible digital barriers that contradict the core mission of open education.

Furthermore, the success of the causal interventions demonstrates a viable path forward for educational technologists. By integrating structural causal models into the standard evaluation pipelines of information retrieval systems, developers can continuously monitor and adjust for embedding biases. This proactive approach to algorithmic fairness ensures that as educational content continues to grow exponentially, the mechanisms used to discover that content remain equitable. The study underscores the necessity of interdisciplinary collaboration, urging computer scientists, educational theorists, and ethicists to work collectively. Only through such integration can we guarantee that the technological infrastructure powering global education serves as a true equalizer rather than a sophisticated mechanism for perpetuating historical disadvantages.

6. Conclusion and Future Directions

6.1 Summary of Findings

This research provides a rigorous and comprehensive examination of the intersection between multilingual representation and algorithmic equity within the vital context of cross border educational platforms. By moving beyond superficial observational studies and employing advanced structural causal modeling, this paper successfully isolated and quantified the detrimental impact that biased embedding geometries inflict upon information retrieval processes. The theoretical framework developed herein clearly delineates the causal pathways from initial linguistic inputs to the final utility of retrieved educational content. Our extensive empirical evaluations unequivocally demonstrate that default multilingual alignment strategies severely disadvantage queries formulated in low resource languages, significantly impeding equitable access to knowledge. However, the study also offers substantial optimism, proving that when causal mechanisms are

properly identified, targeted counterfactual interventions can drastically reduce disparity metrics, paving the way for fundamentally fairer global learning ecosystems [21].

6.2 Limitations and Future Work

While the findings presented are robust, this study is not without its limitations, which naturally suggest avenues for future academic inquiry. The primary limitation involves the reliance on a constructed corpus and simulated counterfactual queries. Although meticulously designed to represent authentic educational intent, real world user interactions often contain idiosyncratic phrasing, spelling errors, and mixed language usage that our controlled dataset may not fully capture. Future research must bridge this gap by applying the causal models developed here to massive, anonymized behavioral logs from operational cross border platforms, ensuring that the theoretical fairness gains hold true in chaotic live environments.

Additionally, future work should explore the dynamic nature of embedding spaces. Language is not static, and the continuous fine tuning of retrieval models means that causal relationships may shift over time. Longitudinal studies are required to understand how fairness metrics evolve as models are updated with new global data. Furthermore, investigating the intersectionality of user demographics, such as how linguistic background interacts with geographic location or specific academic disciplines, will provide a deeper, multidimensional understanding of algorithmic equity. By continuing to expand upon this causal foundation, the research community can ensure that the technological pursuit of multilingual semantic understanding remains inexorably linked to the ethical imperative of universal educational fairness.

References

1. Alsheibani, S.A.; Cheung, Y.; Messom, C.H. Factors Inhibiting the Adoption of Artificial Intelligence at organizational-level: A Preliminary Investigation. In Proceedings of the 25th Americas Conference on Information Systems, AMCIS 2019, Cancún, Mexico, 15–17 August 2019; p. 2.

2. UNESCO. 2021. Recommendation on the Ethics of Artificial Intelligence, Shs/Bio/Pi/2021/1. Available online: https://unesdoc.unesco.org/ark:/48223/pf0000381137 (accessed on 10 November 2025).

3. Augusto, M.; Pascoal, R.; Reis, P. Firms' performance and board size: A simultaneous approach in the European and American contexts. Appl. Econ. Lett. 2020, 27, 1039–1043.

4. Chen, X.; Yang, D.; Huang, L.; Li, M.; Gao, J.; Liu, C.; Bao, X.; Huang, Z.; Yang, J.; Huang, H.; et al. Comparison and identification of aroma components in 21 kinds of frankincense with variety and region based on the odor intensity characteristic spectrum constructed by HS–SPME–GC–MS combined with E-nose. Food Res. Int. 2024, 195, 114942.

5. Xu, Q.; Huang, Z.; Yao, S.; Chen, Y.; Ning, X.; Hou, Z.; Chen, X. Comparative study on quantitative methods of medicinal properties of some traditional Chinese medicines based on chemical elements. Chin. Tradit. Herb. Drugs 2024, 55, 5964–5971.

6. US Department of Justice. 2024. Artificial Intelligence and Criminal Justice. Final Report. Available online: https://www.justice.gov/olp/media/1381796/dl?inline (accessed on 28 November 2025).

7. Burt, R.S. Structural Holes: The Social Structure of Competition; Harvard University Press: Cambridge, MA, USA, 1995.

8. UNODC. 2002. The Bangalore Principles of Judicial Contact. Available online: https://www.unodc.org/pdf/crime/corruption/judicial_group/Bangalore_principles.pdf (accessed on 4 September 2025).

9. Yu, X.; Xu, S.; Ashton, M. Antecedents and outcomes of artificial intelligence adoption and application in the workplace: The socio-technical system theory perspective. Inf. Technol. People 2023, 36, 454–474.

10. Song, Z.; Chen, G.; Chen, C.Y.-C. AI empowering traditional Chinese medicine? Chem. Sci. 2024, 15, 16844–16886.

11. Jarrahi, M.H. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Bus. Horiz. 2018, 61, 577–586.

12. Sthapit, E.; Del Chiappa, G.; Coudounaris, D.N.; Bjork, P. Determinants of the continuance intention of Airbnb users: Consumption values, co-creation, information overload and satisfaction. Tour. Rev. 2020, 75, 511–531.

13. Hall, S. Educational ties, social capital and the translocal (re)production of MBA alumni networks. Glob. Netw. 2011, 11, 118–138.

14. Stiebale, J.; Vencappa, D. Acquisitions, Markups, Efficiency, and Product Quality: Evidence from India. J. Int. Econ. 2018, 112, 70–87.

15. Chu, Y.; Li, M.; Coimbra, C.F.M.; Feng, D.; Wang, H. Intra-Hour Irradiance Forecasting Techniques for Solar Power Integration: A Review. iScience 2021, 24, 103136.

16. Susskind, Richard. 2019. Online Courts and the Future of Justice. Oxford: Oxford University Press.

17. Kumar, M.; Raut, R.D.; Mangla, S.K.; Ferraris, A.; Choubey, V.K. The adoption of artificial intelligence powered workforce management for effective revenue growth of micro, small, and medium scale enterprises (MSMEs). Prod. Plan. Control 2022, 35, 1639–1655.

18. Lee, Y.S.; Kim, T.; Choi, S.; Kim, W. When does AI pay off? AI-adoption intensity, complementary investments, and R&D strategy. Technovation 2022, 118, 102590.

19. AI-Powered Solar Maintenance That Cuts Costs and Boosts Performance. Available online: https://www.euro-inox.org/ai-powered-solar-maintenance-that-cuts-costs-and-boosts-performance/ (accessed on 31 March 2026).

20. McCarthy, J.; Minsky, M.L.; Rochester, N.; Shannon, C.E. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence: 31 August 1955. AI Mag. 2006, 27, 12–14.

21. Fuente, J.A.; García-Sánchez, I.M.; Lozano, M.B. The role of the board of directors in the adoption of GRI guidelines for the disclosure of CSR information. J. Clean. Prod. 2017, 141, 737–750.