University of Qom
Faculty of Literature and Humanities

Research In Instructional Methods

Online ISSN:2981-1074

Homepage: https://jrim.qom.ac.ir

Algorithmic Bias and Educational Justice in the Age of Artificial Intelligence: Social Implications and Policy Solutions in Iran

Faegheh Faghih Moussavi [1], **Faranak Fotouhi Ghazvini** [2]✉

1. Ph.D. Student in Information Technology Engineering, Department of Computer Engineering and Information Technology, Faculty of Engineering, University of Qom, Qom, Iran. E-mail: f.faghihmoussavi@stu.qom.ac.ir
2. Assistant Professor, Department of Computer Engineering and Information Technology, Faculty of Engineering, University of Qom, Qom, Iran.E-mail: f-fotouhi@qom.ac.ir

Article Info

Article type:
Research Article

Article history:
Received 2025-11-13
Received in revised form 2025-12-26
Accepted 2026-01-10
Published online 2026-03-05

Keywords:
algorithmic bias,
Educational justice,
Artificial intelligence.

ABSTRACT

With the expanding use of artificial intelligence—especially large language models (LLMs)—in education, a key question arises: how can these technologies strengthen or undermine educational equity? This study aims to elucidate the dimensions of algorithmic bias within educational systems and to derive social implications and policy responses appropriate to the Iranian context. The present research is qualitative, adopting a documentary–analytical approach. It employs a systematic content analysis of official reports, scholarly articles, and international case studies, guided by a conceptual checklist of educational equity. The study's theoretical framework draws on John Rawls's theory of justice as fairness and Amartya Sen's capability approach. The findings indicate that bias operating at four levels—problem formulation, data, modelling, and interpretation/implementation—can reproduce educational inequalities and, in Iran's diverse context marked by a digital divide, further intensify them. Accordingly, six policy directions are proposed: equity-oriented algorithm design; monitoring and ensuring data diversity; providing algorithmic ethics education for stakeholders; strengthening transparency and accountability; developing indigenous models; and reducing the digital divide. The article's contribution lies in linking theories of justice with the literature on algorithmic bias and in advancing a locally grounded framework for educational equity policymaking in Iran in the age of artificial intelligence.

How To Cite: Faghih Moussavi, F., & Fotouhi Ghazvini, F. (2026). Algorithmic bias and educational justice in the age of artificial intelligence: social implications and policy solutions in Iran, *Research in Instructional Methods*, 3 (5), 196-213. https://doi.org/10.22091/jrim.2026.14502.1436

Publisher: University of Qom
DOI:https://doi.org/10.22091/jrim.2026.14502.1436

Introduction

The integration of artificial intelligence (AI) and, more specifically, large language models (LLMs) into educational systems has marked one of the most transformative developments of the past decade. These technologies, with their capacity for personalized learning, enhanced efficiency, and reduced human error, have opened new horizons in educational practice and policy. However, global experiences have shown that the use of AI in education, when detached from social, cultural, and historical considerations, can reproduce or even exacerbate existing inequalities. Among the most pressing ethical and pedagogical challenges is algorithmic bias, a phenomenon through which unequal data, unfair design, and insufficient human oversight lead to discriminatory decisions and inequitable distribution of learning opportunities.

This study aims to examine the various dimensions of algorithmic bias in educational systems and its implications for educational justice, with a particular focus on the Iranian context. The theoretical framework is grounded in principles of social and educational justice as well as capability-based approaches. From this perspective, educational justice is not merely equality of access or outcomes but the creation of diverse opportunities that enable each learner to realize their human potential within their own social and cultural setting. Accordingly, if designed and managed conscientiously, AI can serve as a powerful tool for empowerment; yet, if implemented without ethical oversight, it risks reinforcing historical patterns of inequality under the guise of technological neutrality.

Methods

This research employs a qualitative, document-based, and content analysis approach. The data include official reports from educational institutions, peer-reviewed academic publications, and international case studies of AI applications in education. The analysis proceeded through three main phases.

In the first phase, global examples such as the Ofqual grading algorithm in the United Kingdom, AI-driven recruitment systems in major technology companies, and large language models like ChatGPT were reviewed as representative cases of algorithmic bias in intelligent systems.

The second phase involved a comparative analysis between these international experiences and the structural characteristics of Iran's educational system, with particular attention to cultural, linguistic, economic, and digital disparities.

In the third phase, the findings were synthesized and categorized thematically to identify the social and policy implications relevant to the Iranian educational landscape. To enhance the reliability of results, triangulation of data sources and expert review from specialists in educational technology, philosophy of justice, and policy studies were employed. This multi-layered analysis made it possible to move beyond mere description toward a deeper interpretation of how technology, power structures, and justice interact within educational systems.

Results

The results reveal that algorithmic bias in education is a complex and multi-dimensional phenomenon that extends far beyond technical errors. It operates at every stage—from problem formulation to final implementation. Four primary types of bias were identified:

1. Problem Definition Bias: Occurs when the goals and variables of an algorithm are shaped by particular value judgments or limited conceptions of educational success, thereby marginalizing certain learners or objectives.
2. Data Bias: Arises from unbalanced or partial datasets that reflect historical inequities, measurement errors, or structural disadvantages, leading algorithms to reproduce those inequities.

3. Modeling and Evaluation Bias: Emerges when optimization criteria focus exclusively on overall accuracy or error minimization while disregarding distributive fairness across different learner groups.
4. Interpretation and Implementation Bias: Appears when algorithmic outputs are accepted as objective truths without critical analysis or contextual interpretation, diminishing the role of human judgment.

These forms of bias, when combined, can significantly deepen educational inequality. For instance, automated grading systems or AI-based academic placement tools, when trained on historically biased data, may unintentionally perpetuate social and regional disparities. Furthermore, the study identified automation bias, a behavioral tendency in which educators and administrators place excessive trust in algorithmic outputs and underestimate the importance of contextual, human-based evaluation. Over time, this tendency risks weakening teachers' professional autonomy and diminishing their interpretive role in student assessment.

In Iran, these challenges are compounded by the country's cultural and linguistic diversity, economic inequalities, and digital divide. Educational algorithms trained on predominantly Western data may fail to reflect local values, languages, and learning needs, resulting in culturally irrelevant or exclusionary outputs. In addition, unequal access to digital infrastructure and internet connectivity limits the benefits of AI for students in underprivileged or rural areas, creating new layers of learning inequality.

The analysis identified six key policy priorities for addressing these issues within Iran's educational context:

1. Designing educational algorithms that integrate distributive justice indicators and cultural-linguistic sensitivity.
2. Monitoring data diversity to ensure the fair representation of all ethnic, linguistic, and regional groups.
3. Providing algorithmic ethics and digital literacy training for educators, policymakers, and developers.
4. Establishing legal frameworks for transparency, accountability, and the right to human review in algorithmic decisions.
5. Supporting the development of localized AI and Persian-language models to mitigate cultural bias and foster relevant educational content.
6. Investing in digital infrastructure and equitable access to technology to narrow the digital divide between urban and rural regions.

Conclusions

The study concludes that achieving educational justice in the age of AI requires more than equal access to technology; it demands the conscious integration of ethical and justice-based principles at every stage of algorithmic design and implementation. Focusing solely on technical accuracy or efficiency cannot ensure fairness; in fact, it may conceal deeper structural inequities beneath the surface of quantitative success. For this reason, policymakers and designers of intelligent educational systems must embed fairness metrics as a core element of algorithm development and evaluation.

In light of these findings, several strategic actions are proposed for Iran's educational system. These include embedding justice indicators in algorithm design, ensuring equitable data representation, integrating ethics education into teacher and policymaker training programs, and enforcing transparency and accountability in algorithmic governance. Moreover, prioritizing the creation of localized AI models using culturally and linguistically diverse data can help reduce dependence on Western content and enhance educational relevance.

The study further emphasizes that equitable access to technology is foundational for all other reforms. Without adequate infrastructure and inclusive access, justice in algorithmic education remains unattainable. Therefore, simultaneous investment in both human capacity and technological systems is essential.

Ultimately, the future of educational justice in Iran—and globally—depends on the balance between technological progress and ethical responsibility. Artificial intelligence should not replace teachers but should instead serve as a tool that strengthens human connection, critical thinking, and pedagogical fairness. When guided by transparency, inclusivity, and sustained human oversight, AI can transform from a reproducer of inequality into a means of empowerment and opportunity for all learners.

Author Contributions

All authors contributed equally to the conceptualization of the article and writing of the original and subsequent drafts.

Data Availability Statement

Data available on request from the authors.

Acknowledgements

The authors would like to thank all participants in the present study

Ethical Considerations

The authors avoided data fabrication, falsification, plagiarism, and misconduct

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Conflict of Interest

The authors declare no conflict of interest.

پژوهش در روش‌های آموزش
شاپا الکترونیکی: ۲۹۸۱-۱۰۷۴

سوگیری الگوریتمی و عدالت آموزشی در عصر هوش مصنوعی، پیامدهای اجتماعی و راهکارهای سیاستی در ایران

فائقه فقیه موسوی[۱]، فرانک فتوحی قزوینی[۲] ✉

۱. دانشجوی دکتری مهندسی فناوری اطلاعات، گروه مهندسی کامپیوتر و فناوری اطلاعات ، دانشگاه قم، قم، ایران. رایانامه: faghihmoussavi@stu.qom.ac.ir
۲. استادیار گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه قم، قم، ایران. (نویسنده مسئول) رایانامه: f-fotouhi@qom.ac.ir

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۴/۰۸/۲۲
تاریخ بازنگری: ۱۴۰۴/۱۰/۰۵
تاریخ پذیرش: ۱۴۰۴/۱۰/۲۰
تاریخ انتشار: ۱۴۰۴/۱۲/۱۴

کلیدواژه‌ها:
سوگیری الگوریتمی، عدالت آموزشی، هوش مصنوعی، مدل‌های زبانی بزرگ، نظام آموزشی..

چکیده

با گسترش کاربرد هوش مصنوعی و به‌ویژه مدل‌های زبانی بزرگ در آموزش، این پرسش مطرح می‌شود که این فناوری‌ها چگونه می‌توانند عدالت آموزشی را تقویت یا تضعیف کنند. هدف این پژوهش، تبیین ابعاد سوگیری الگوریتمی در نظام‌های آموزشی و استخراج پیامدهای اجتماعی و راهکارهای سیاستی متناسب با بافت ایران است. پژوهش حاضر کیفی و از نوع تحلیل اسنادی ـ تحلیلی است و با استفاده از تحلیل محتوای نظام‌مند گزارش‌های رسمی، مقالات علمی و مطالعات موردی بین‌المللی، براساس چک‌لیست مفهومی عدالت آموزشی، انجام شده است. چارچوب نظری مطالعه بر نظریه عدالت به‌مثابه انصافِ جان راولز و رویکرد قابلیت‌های آمارتیا سن استوار است. یافته‌ها نشان می‌دهد سوگیری در چهار سطح تعریف مسئله، داده، مدل‌سازی و تفسیر/ پیاده‌سازی می‌تواند نابرابری‌های آموزشی را بازتولید کند و در بافت متنوع و دارای شکاف دیجیتال ایران این نابرابری‌ها را تعمیق بخشد. بر این اساس، شش محور سیاستی شامل طراحی عدالت‌محور الگوریتم‌ها، پایش تنوع داده‌ها، آموزش اخلاق الگوریتمی به ذی‌نفعان، شفافیت و پاسخ‌گویی، توسعه مدل‌های بومی و کاهش شکاف دیجیتال پیشنهاد می‌شود. نوآوری مقاله در پیوند دادن نظریه‌های عدالت با ادبیات سوگیری الگوریتمی و بسط چارچوبی بومی برای سیاست‌گذاری عدالت آموزشی در عصر هوش مصنوعی در ایران است.

استناد: فقیه‌موسوی، فائقه؛ فتوحی‌قزوینی، فرانک. (۱٤۰٤). سوگیری الگوریتمی و عدالت آموزشی در عصر هوش مصنوعی، پیامدهای اجتماعی و راهکارهای سیاستی در ایران، پژوهش در روش‌های آموزش، ۳ (۵)، ۲۱۳-۱۹۶. https://doi.org/10.22091/jrim.2026.14502.1436

DOI: https://doi.org/10.22091/jrim.2026.14502.1436
ناشر: دانشگاه قم

مقدمه

ورود هوش مصنوعی و به‌ویژه مدل‌های زبانی بزرگ به نظام‌های آموزشی، افق‌های تازه‌ای برای شخصی‌سازی یادگیری، افزایش بهره‌وری و کاهش خطای انسانی گشوده است (Holmes & Miao, 2023 ; Nabipour Gisi et al., 2024). سامانه‌های توصیه‌گر آموزشی، ابزارهای تصحیح خودکار، و چت‌بات‌های یادگیری می‌توانند به معلمان در هدایت یادگیرندگان و به دانش‌آموزان در دسترسی به حمایت فردی کمک کنند. با این حال، تجربه‌های جهانی نشان داده است که به‌کارگیری این فناوری‌ها بدون توجه به عدالت آموزشی، می‌تواند نابرابری‌های موجود را نه‌تنها بازتولید، بلکه تعمیق کند (Baker & Hawn, 2022; Bulathwela et al., 2024) .

در این مقاله سه مفهوم کلیدی در کانون بحث قرار دارند: «عدالت آموزشی»، «سوگیری الگوریتمی» و «مدل‌های زبانی بزرگ». عدالت آموزشی به‌طور خلاصه به فراهم‌سازی فرصت‌های یادگیری معنادار و باکیفیت برای همه یادگیرندگان، با توجه به تفاوت‌های اجتماعی، اقتصادی و فرهنگی آنان اشاره دارد (Holmes & Miao, 2023; Nazari et al., 2022). سوگیری الگوریتمی به وضعیتی اطلاق می‌شود که در آن، سیستم‌های هوشمند به‌طور سیستماتیک به نفع یا ضرر گروهی خاص عمل می‌کنند، بدون آن‌که تفاوت واقعی در شایستگی وجود داشته باشد (Baker & Hawn, 2022; Boateng & Boateng, 2025). مدل‌های زبانی بزرگ نیز خانواده‌ای از سامانه‌های هوش مصنوعی مبتنی بر معماری ترنسفورمر هستند که بر حجم عظیمی از داده‌های متنی ـ عمدتاً غربی و غیربومی ـ آموزش دیده‌اند و در تولید و تحلیل محتوا نقشی فزاینده دارند (Guo et al., 2024; Tao et al., 2024). ترکیب این سه عنصر در نظام آموزشی، پرسش‌های جدی درباره عدالت، بازنمایی فرهنگی و پیامدهای اجتماعی برمی‌انگیزد.

مطالعات بین‌المللی متعددی به سوگیری الگوریتمی در حوزه‌های حساس مانند عدالت کیفری، استخدام و آموزش پرداخته‌اند. برای نمونه، ماجرای الگوریتم نمره‌دهی Ofqual در بریتانیا نشان داد که تکیه بر داده‌های تاریخی می‌تواند نمرات دانش‌آموزان مناطق محروم را به‌طور ناعادلانه کاهش دهد (Hern, 2020; Mallett, 2023). تجربه ابزار استخدامی آمازون، تبعیض سیستماتیک علیه زنان را آشکار کرد (Dastin, 2022)، و تحلیل‌ها درباره الگوریتم COMPAS[1] نشان داد که متهمان سیاه‌پوست با نرخ بالاتری به اشتباه «پرخطر» برچسب‌گذاری می‌شوند (Angwin et al., 2022). در حوزه آموزش، پژوهش‌هایی مانند (Baker & Hawn, 2022; Bulathwela et al., 2024; Kizilcec & Lee, 2022). به این نتیجه رسیده‌اند که الگوریتم‌های آموزشی، حتی در صورت دقت بالا، ممکن است برای گروه‌های مختلف عملکرد نابرابر داشته باشند و در غیاب شاخص‌های عدالت، نابرابری‌ها را تشدید کنند.

در زمینه مدل‌های زبانی بزرگ نیز پژوهش‌هایی نشان داده‌اند که این مدل‌ها، ارزش‌ها و الگوهای فرهنگی خاصی را بازتاب می‌دهند و در مواجهه با نام‌ها یا هویت‌های اقلیت، پاسخ‌های تبعیض‌آمیز یا کم‌کیفیت‌تر ارائه می‌کنند (An et al., 2024; Cheng et al., 2024; Guo et al., 2024). در ایران، مطالعاتی مانند نبی پور گیسی و نبی (Nabi et al., 2024; Nabipour Gisi et al., 2024) بر ظرفیت‌های هوش مصنوعی برای کاهش نابرابری آموزشی تأکید کرده‌اند، اما هم‌زمان هشدار داده‌اند که در غیاب داده‌های متنوع از اقوام، زبان‌ها و مناطق مختلف کشور، الگوریتم‌ها می‌توانند شکاف‌های منطقه‌ای و فرهنگی را بازتولید کنند. با این حال، هنوز چارچوبی منسجم که نظریه‌های عدالت (مانند راولز[2] و سن[3] را با ادبیات سوگیری الگوریتمی پیوند دهد و براساس آن، پیامدهای اجتماعی و سیاستی هوش مصنوعی در نظام آموزشی ایران را تحلیل کند، در دسترس نیست. این «گپ پژوهشی» به‌ویژه در حوزه مدل‌های زبانی بزرگ و کاربرد آن‌ها در آموزش زبان فارسی محسوس است.

1. Correctional Offender Management Profiling for Alternative Sanctions
2. John Rawls
3. Amartya Sen

اهمیت و ضرورت این پژوهش از دو جهت قابل تبیین است. از منظر نظری، ترکیب دیدگاه‌های عدالت به‌مثابه انصافِ راولز (Rawls, 2017) و رویکرد قابلیت‌های آمارتیا سن (Esmer, 2021)، امکان ارزیابی انتقادی فناوری‌های آموزشی را فراتر از معیارهای کارایی فنی فراهم می‌کند و به پرسش از این‌که «چه کسانی از هوش مصنوعی سود می‌برند؟ و چه کسانی به حاشیه رانده می‌شوند؟» پاسخ می‌دهد (Esmer, 2021; Rawls, 2017; Sarafa & Oyewole, 2023; Sen, 2008). از منظر عملی و سیاستی، تنوع فرهنگی و زبانی، نابرابری‌های منطقه‌ای و شکاف دیجیتال در ایران (Nabipour Gisi et al., 2024; Nazari et al., 2022) سبب می‌شود که هرگونه استفاده شتاب‌زده از الگوریتم‌ها در نمره‌دهی، هدایت تحصیلی یا تخصیص منابع آموزشی، خطر تشدید محرومیت گروه‌های حاشیه‌ای را در پی داشته باشد. بنابراین، تحلیل عدالت‌محور سوگیری الگوریتمی، پیش‌نیاز شکل‌گیری سیاست‌های مسئولانه در حوزه «هوش مصنوعی و آموزش» است.

بر این اساس، هدف پژوهش حاضر آن است که با تکیه بر نظریه‌های عدالت راولز و سن (Esmer, 2024; Rawls, 2017)، ادبیات سوگیری الگوریتمی و نمونه‌های جهانی، چارچوبی مفهومی برای تحلیل پیامدهای اجتماعی هوش مصنوعی در آموزش ارائه دهد و بر مبنای آن، راهکارهای سیاستی متناسب با نظام آموزشی ایران پیشنهاد کند. از نظر روش‌شناختی، پژوهش حاضر کیفی و از نوع تحلیل اسنادی ـ تحلیلی است و داده‌های آن از گزارش‌های رسمی، مقالات علمی و مطالعات موردی بین‌المللی گردآوری و با استفاده از تحلیل محتوای نظام‌مند و چک‌لیست مفهومی، عدالت آموزشی بررسی شده‌اند.

در پرتو این مباحث، پرسش‌های پژوهش به‌صورت زیر صورت‌بندی می‌شوند:

۱. سوگیری الگوریتمی در نظام‌های آموزشی، با تأکید بر مدل‌های زبانی بزرگ، چگونه و در چه سطوحی می‌تواند عدالت آموزشی را تهدید یا تضعیف کند؟

۲. با توجه به ویژگی‌های فرهنگی، زبانی و ساختاری نظام آموزشی ایران، چه اصول و راهبردهای سیاستی برای کاهش سوگیری الگوریتمی و تقویت عدالت آموزشی در استفاده از هوش مصنوعی قابل پیشنهاد است؟

مبانی نظری پژوهش

تحقیقات اخیر، به‌ویژه در حوزه آموزش، شواهد فزاینده‌ای از تأثیر منفی سوگیری الگوریتمی بر عدالت آموزشی ارائه داده‌اند. بیکر و هان (Baker & Hawn, 2022) با تحلیل تجربی چندین الگوریتم آموزشی، نشان داده‌اند که حتی الگوریتم‌هایی با دقت بالا می‌توانند برای گروه‌های مختلف نتایج نابرابر تولید کنند. آن‌ها خواستار استفاده از شاخص‌های توزیعی به‌جای معیارهای صرفاً میانگین‌محور در آموزش هوشمند شده‌اند. داستین (Dastin, 2022) تجربه آمازون در حوزه استخدام را بررسی کرد که الگوریتم غربال رزومه‌ها، رزومه‌های زنان را به‌طور سیستماتیک رد می‌کرد. این مطالعه نشان داد که داده‌های تاریخی می‌توانند منشأ اصلی سوگیری باشند. انگوین و همکاران (Angwin et al., 2022) در مطالعه‌ای کلاسیک نشان دادند که الگوریتم قضایی COMPAS، متهمان سیاه‌پوست را با نرخ بالاتری به اشتباه «پرخطر» ارزیابی می‌کند. چنگ و همکاران (Cheng et al., 2024) در بررسی سوگیری در مدل‌های زبانی بزرگ، تأکید می‌کنند که منابع آموزشی ورودی این مدل‌ها مانند ChatGPT اغلب مبتنی بر داده‌های غربی هستند. تائو و همکاران (Tao et al., 2024) نیز نشان داده‌اند که مدل‌های زبانی مانند GPT-4 عمدتاً ارزش‌ها و پاسخ‌های مشابه فرهنگ‌های اروپای شمالی را بازتاب می‌دهند که در آموزش چندفرهنگی ممکن است به حذف یا بی‌اهمیتی دیگر فرهنگ‌ها منجر شود. در ایران نبی پورگیسی و همکاران (Nabipour Gisi et al., 2024)، هشدار می‌دهند که در غیاب داده‌های متنوع از مناطق و اقوام مختلف کشور، هرگونه استفاده از الگوریتم‌های آموزشی می‌تواند به تعمیق نابرابری منطقه‌ای منجر شود.

۱. عدالت آموزشی

عدالت آموزشی به معنای فراهم‌سازی فرصت‌های یادگیری برابر برای همه یادگیرندگان، صرف‌نظر از پیشینه اجتماعی، اقتصادی یا فرهنگی آن‌ها، با تأکید بر دسترسی عادلانه به فناوری، محتوای بدون سوگیری و مهارت‌های دیجیتالی ضروری است (Holmes & Miao, 2023). جان راولز[1] در نظریه «عدالت به‌مثابه انصاف[2]» به برابری حقوق و آزادی‌های اساسی و بهبود وضعیت محروم‌ترین گروه‌ها تأکید می‌کند (Sarafa & Oyewole, 2023). آمارتیا سن[3] نیز با "رویکرد قابلیت‌ها[4]"، عدالت را در توانمندسازی افراد برای تحقق توانایی‌هایشان می‌بیند (Esmer, 2021). این دیدگاه‌ها معیارهایی برای ارزیابی فناوری‌های آموزشی ارائه می‌دهند. برای مثال، یک الگوریتم آموزشی عادلانه باید منابع را به‌گونه‌ای تخصیص دهد که تا حد امکان شکاف‌های موجود کاهش یابد. در ایران، عدالت آموزشی با چالش‌هایی مانند نابرابری منطقه‌ای و کمبود زیرساخت‌های دیجیتال مواجه است (Nabipour Gisi et al., 2024). نابرابری آموزشی در ایران نتیجه ترکیبی از عوامل ساختاری درون نظام آموزشی و نابرابری‌های اجتماعی-اقتصادی است (Nazari et al., 2022). نظریه‌های راولز و سن (Esmer, 2021; Rawls, 2017) نشان می‌دهند که فناوری‌ها باید به گونه‌ای طراحی شوند که نه‌تنها برابری فرصت‌ها را تضمین کنند، بلکه توانمندی‌های فردی را نیز تقویت کنند. این چارچوب برای تحلیل سوگیری الگوریتمی و ارائه راهکارهای سیاستی در این مقاله به کار می‌رود.

۲. سوگیری الگوریتمی

سوگیری الگوریتمی[5] به وضعیتی اطلاق می‌شود که سیستم‌های خودکار به‌طور سیستماتیک به نفع یا ضرر گروهی خاص عمل کنند بدون آن‌که چنین تفاوتی در شایستگی‌های واقعی افراد وجود داشته باشد (Boateng & Boateng, 2025). این پدیده که از داده‌های آموزشی نامتوازن، طراحی نادرست یا معیارهای بهینه‌سازی غیرمنصفانه ناشی می‌شود ممکن است ناشی از چهار عامل کلیدی باشد (Baker & Hawn, 2022):

سوگیری در تعریف مسئله[6]: این نوع سوگیری ناشی از تصمیم‌گیری‌های مفهومی و ارزشی در مرحله تعریف هدف الگوریتم است، جایی که نحوه صورت‌بندی مسئله و انتخاب متغیرهای هدف می‌تواند برداشت‌های خاصی از مطلوبیت آموزشی را تثبیت یا حذف کند. در این مرحله، نگاه طراح به آنچه «موفقیت» یا «پیشرفت آموزشی» تلقی می‌شود، می‌تواند مسیر کل سامانه را تعیین کند. اگر ارزش‌های عدالت‌محور در این نقطه لحاظ نشوند، تمام فرآیند یادگیری ماشینی بر پایه تعریفی ناقص از عدالت بنا خواهد شد.

سوگیری در داده[7]: این سوگیری از داده‌هایی نشئت می‌گیرد که بازتاب‌دهنده نابرابری‌های ساختاری، خطاهای نمونه‌گیری یا رویه‌های سنجش مغرضانه‌اند و به‌صورت نظام‌مند موجب انتقال تعصبات پیشین به مدل‌های یادگیری می‌شوند. به بیان دیگر، داده‌ها نه آینه‌ای خنثی، بلکه محصول تاریخ و سیاست آموزشی‌اند. بدون پالایش انتقادی داده‌های ورودی، حتی بهترین الگوریتم‌ها نیز به ابزار بازتولید تبعیض تبدیل می‌شوند.

سوگیری در مدل‌سازی و ارزیابی[8]: در این مرحله، سوگیری به‌واسطه انتخاب تابع هدف، معیارهای عملکرد، و تصمیم‌گیری‌های فنی با بار ارزشی وارد می‌شود؛ انتخاب‌هایی که پیامدهای مستقیم بر نحوه بهینه‌سازی، تعمیم‌پذیری و عدالت الگوریتم دارند. به طور

1. John Rawls
2. Justice as Fairness
3. Amartya Sen
4. Capability Approach
5. Algorithmic Bias
6. Biases in problem specification
7. Biases in data
8. Biases in modeling & validation

مثال، تمرکز بیش از حد بر دقت آماری ممکن است موجب نادیده‌گرفتن گروه‌های کوچک اما آسیب‌پذیر شود. بنابراین، طراحی مدل باید علاوه بر کارایی ریاضی، با اصول اخلاقی و عدالت آموزشی هم‌تراز باشد.

سوگیری در تفسیر و پیاده‌سازی[1]: این نوع سوگیری زمانی پدید می‌آید که نحوه استفاده از الگوریتم با زمینه طراحی‌شده آن ناسازگار باشد، یا خروجی‌های آن بدون تحلیل انتقادی و درک صحیح از محدودیت‌ها واقعیت قطعی تلقی شوند. در چنین شرایطی، خطر جایگزینی قضاوت انسانی با تصمیم ماشینی افزایش می‌یابد و پاسخ‌گویی در فرآیندهای آموزشی تضعیف می‌شود. برای پیشگیری، کاربران باید آموزش ببینند تا خروجی الگوریتم را به‌عنوان «ابزار تصمیم‌سازی»، نه «داور نهایی»، تلقی کنند.

در آموزش، سوگیری می‌تواند به تخصیص نابرابر منابع، ارزیابی تبعیض‌آمیز یا بازتولید کلیشه‌های فرهنگی منجر شود. پژوهش بولستولا و همکاران (Bulathwela et al., 2024) هشدار می‌دهد که اگر ابزارهای هوش مصنوعی بدون توجه به زمینه‌های لازم پیاده‌سازی شوند، ممکن است به طور غیر عمدی نابرابری‌های آموزشی را عمیق‌تر کنند، سوگیری‌ها را تقویت کنند و نابرابری‌ها را بیشتر کنند. برای نمونه، اگر داده‌های آموزشی عمدتاً از گروه‌های شهری جمع‌آوری شوند، الگوریتم‌ها ممکن است نیازهای دانش‌آموزان روستایی را نادیده بگیرند. در ایران، تنوع زبانی و فرهنگی این مشکل را تشدید می‌کند، زیرا مدل‌های زبانی آموزش‌دیده بر داده‌های غیربومی ممکن است محتوای نامناسب تولید کنند (Nazari et al., 2022). در زمینه آموزش، این نوع سوگیری‌ها به‌ویژه نگران‌کننده‌اند، زیرا خروجی‌های ناعادلانه الگوریتم‌ها می‌توانند به نابرابری در دسترسی به فرصت‌های تحصیلی، تخصیص منابع یا حتی شکل‌گیری تصویر منفی از خود در دانش‌آموزان منجر شوند (Baker & Hawn, 2022). سوگیری الگوریتمی نه‌تنها عدالت آموزشی را تضعیف می‌کند، بلکه اعتماد عمومی به فناوری‌ها را نیز کاهش می‌دهد. در چنین شرایطی، هوش مصنوعی به جای ابزاری برای کاهش شکاف‌های آموزشی، می‌تواند به عاملی برای تعمیق تبعیض بدل شود. بنابراین، رویکردهای عدالت‌محور در طراحی و پیاده‌سازی سیستم‌های آموزشی مبتنی بر هوش مصنوعی ضروری است. همچنین، ایجاد چارچوب‌های اخلاقی برای نظارت بر عملکرد الگوریتم‌ها می‌تواند از پیامدهای ناخواسته پیشگیری کند. از سوی دیگر، آموزش ذی‌نفعان درباره ماهیت سوگیری و روش‌های شناسایی آن می‌تواند نقش مهمی در افزایش آگاهی و توانمندی کاربران ایفا کند. در نهایت، ترکیب داده‌های بومی، سیاست‌گذاری شفاف و مشارکت فعال معلمان و پژوهشگران می‌تواند مسیر حرکت به سوی نظام آموزشی عادلانه‌تر را هموار سازد.

۳. نمونه‌های جهانی سوگیری الگوریتمی

۳-۱. الگوریتم Ofqual

در سال ۲۰۲۰، الگوریتم Ofqual در انگلستان برای استانداردسازی نمرات امتحانات به دلیل اتکا به داده‌های تاریخی مدارس، به‌طور ناعادلانه نمرات دانش‌آموزان کم‌برخوردار را کاهش داد (Mallett, 2023). این الگوریتم از عملکرد گذشته مدارس برای پیش‌بینی نمرات استفاده کرد، اما داده‌های تاریخی نابرابری‌های ساختاری را بازتولید کردند. نتیجه، کاهش فرصت‌های دانش‌آموزان مناطق محروم برای ورود به دانشگاه بود که اعتراضات گسترده‌ای را برانگیخت. این مورد نشان می‌دهد که الگوریتم‌های بدون نظارت انسانی می‌توانند به تبعیض سیستماتیک منجر شوند. در ایران، استفاده از الگوریتم‌های مشابه برای کنکور یا ارزیابی تحصیلی می‌تواند نابرابری‌های منطقه‌ای را تشدید کند. این موارد بر لزوم شفافیت و نظارت در کاربرد هوش مصنوعی تأکید می‌کند.

1. Biases in deployment

۳-۲. ابزار استخدامی آمازون

الگوریتم استخدامی آمازون به دلیل آموزش بر رزومه‌های مردسالارانه، علیه زنان تبعیض قائل شد و در نهایت کنار گذاشته شد (Dastin, 2022). داده‌های آموزشی این الگوریتم از رزومه‌های مردان در صنعت فناوری تشکیل شده بود، که منجر به امتیازدهی پایین‌تر به زنان شد. این مورد اهمیت کیفیت و تنوع داده‌های آموزشی را نشان می‌دهد. در آموزش، الگوریتم‌های مشابه ممکن است به تبعیض علیه گروه‌های حاشیه‌ای منجر شوند، مانند دانش‌آموزان اقلیت‌های قومی در ایران. آمازون پس از شناسایی سوگیری، الگوریتم را بازطراحی کرد، اما این تجربه نشان داد که پیشگیری از سوگیری نیازمند آزمون‌های عدالت پیش از اجرا است. این نمونه بر لزوم طراحی الگوریتم‌هایی با داده‌های منعکس‌کننده ترکیب واقعی جمعیت و معیارهای عادلانه تأکید می‌کند.

۳-۳. الگوریتم قضاییCOMPAS

الگوریتم قضایی COMPAS که در ایالات متحده برای پیش‌بینی خطر تکرار جرم مورد استفاده قرار می‌گیرد، یکی از نمونه‌های شاخص سوگیری الگوریتمی در تصمیم‌گیری‌های حساس اجتماعی است. تحلیل‌های مستقل (Angwin et al., 2022) نشان دادند که این الگوریتم، با وجود دقت آماری نسبتاً بالا، نرخ بالاتری از برچسب‌گذاری نادرست[1] را برای متهمان سیاه‌پوست نسبت به سفیدپوستان تولید می‌کرد. این نابرابری آماری، که ریشه در داده‌های آموزشی مغرضانه داشت، تضاد میان معیارهای فنی عدالت و الزامات عدالت را برجسته می‌سازد و لزوم بازنگری در مفاهیم سنتی ارزیابی الگوریتم‌ها در حوزه‌های حساس را نشان می‌دهد.

۳-۴. رفتار مدل های زبانی بزرگ مانند ChatGPT

مطالعات نشان داده‌اند که ChatGPT در پاسخ به اسامی با بار نژادی متفاوت، پیشنهادهای تبعیض‌آمیزی ارائه می‌کند و گرایش به بازتولید ارزش‌های غربی و کلیشه‌های جنسیتی دارد (An et al., 2024). برای مثال، این مدل ممکن است برای اسامی مرتبط با اقلیت‌ها پاسخ‌های کم‌کیفیت‌تر ارائه دهد. در ایران، مدل‌های زبانی آموزش‌دیده بر داده‌های غربی ممکن است محتوای نامناسب با فرهنگ محلی تولید کنند (Nabi et al., 2024). کاهش چنین سوگیری‌هایی نیازمند توسعه مدل‌های بومی و نظارت بر خروجی‌ها است. این مورد نشان می‌دهد که بدون توجه به تنوع فرهنگی، هوش مصنوعی می‌تواند نابرابری‌ها را تقویت کند و اثرات سو بر نتایج حاصل داشته باشد.

پژوهش‌های پیشین در زمینه سوگیری الگوریتمی

پژوهش‌های بین‌المللی

بخش قابل‌توجهی از ادبیات جهانی به بررسی سوگیری الگوریتمی در سامانه‌های آموزشی و پیامدهای آن برای عدالت پرداخته است (Baker & Hawn, 2022; Boateng & Boateng, 2025; Kizilcec & Lee, 2022) که نشان داده‌اند: حتی زمانی که دقت کلی الگوریتم‌های پیش‌بینی عملکرد، هدایت تحصیلی یا سیستم‌های توصیه‌گر بالاست، خطای آن‌ها میان گروه‌های مختلف یادگیرندگان (برحسب جنسیت، قومیت یا وضعیت اقتصادی) یکسان نیست. این مطالعات تأکید می‌کنند که معیارهای میانگین‌محور مانند دقت[2] به‌تنهایی کافی نیستند و باید در کنار آن‌ها شاخص‌های عدالت توزیعی و برابری فرصت‌ها نیز به کار گرفته شود. مقاله بوثولا و همکاران (Bulathwela et al., 2024) با نقد رویکرد «تکنو-سولوشنیستی[3]»، استدلال می‌کند که اتکا به هوش مصنوعی بدون توجه به زمینه‌های

1. false positives
2. Accuracy
3. Techno-Solutionist

ساختاری نابرابر، می‌تواند به تعمیق شکاف‌های آموزشی منجر شود. راهنمای یونسکو نیز بر همین نکته تأکید دارد که بهره‌گیری از هوش مصنوعی در آموزش باید همراه با ملاحظات اخلاقی و عدالت‌محور باشد (Holmes & Miao, 2023).

در کنار این خط پژوهشی، مطالعات موردی متعددی، سوگیری الگوریتمی را در تصمیم‌گیری‌های حساس مستند کرده‌اند. ماجرای الگوریتم نمره‌دهی Ofqual در بریتانیا طی دوران همه‌گیری کووید-۱۹ نشان داد که تکیه بر داده‌های تاریخی مدارس می‌تواند به کاهش ناعادلانه نمرات دانش‌آموزان مناطق محروم و اعتراضات گسترده منجر شود (Hern, 2020; Mallett, 2023). ابزار استخدامی آمازون نیز، به‌دلیل آموزش بر رزومه‌های مردسالارانه، به‌طور سیستماتیک امتیاز زنان را کاهش داد و نهایتاً کنار گذاشته شد (Dastin, 2022). در حوزه عدالت کیفری، تحلیل انگوین و همکاران (Angwin et al., 2022) از الگوریتم قضایی COMPAS نشان داد که متهمان سیاه‌پوست با نرخ بالاتری نسبت به سفیدپوستان به اشتباه «پرخطر» برچسب‌گذاری می‌شوند. وجه مشترک این مطالعات آن است که الگوریتم‌ها، در صورت آموزش بر داده‌های نابرابر یا تعریف مسئله مغرضانه، نابرابری‌های ساختاری پیشین را بازتولید می‌کنند.

ادبیات تازه‌ای نیز بر سوگیری در مدل‌های زبانی بزرگ تمرکز دارد. گوئو و همکاران (Guo et al., 2024)، تائو و همکاران (Tao et al., 2024)، چنگ و همکاران (Cheng et al., 2024) و آن و همکاران (An et al., 2024) نشان داده‌اند که مدل‌های زبانی بزرگ، از جمله ChatGPT، اغلب ارزش‌ها و هنجارهای فرهنگی غربی را بازتاب می‌دهند، نسبت به نام‌ها و هویت‌های اقلیت پاسخ‌های کم‌کیفیت‌تر یا کلیشه‌ای تولید می‌کنند و می‌توانند کلیشه‌های جنسیتی و نژادی را تقویت کنند. این مطالعات هشدار می‌دهند که در بافت‌های چندفرهنگی، استفاده آموزشی از مدل‌های زبانی بدون سازوکارهای تعدیل و بومی‌سازی، ممکن است به حاشیه‌راندن فرهنگ‌ها و زبان‌های غیرغربی بینجامد. در مجموع، پژوهش‌های بین‌المللی نشان می‌دهند که سوگیری الگوریتمی در آموزش پدیده‌ای واقعی و چندسطحی است، اما عمده این شواهد در کشورهای صنعتی و نظام‌های آموزشی غربی تولید شده و کمتر به بسترهای غیرغربی و دارای شکاف دیجیتال پرداخته شده است.

پژوهش‌های داخلی

در ایران، بخش مهمی از ادبیات به توصیف و تحلیل نابرابری‌های آموزشی در سطح ساختار نظام آموزشی می‌پردازد. نظری و همکاران (Nazari et al., 2022) با استفاده از دیدگاه‌های خبرگان و معلمان، نشان داده‌اند که نابرابری آموزشی در ایران چندبعدی است و از ترکیب عواملی مانند تفاوت در کیفیت مدارس، توزیع نامتوازن منابع، تمرکزگرایی، شرایط اقتصادی و شکاف‌های منطقه‌ای شکل می‌گیرد. این یافته‌ها نشان می‌دهد که حتی پیش از ورود هوش مصنوعی، زمینه‌ای از نابرابری ساختاری در نظام آموزشی وجود دارد که می‌تواند بر هر فناوری جدیدی، از جمله الگوریتم‌ها، سایه بیندازد.

در سال‌های اخیر، برخی مطالعات داخلی به‌طور خاص‌تر به نسبت میان هوش مصنوعی و عدالت آموزشی پرداخته‌اند. نبی‌پورگیسی (Nabipour Gisi et al., 2024) با تمرکز بر ظرفیت‌ها و چالش‌های هوش مصنوعی در کاهش نابرابری‌ها، تأکید می‌کند که بدون دسترسی عادلانه به زیرساخت دیجیتال و بدون تنوع داده‌ای از مناطق و اقوام مختلف، الگوریتم‌های آموزشی می‌توانند شکاف‌های منطقه‌ای و طبقاتی را تشدید کنند. نبی (Nabi et al., 2024) نیز با تمرکز بر نقش هوش مصنوعی در کاهش تبعیض آموزشی، بر ضرورت توسعه مدل‌های بومی و توجه به ویژگی‌های فرهنگی-زبانی ایران تأکید می‌کند. با این حال، این مطالعات عمدتاً ماهیتی مفهومی و هشداردهنده دارند و کمتر به تحلیل نظام‌مند سطوح سوگیری الگوریتمی (در تعریف مسئله، داده، مدل و تفسیر) و پیوند آن با نظریه‌های عدالت پرداخته‌اند.

جمع‌بندی پژوهش‌های پیشین نشان می‌دهد که در سطح بین‌المللی، سوگیری الگوریتمی در آموزش و مدل‌های زبانی بزرگ تا حد زیادی مستند شده است، اما این ادبیات در بافت نظام‌های آموزشی غربی متمرکز است و کمتر به شرایط کشورهایی با تنوع زبانی،

فرهنگی و شکاف دیجیتال بالا پرداخته است. در ایران نیز اگرچه نابرابری آموزشی و پیامدهای بالقوه هوش مصنوعی مورد بحث قرار گرفته است (Nabi et al., 2024; Nabipour Gisi et al., 2024; Nazari et al., 2022)، هنوز مطالعه‌ای که به‌طور منسجم نظریه‌های عدالت راولز و سن (Esmer, 2021; Rawls, 2017) را با ادبیات سوگیری الگوریتمی و شواهد جهانی ترکیب کند و از دل آن، چارچوبی مفهومی و مجموعه‌ای از راهکارهای سیاستی مشخص برای استفاده از هوش مصنوعی و به‌ویژه مدل‌های زبانی بزرگ در نظام آموزشی ایران ارائه دهد، گزارش نشده است. پژوهش حاضر درصدد پر کردن این خلأ است.

روش‌شناسی پژوهش

پژوهش حاضر از نظر هدف، توصیفی-تحلیلی و از نظر ماهیت، کیفی و مبتنی بر تحلیل اسنادی و تحلیل محتوای جهت‌دار است. در این نوع مطالعه، «داده‌ها» به‌صورت ثانویه و از دل اسناد مکتوب (مقالات علمی، کتاب‌ها، گزارش‌های سیاستی و اسناد بین‌المللی) استخراج می‌شوند و تمرکز بر تبیین مفهومی پدیده‌ها و پیامدهای آن‌هاست، نه آزمون آماری فرضیه‌ها.

جامعه اسنادی و ملاک انتخاب منابع

جامعه اسنادی این پژوهش شامل کلیه منابع علمی و اسناد بالادستی و گزارش‌های رسمی نهادهای سیاست‌گذار آموزشی منتشرشده به زبان‌های فارسی و انگلیسی است که به یکی از حوزه‌های «هوش مصنوعی و آموزش»، «سوگیری الگوریتمی»، «عدالت آموزشی» و «وضعیت نابرابری آموزشی در ایران» می‌پردازند. برای انتخاب نمونه، از نمونه‌گیری هدفمند استفاده شد. معیارهای ورود منابع عبارت بودند از:

- ارتباط مستقیم با حداقل یکی از مفاهیم عدالت آموزشی، سوگیری الگوریتمی، مدل‌های زبانی بزرگ یا سیاست‌گذاری آموزشی
- برخورداری از اعتبار علمی انتشار در مجلات علمی - پژوهشی، کتاب‌های دانشگاهی یا گزارش‌های نهادهای معتبر مانند یونسکو و OECD
- دسترسی کامل به متن
- انتشار ترجیحاً در دو دهه اخیر، به‌استثنای متون کلاسیک نظری مانند (Rawls, 2017; Sen, 2008) که به‌طور هدفمند به‌عنوان مبنای نظری وارد شدند.

در گام نخست، با استفاده از جست‌وجوی کلیدواژه‌ای و مرور فهرست منابع مقالات کلیدی، فهرست اولیه‌ای از اسناد تهیه شد. سپس با حذف موارد تکراری یا کم‌ارتباط، در نهایت تعداد نهایی اسناد، ۲۲ سند برای تحلیل انتخاب گردید.

روش جمع‌آوری داده‌ها

داده‌ها به‌صورت ثانویه و از طریق جست‌وجوی نظام‌مند در پایگاه‌های علمی بین‌المللی مانند Google Scholar وScopus، در حد دسترسی و پایگاه‌های فارسی مانند SID و نورمگز، و نیز مرور هدفمند گزارش‌های سازمان‌های بین‌المللی یونسکو، OECDو اسناد ملی در حوزه آموزش در ایران گردآوری شد. در جست‌وجو از ترکیب کلیدواژه‌های فارسی و انگلیسی مانند: سوگیری الگوریتمی در آموزش، عدالت آموزشی و هوش مصنوعی، algorithmic bias in education, educational justice,

large language models, AI in education in Iran استفاده گردید. علاوه بر این، از روش گلوله‌برفی[1] نیز بهره گرفته شد؛ بدین معنا که از فهرست منابع مقالات و گزارش‌های کلیدی، منابع مرتبط جدید شناسایی و به فهرست اضافه شدند.

ابزار و چارچوب تحلیل اسناد

ابزار اصلی پژوهش یک چک‌لیست مفهومی عدالت آموزشی در عصر هوش مصنوعی بود که توسط پژوهشگر و براساس ترکیب نظریه عدالت به‌مثابه انصافِ راولز، رویکرد قابلیت‌های آمارتیا سن، و ادبیات سوگیری الگوریتمی در آموزش (Baker & Hawn, 2022; Boateng & Boateng, 2025; Holmes & Miao, 2023; Kizilcec & Lee, 2022) طراحی شد. این چک‌لیست شامل محورهای زیر بود:

- سطح و نوع سوگیری الگوریتمی (در تعریف مسئله، داده، مدل‌سازی/ارزیابی، تفسیر/پیاده‌سازی)
- تأثیر تصمیمات الگوریتمی بر عدالت توزیعی و برابری فرصت‌های آموزشی
- اثر بر قابلیت‌ها و توانمندی‌های یادگیرندگان (در چارچوب رویکرد قابلیت‌ها)
- میزان حساسیت فرهنگی-زبانی و نحوه بازنمایی اقلیت‌ها و گروه‌های حاشیه‌ای
- نوع پاسخ سیاستی، مقرراتی یا راهکار اصلاحی پیشنهادشده در هر منبع.

این چک‌لیست به‌عنوان ابزار استخراج و سازمان‌دهی داده‌ها از اسناد به کار رفت و نقش چارچوب تحلیلی را در کل مطالعه ایفا کرد.

شیوه تجزیه و تحلیل داده‌ها

برای تحلیل داده‌ها از تحلیل محتوای کیفی با رویکرد استقرایی - قیاسی استفاده شد. در گام نخست، متن کامل اسناد منتخب چندین بار مطالعه شد و واحدهای معنایی مرتبط با سوگیری، عدالت و پیامدهای آموزشی شناسایی و به‌صورت کدهای اولیه ثبت گردید. در گام دوم، این کدها با استفاده از چک‌لیست مفهومی و چارچوب نظری راولز و سن، در قالب مقوله‌های بالاتر مانند «بازتولید نابرابری‌های ساختاری»، «تضاد میان دقت الگوریتمی و عدالت آموزشی»، «سوگیری اتوماسیون» و «ضررهای تخصیصی و بازنمایی» گروه‌بندی شدند. در گام سوم، مقوله‌ها در دو سطح «یافته‌های جهانی» و «پیامدهای خاص برای ایران» سازمان‌دهی و به‌صورت روایی در بخش‌های یافته‌ها و بحث گزارش شد.

برای تقویت روایی و پایایی تحلیل، از راهبردهایی مانند استفاده از منابع متنوع علمی، اسناد بالادستی، بین‌المللی و بومی، مقایسه دیدگاه‌های چند نویسنده درباره یک پدیده و بازبینی مکرر دسته‌بندی‌ها براساس چارچوب نظری استفاده شد. در نتیجه، اگرچه مطالعه از نظر ماهیت تفسیری است، اما تلاش شده است استنتاج‌ها بر پایه شواهد مستند و آشکار تکیه داشته باشد.

یافته‌ها

نمای کلی فرایند تحلیل

همان‌گونه که در بخش روش‌شناسی اشاره شد، متن کامل ۲۲ سند منتخب چند بار خوانده شد و واحدهای تحلیل به‌صورت جملات یا بندهایی که به یکی از محورهای «سوگیری الگوریتمی»، «عدالت آموزشی»، «مدل‌های زبانی بزرگ» یا «سیاست‌گذاری آموزشی» اشاره داشتند، استخراج گردید. این واحدهای معنایی ابتدا به‌صورت کدهای باز ثبت شدند مانند «استفاده از داده‌های تاریخی نابرابر در

1. snowball

Ofqualنرخ خطای بیشتر برای اقلیت‌ها در COMPAS الگوریتم استخدامی تبعیض‌آمیز، بازنمایی کلیشه‌ای زنان در متن‌های تولیدی LLM، اعتماد بیش از حد معلمان به خروجی سیستم و .... در گام بعد، این کدها براساس چک‌لیست مفهومی عدالت آموزشی و چهار سطح سوگیری **تعریف مسئله، داده، مدل‌سازی - ارزیابی و تفسیر - پیاده‌سازی** تجمیع و در قالب مضامین بالاتر سازمان‌دهی شدند. نتیجه این فرایند، استخراج چهار مضمون اصلی بود که در اکثر مستندات تکرار می‌شدند:

۱. بازتولید نابرابری‌های ساختاری،
۲. تضاد میان دقت الگوریتمی و عدالت آموزشی،
۳. سوگیری اتوماسیون و تضعیف قضاوت انسانی،
٤. ضررهای تخصیصی و بازنمایی.

جدول ۱ این مضامین، تعریف مفهومی هر یک و نمونه‌هایی از واحدهای تحلیل را به‌طور خلاصه نشان می‌دهد.

جدول ۱. مضامین اصلی استخراج‌شده از تحلیل اسنادی

مضمون اصلی	تعریف خلاصه	نمونه واحد تحلیل / سند
بازتولید نابرابری‌های ساختاری	موقعیت‌هایی که الگوریتم‌ها نابرابری‌های پیشین (منطقه‌ای، طبقاتی، جنسیتی...) را از طریق داده یا منطق تصمیم تکرار می‌کنند.	گزارش Ofqual درباره کاهش نامتناسب نمرات دانش‌آموزان مدارس محروم (Hern, 2020; Mallett, 2023)؛ تحلیل Baker & Hawn, 2022)از عملکرد نابرابر الگوریتم‌های آموزشی برای گروه‌های مختلف.
تضاد میان دقت الگوریتمی و عدالت آموزشی	مواردی که الگوریتم از نظر دقت کلی قابل قبول است، اما خطا یا پیامد آن برای برخی گروه‌ها ناعادلانه است.	تحلیل COMPAS و تفاوت نرخ خطای مثبت کاذب برای متهمان سیاه‌پوست (Angwin et al., 2022)؛ بحث Kizilcec & Lee, 2022) درباره ضرورت شاخص‌های عدالت در کنارAccuracy.
سوگیری اتوماسیون و تضعیف قضاوت انسانی	موقعیت‌هایی که کاربران به‌طور افراطی به خروجی سیستم اعتماد می‌کنند و نقش قضاوت انتقادی انسانی کاهش می‌یابد.	گزارش‌ها درباره تمایل تصمیم‌گیران عمومی به تبعیت از توصیه الگوریتم حتی در صورت شک به خطا Alon-Barkat & Busuioc, 2023; Carragher et al., 2024).
ضررهای تخصیصی و بازنمایی	پیامدهای ناعادلانه در تخصیص فرصت‌ها - منابع و بازتولید کلیشه‌ها در محتواهای تولیدی.	مثال‌های مربوط به حذف رزومه زنان در الگوریتم استخدامی آمازون (Dastin, 2022)؛ بازنمایی کلیشه‌ای جنسیتی و فرهنگی در پاسخ‌های مدل‌های زبانی بزرگ (An et al., 2024; Guo et al., 2024; Tao et al., 2024).

بازتولید نابرابری‌های ساختاری

نخستین مضمون غالب در اسناد، بازتولید نابرابری‌های ساختاری بود. در بسیاری از متون، الگوریتم‌ها نه به‌عنوان ابزار خنثی، بلکه به‌عنوان سازوکارهایی توصیف شده‌اند که نابرابری‌های موجود در داده‌ها و ساختارهای آموزشی را تثبیت می‌کنند. برای مثال، در تحلیل ماجرای Ofqual، گزارش‌های هِرن و مالِت (Hern, 2020; Mallett, 2023) نشان می‌دهند که الگوریتم استانداردسازی نمرات، با تکیه بر عملکرد تاریخی مدارس، نمرات دانش‌آموزان مدارس کم‌برخوردار را به‌طور نامتناسب کاهش داده و در نتیجه، فرصت ورود آن‌ها به دانشگاه را محدود کرده است. این واحدهای تحلیل تحت کدهایی مانند «اتکای الگوریتم به دادهٔ تاریخی نابرابر» و «کاهش فرصت‌های تحصیلی برای مناطق محروم» ثبت شدند.

به‌طور مشابه، بیکر و هان (Baker & Hawn, 2022) در تحلیل چندین الگوریتم آموزشی نشان می‌دهند که مدل‌هایی با دقت کلی بالا، برای گروه‌های دارای سابقه عملکرد ضعیف‌تر (که اغلب از اقشار کم‌برخوردار هستند) خطای بیشتری تولید می‌کنند. این الگو

در مطالعه بولستولا و همکاران (Bulathwela et al., 2024) نیز دیده می‌شود؛ جایی که تأکید می‌شود ابزارهای هوش مصنوعی در صورت بی‌توجهی به زمینه اجتماعی، به «تثبیت‌کننده» شکاف‌های موجود بدل می‌شوند. در چارچوب عدالت آموزشی، این یافته‌ها نشان می‌دهد که بدون بازنگری انتقادی در داده‌های ورودی و منطق تصمیم‌گیری، الگوریتم‌ها می‌توانند نابرابری‌های منطقه‌ای و طبقاتی را بازتولید و به نسل‌های بعد منتقل کنند؛ وضعیتی که با ساختار نابرابر نظام آموزشی ایران (Nazari et al., 2022) هم‌پوشانی مفهومی دارد.

تضاد میان دقت الگوریتمی و عدالت آموزشی

مضمون دوم، تضاد میان دقت الگوریتمی و عدالت آموزشی بود. در بخشی از اسناد، الگوریتم‌ها از نظر شاخص‌های فنی مانند Accuracy یا خطای میانگین، «موفق» ارزیابی شده‌اند، اما همان اسناد نشان می‌دهند که توزیع خطا میان گروه‌های مختلف یادگیرندگان نابرابر است. تحلیل کلاسیک انگوین و همکاران (Angwin et al., 2022) از الگوریتم قضایی COMPAS نمونه‌ای از این وضعیت است: این الگوریتم در مجموع از نظر پیش‌بینی تکرار جرم عملکرد قابل قبولی دارد، اما نرخ «مثبت کاذب» برای متهمان سیاه‌پوست به‌طور معناداری بالاتر از سفیدپوستان گزارش شده است. در کدگذاری، این موارد زیر برچسب‌هایی مانند «دقت کلی بالا، بی‌عدالتی توزیعی» و «فاصله میان معیارهای فنی و معیارهای عدالت» قرار گرفتند.

در ادبیات آموزشی نیز بیکر و هاون (Baker & Hawn, 2022) و کیزیلسک و لی (Kizilcec & Lee, 2022) تأکید می‌کنند که تمرکز صرف بر دقت متوسط می‌تواند گروه‌های کوچک اما آسیب‌پذیر را نادیده بگذارد. این اسناد به‌گونه‌ای کدگذاری شدند که نشان می‌دهند معیارهای سنتی ارزیابی مدل، با دیدگاه‌های عدالت راولز (Rawls, 2017) (بهبود وضعیت محروم‌ترین‌ها) و رویکرد قابلیت‌های سن (توجه به توانمندسازی واقعی افراد) در تعارض قرار می‌گیرد. بر این اساس، مضمون «تضاد دقت – عدالت» حاصل ادغام کدهایی بود که به شکاف میان موفقیت فنی الگوریتم و پیامدهای نابرابر آن برای گروه‌های مختلف اشاره داشتند.

سوگیری اتوماسیون و تضعیف قضاوت انسانی

مضمون سوم، سوگیری اتوماسیون و تضعیف قضاوت انسانی است. در تحلیل مطالعات (Alon-Barkat & Busuioc, 2023)، واحدهای تحلیلی متعددی شناسایی شد که نشان می‌دادند افراد در مواجهه با توصیه‌های الگوریتمی، تمایل دارند حتی در صورت تردید نسبت به صحت آن، به‌طور خودکار از خروجی تبعیت کنند. این پدیده تحت کدهایی مانند «اعتماد افراطی به سیستم»، «نادیده‌گرفتن شواهد زمینه‌ای» و «کاهش نقش قضاوت حرفه‌ای» طبقه‌بندی شد.

در حوزه آموزش، این الگو به معنای آن است که معلمان، مشاوران یا مدیران ممکن است خروجی سامانه‌های هدایت تحصیلی، پیش‌بینی عملکرد یا تخصیص منابع را به‌عنوان «حکم نهایی» تلقی کنند و از سنجش انتقادی وضعیت خاص هر دانش‌آموز غفلت ورزند. برخی اسناد سیاستی (Holmes & Miao, 2023) نیز به‌طور مستقیم هشدار می‌دهند که استفاده از هوش مصنوعی در آموزش، بدون تقویت سواد الگوریتمی و مهارت‌های نقد فناوری در میان ذی‌نفعان، می‌تواند به کاهش استقلال حرفه‌ای معلمان و تضعیف نقش تربیتی آن‌ها منجر شود. مجموعه این کدها، مضمون «سوگیری اتوماسیون» را شکل داد که در بحث نتایج، با دغدغه حفظ «جایگاه قضاوت انسانی» در نظریه‌های عدالت پیوند داده می‌شود.

ضررهای تخصیصی و بازنمایی

چهارمین مضمون، «ضررهای تخصیصی و بازنمایی» است که دو سطح متمایز از آسیب را پوشش می‌دهد:

اول: تخصیص ناعادلانه فرصت‌ها، منابع و تصمیم‌های آموزشی میان گروه‌های مختلف یادگیرندگان مانند پذیرش دانشگاهی، دسترسی به دوره‌های پیشرفته یا حمایت‌های جبرانی؛

دوم: بازنمایی کلیشه‌ای، تبعیض‌آمیز یا حاشیه‌ای گروه‌ها در محتواهای تولیدی و خروجی مدل‌های زبانی، به‌گونه‌ای که هویت، فرهنگ یا توانمندی برخی گروه‌ها کم‌رنگ، تحریف‌شده یا در نقش‌های محدود و کلیشه‌ای نشان داده می‌شود.

در سطح نخست، مطالعات مربوط به ابزار استخدامی آمازون (Dastin, 2022) و برخی الگوریتم‌های پذیرش یا نمره‌دهی نشان می‌دهد که داده‌های آموزشی با سوگیری جنسیتی یا طبقاتی می‌توانند به «حذف سیستماتیک[1]» گروه‌هایی مانند زنان یا اقلیت‌ها از فرایند تصمیم‌گیری منجر شوند. در کدگذاری، این موارد زیر برچسب‌هایی چون «حذف نامرئی» و «ضرر تخصیصی» ثبت شدند. در سطح دوم، اسنادی مانند (An et al., 2024; Guo et al., 2024; Tao et al., 2024) نشان می‌دهند که مدل‌های زبانی بزرگ در تولید مثال‌ها، توصیف نقش‌های اجتماعی یا پاسخ به نام‌های دارای بار نژادی – جنسیتی، تمایل دارند نقش‌های سنتی و کلیشه‌ای مانند تمرکز بر مردان در نقش‌های علمی و زنان در نقش‌های مراقبتی را بازتولید کنند. این واحدهای تحلیل با کدهایی مانند «بازنمایی کلیشه‌ای»، «غیبت فرهنگ‌های غیرغربی» و «حذف نمادین اقلیت‌ها» دسته‌بندی شدند.

با تجمیع این کدها، مضمون «ضررهای تخصیصی و بازنمایی» شکل گرفت که نشان می‌دهد سوگیری الگوریتمی تنها در سطح عددی و تخصیص فرصت‌ها رخ نمی‌دهد، بلکه بر خودتصوری یادگیرندگان، احساس تعلق فرهنگی و توقعات آن‌ها از آینده تحصیلی نیز تأثیر می‌گذارد. این مضمون، به‌ویژه در بافت ایران که تنوع فرهنگی و زبانی بالاست و مدل‌های زبانی غالباً بر داده‌های غیربومی آموزش دیده‌اند (Nabi et al., 2024; Nabipour Gisi et al., 2024)، اهمیت مضاعف پیدا می‌کند.

بحث و نتیجه‌گیری

براساس تحلیل اسنادی و تحلیل محتوای کیفی، این پژوهش نشان داد که سوگیری الگوریتمی در آموزش، پدیده‌ای محدود به خطای فنی یا نقص در طراحی یک مدل خاص نیست، بلکه در چند سطح به عدالت آموزشی آسیب می‌زند. مضامین استخراج‌شده در بخش یافته‌ها - شامل بازتولید نابرابری‌های ساختاری، تعارض میان دقت الگوریتمی و عدالت آموزشی، سوگیری اتوماسیون و تفکیک میان ضررهای تخصیصی و بازنمایی - در مجموع ترسیم می‌کنند که چگونه هوش مصنوعی می‌تواند نابرابری‌های پیشین را در قالبی فناورانه بازتولید کند. نمونه‌هایی مانند الگوریتم نمره‌دهیOfqual، ابزار استخدامی آمازون و الگوریتم قضایی COMPAS نشان می‌دهند که اتکای ساده‌انگارانه به داده‌های تاریخی و شاخص‌های دقت میانگین می‌تواند به کاهش فرصت‌های آموزشی و شغلی برای گروه‌های کم‌برخوردار منجر شود (Angwin et al., 2022; Baker & Hawn, 2022; Dastin, 2022; Hern, 2020; Mallett, 2023).

این الگو هنگامی نگران‌کننده‌تر می‌شود که آن را در کنار شواهد داخلی درباره نابرابری‌های منطقه‌ای، تفاوت کیفیت مدارس و شکاف دیجیتال در ایران قرار دهیم (Nabipour Gisi et al., 2024; Nazari et al., 2022)، زیرا در چنین زمینه‌ای هر الگوریتم آموزشیِ «بی‌توجه به عدالت» با احتمال بالا همان ساختارهای نابرابر را در تصمیم‌های خود بازتاب خواهد داد.

این یافته‌ها در پرتو چارچوب‌های نظری عدالت، معنا و عمق بیشتری پیدا می‌کنند. از منظر ، نظام عادلانه نظامی است که هم حقوق پایه را برای همه تضمین می‌کند و هم وضعیت محروم‌ترین گروه‌ها را بهبود می‌بخشد(Rawls, 2017; Sarafa & Oyewole, 2023). رویکرد قابلیت‌های آمارتیا سن نیز عدالت را در امکان بالفعل‌سازی توانمندی‌های افراد می‌بیند، نه در صرف برابری صوری یا میانگین‌های آماری (Esmer, 2021; Sen, 2008). مقایسه این چارچوب‌ها با شواهد مرور شده نشان می‌دهد که بسیاری از کاربردهای

1. systematic exclusion

فعلی هوش مصنوعی - حتی زمانی که از نظر دقت کلی موفق‌اند - با این معیارهای عدالت هم‌راستا نیستند، زیرا خطای بیشتری برای گروه‌های حاشیه‌ای تولید می‌کنند یا دسترسی آن‌ها به فرصت‌های آموزشی را محدود می‌سازند. به بیان دیگر، «موفقیت فنی» الگوریتم‌ها لزوماً به «عدالت آموزشی» منجر نمی‌شود و بدون وارد کردن شاخص‌های عدالت توزیعی، ارزیابی صرفاً براساس دقت می‌تواند گمراه‌کننده باشد.

وجه دیگر نتایج، به سطح کنش انسانی و تفسیر الگوریتم‌ها مربوط است. شواهد مربوط به سوگیری اتوماسیون نشان داد که معلمان، مدیران و سیاست‌گذاران در مواجهه با سامانه‌های هوش مصنوعی ممکن است تصمیم‌های الگوریتمی را به‌طور پیش‌فرض معتبرتر از قضاوت انسانی تلقی کنند و در عمل نقش انتقادی خود را به «تاییدکننده خروجی سیستم» تقلیل دهند. در کنار این، مطالعات اخیر درباره مدل‌های زبانی بزرگ نشان می‌دهد که این مدل‌ها، بر اثر آموزش بر داده‌های عمدتاً غربی، گرایش به بازتولید کلیشه‌های جنسیتی و فرهنگی دارند و حضور زبان‌ها و فرهنگ‌های غیرغربی را به‌طور محدودتر بازنمایی می‌کنند. ترکیب این دو روند در زمینه‌ای مانند ایران با تنوع زبانی و فرهنگی بالا می‌تواند هم در سطح تخصیص فرصت‌ها مانند پیشنهاد دوره‌ها، بورسیه‌ها یا سطح‌بندی تحصیلی و هم در سطح بازنمایی نمادین مانند تصویری که دانش‌آموزان از خود و دیگران می‌سازند پیامدهای نابرابر ایجاد کند. این نتایج نشان می‌دهد که حفظ نقش قضاوت حرفه‌ای معلم و حساسیت فرهنگی در استفاده از مدل‌های زبانی، پیش‌شرط استفاده مسئولانه از هوش مصنوعی در آموزش است، نه مسئله‌ای حاشیه‌ای.

در سطح سیاست‌گذاری، تحلیل اسناد حاکی از آن است که پاسخ به مسئله سوگیری الگوریتمی نیازمند مداخله هم‌زمان در چند لایه است. در سطح ملی، نهادهایی مانند وزارت آموزش‌وپرورش، سازمان سنجش آموزش کشور و شوراهای سیاست‌گذار آموزشی می‌توانند تدوین مقررات شفاف برای استفاده از الگوریتم‌ها در تصمیم‌های حساس همچون نمره‌دهی، هدایت تحصیلی، پذیرش دانشگاهی را در دستور کار قرار دهند و الزام به گزارش دهی درباره خطا و عملکرد الگوریتم‌ها در گروه‌های مختلف جمعیتی را نهادینه کنند. در سطح نهادی، دانشگاه‌ها و ادارات آموزش‌وپرورش استانی می‌توانند بر تنوع داده‌های آموزشی، نمایندگی اقوام و زبان‌ها در مجموعه‌داده‌ها، و آموزش سواد الگوریتمی به معلمان و کارشناسان نظارت کنند. در سطح مدرسه و کلاس درس نیز، نقش معلم به‌عنوان «کاربرِ آگاه و نقاد» اهمیت می‌یابد؛ بدین معنا که معلمان خروجی‌های سامانه‌های هوش مصنوعی را نقطه شروع گفت‌وگو و تصمیم‌گیری بدانند، نه جایگزین قضاوت حرفه‌ای خود. این تقسیم کار چندسطحی، راهکارهای مقاله را از سطح توصیه‌های کلی به سطح پیشنهادهای قابل‌پیگیری برای بازیگران واقعی نظام آموزشی نزدیک‌تر می‌کند.

به عنوان جمع‌بندی، مقاله حاضر با تکیه بر تحلیل اسنادی نظام‌مند، سوگیری الگوریتمی را از سطح یک مسئله تکنیکی به سطح یک مسئله عدالت‌محور و سیاستی ارتقا می‌دهد و آن را در چارچوب شرایط ایران بازتفسیر می‌کند. نوآوری اصلی پژوهش در دو بعد است: **نخست**، صورت‌بندی چهار خاستگاه کلیدی سوگیری (تعریف مسئله، داده‌ها، مدل‌سازی و تفسیر - پیاده‌سازی) در پیوند با عدالت آموزشی؛ **دوم**، ترجمه این چارچوب به بستر ایرانی با توجه به نابرابری‌های آموزشی، شکاف دیجیتال و چالش‌های فرهنگی-زبانی.

براساس تحلیل انجام‌شده و در همین راستا، چند راهکار کلیدی برای نظام آموزشی ایران پیشنهاد می‌شود:

- طراحی الگوریتم‌های آموزشی با در نظر گرفتن عدالت توزیعی، برابری فرصت‌ها و حساسیت فرهنگی - زبانی.
- پایش مداوم تنوع داده‌های آموزشی برای تضمین حضور برابر اقوام، زبان‌ها و مناطق کشور.
- آموزش مفاهیم اخلاق الگوریتمی و سواد دیجیتال به معلمان، دانش‌آموزان و سیاست‌گذاران.
- تدوین مقررات الزام‌آور برای شفافیت در تصمیم‌گیری‌های الگوریتمی و ایجاد سازوکار اعتراض و بازبینی انسانی.

- حمایت از توسعه مدل‌های زبانی و هوش مصنوعی بومی با داده‌های محلی برای کاهش سوگیری فرهنگی.
- سرمایه‌گذاری در زیرساخت دیجیتال و دسترسی برابر به فناوری در مناطق محروم به منظور کاهش شکاف دیجیتال.

بر این اساس، نتیجه‌گیری می‌شود که استفاده عادلانه از هوش مصنوعی در آموزش ایران تنها زمانی امکان‌پذیر است که طراحی و ارزیابی الگوریتم‌ها صریحاً شامل شاخص‌های عدالت توزیعی و حساسیت فرهنگی باشد، داده‌های آموزشی نماینده همه گروه‌ها باشند، و جایگاه قضاوت انسانی در تصمیم‌گیری‌های آموزشی حفظ و تقویت شود. در غیر این صورت، این خطر وجود دارد که سامانه‌های هوشمند، به‌جای کاهش نابرابری‌ها، شکاف‌های موجود را در قالبی ظاهراً خنثی و فناورانه بازتولید کنند.

منابع

Alon-Barkat, S., & Busuioc, M. (2023). Human–AI interactions in public sector decision making:“automation bias” and “selective adherence” to algorithmic advice. *Journal of Public Administration Research and Theory*, *33*(1), 153-169.

An, H., Acquaye, C., Wang, C., Li, Z., & Rudinger, R. (2024). Do Large Language Models Discriminate in Hiring Decisions on the Basis of Race, Ethnicity, and Gender? *arXiv preprint arXiv:2406.10486*.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2022). Machine bias. In *Ethics of data and analytics* (pp. 254-264). Auerbach Publications.

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 1-41.

Boateng, O., & Boateng, B. (2025). Algorithmic bias in educational systems: Examining the impact of AI-driven decision making in modern education. *World Journal of Advanced Research and Reviews*, *25*(1), 2012-2017.

Bulathwela, S., Pérez-Ortiz, M., Holloway, C., Cukurova, M., & Shawe-Taylor, J. (2024). Artificial intelligence alone will not democratise education: On educational inequality, techno-solutionism and inclusive tools. *Sustainability*, *16*(2), 781.

Carragher, D. J., Sturman, D., & Hancock, P. J. (2024). Trust in automation and the accuracy of human–algorithm teams performing one-to-one face matching tasks. *Cognitive Research: Principles and Implications*, *9*(1), 41.

Cheng, H., Guo, Y., Guo, Q., Yang, M., Gan, T., & Nie, L. (2024). Social debiasing for fair multi-modal llms. *arXiv preprint arXiv:2408.06569*.

Dastin, J. (2022). Amazon scraps secret AI recruiting tool that showed bias against women. In *Ethics of data and analytics* (pp. 296-299). Auerbach Publications.

Esmer, S. (2021). *Amartya Sen's capability approach and its relation with John Rawls 'justice as fairness*. Middle East Technical University.

Guo, Y., Guo, M., Su, J., Yang, Z., Zhu, M., Li, H., Qiu, M., & Liu, S. S. (2024). Bias in large language models: Origin, evaluation, and mitigation. *arXiv preprint arXiv:2411.10915*.

Hern, A. (2020). Ofqual's A-level algorithm: Why did it fail to make the grade. *The Guardian*, *21*.

Holmes, W., & Miao, F. (2023). *Guidance for generative AI in education and research*. UNESCO Publishing.

Kizilcec, R. F., & Lee, H. (2022). Algorithmic fairness in education. In *The ethics of artificial intelligence in education* (pp. 174-202). Routledge.

Mallett, B. (2023). Reviewing the impact of OFQUAL's assessment 'algorithm'on racial inequalities. In *COVID-19 and Racism* (pp. 187-198). Policy Press.

Nabi, d., Shahraki, H., Ghofran Mazloom, I., & Absalan, R. (2024). Artificial Intelligence and Reducing Educational Discrimination. *The First National Conference on Modern Perspectives on Educational Issues*.

Nabipour Gisi, E., Ahmadi, A., Darabi, J., & Sharifi, R. (2024). *Artificial intelligence and educational equity: how can technology reduce inequalities?* The First National Conference on New Approaches to Educational Issues, Ramshir.

Nazari, F., Pirootiaghdam, M., & Zovko, M.-E. (2022). Educational inequalities in Iran based on the viewpoints of educational experts and qualified high school teachers. *Distinctio: Journal of Intersubjective Studies*, *1*(2), 73-93.

Rawls, J. (2017). A theory of justice. In *Applied ethics* (pp. 21-29). Routledge.

Sarafa, O. I., & Oyewole, S. (2023). John Rawls on the theory of justice. *Classical Theorists in the Social Sciences: From Western Ideas to African Realities*, 347-375.

Sen, A. (2008). The idea of justice. *Journal of Human Development*, *9*(3), 331-342.

Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS nexus*, *3*(9), pgae346.