https://research.adra.ac.id/index.php/ijlul/
P - ISSN: 3026-7102
E - ISSN: 3030-8372

The Ethics of AI in Language Assessment: A Critical Examination of Algorithmic Bias in Automated Speaking and Writing Tests

Roni Subhan[1], **Alejandro Díaz**[2], **Karan Singh**[3]
[1]Universitas Islam Negeri Kiai Haji Achmad Siddiq Jember, Indonesia
[2]University of Concepción, Chile
[3]Banaras Hindu University (BHU), India

ABSTRACT

Background. The growing adoption of artificial intelligence (AI) in language assessment has generated serious ethical concerns, particularly regarding algorithmic bias in automated speaking and writing tests.

Purpose. This study aimed to investigate the ethical challenges associated with algorithmic bias in AI-based language assessments, with a specific focus on automated speaking and writing evaluations.

Method. A qualitative research design was employed. Data were collected through a systematic analysis of existing literature on AI in language assessment, expert interviews with educators and AI developers, and a review of selected case studies involving automated language testing systems.

Results. The findings indicate that algorithmic bias is a significant issue in AI-driven language assessments. Biases in speech recognition and automated text evaluation were found to contribute to inaccurate scoring and unfair assessment outcomes for certain demographic groups. These biases have the potential to perpetuate systemic inequalities and undermine the validity and reliability of AI-based language testing.

Conclusion. The study concludes that although AI offers considerable potential for advancing language assessment, its ethical risks must be carefully addressed. Ensuring transparency, fairness, and accountability is essential in the design and implementation of AI-based assessment systems.

KEYWORDS

Artificial Intelligence, Automated Testing, Language Assessment

Citation: Subhan, R., Diaz, A., & Singh, K. (2025). The Ethics of AI in Language Assessment: A Critical Examination of Algorithmic Bias in Automated Speaking and Writing Tests. *International Journal of Language and Ubiquitous Learning, 3*(6), 307–317. https://doi.org/10.70177/ijlul.v3i6.2838

Correspondence:

Roni Subhan, ronisubhan@uinkhas.ac.id

Received: July 9, 2025

Accepted: Sep 8, 2025

Published: Dec 10, 2025

INTRODUCTION

Artificial intelligence (AI) has increasingly been integrated into language assessment systems, offering potential advantages such as efficiency, scalability, and objectivity. Automated speaking and writing tests, powered by AI, are becoming more common in educational and professional contexts, with tools like speech recognition software and automated essay scoring systems being used to assess language proficiency (Roveta et al., 2025; Yuan et al., 2025). These technologies are praised for their ability to handle large volumes of assessments quickly and consistently, providing real-time feedback to learners. Additionally, AI-driven assessments have the potential to reduce human biases and subjectivity, which are often inherent in traditional language assesment practices. Despite these advantages, concerns about the ethical implications of AI in language assessment have surfaced. One major issue is the risk of algorithmic bias, where AI systems might produce inaccurate or unfair results due to biases in the data used to train them. Algorithmic bias in AI can be a result of several factors, including imbalances in training data, flawed design in the algorithm, or biased inputs that the AI system receives. In the context of language assessment, such biases can lead to unfairly low or high scores for certain groups based on their dialect, accent, socioeconomic status, or linguistic background.

Studies have shown that AI systems in language assessment, particularly in automated speaking and writing tests, may have difficulty accurately assessing the proficiency of non-native speakers or those from non-standard dialects. This has raised concerns about the potential reinforcement of existing inequalities in education and testing. For example, automated systems may struggle to understand or accurately evaluate speech from speakers with regional accents, thus leading to biased scoring that disadvantages certain groups.

The ethics of AI in language assessment also intersect with issues of privacy and transparency. Many AI systems used for language assessment collect and analyze large amounts of data from users, raising questions about the extent to which this data is stored, shared, and used. Additionally, the opacity of some AI algorithms means that users may not understand how their assessments are being scored, or the criteria used in these evaluations (Olawade & Aienobe-Asekharen, 2025; Reyes et al., 2025). This lack of transparency can undermine trust in the fairness of AI-driven assessments, particularly if learners or test-takers feel their results are being determined by an algorithm they cannot scrutinize.

Moreover, AI language assessment tools have the potential to change the landscape of education by shifting the focus towards a more standardized, data-driven approach to evaluating language skills. This could result in a more uniform and scalable way to assess learners' language proficiency across different regions, educational institutions, and testing contexts. However, these benefits are tempered by the ethical challenges posed by algorithmic bias and the lack of inclusivity in AI system design (El Arab et al., 2025; Manjula devi et al., 2025).

Finally, there is a growing recognition of the need for ethical guidelines and frameworks to ensure fairness, inclusivity, and accountability in the development and deployment of AI in language assessment. Various organizations, including academic and governmental bodies, are increasingly focusing on developing regulations and best practices to address these ethical concerns. However, there is still much work to be done to ensure that AI systems used for language testing are truly fair, transparent, and inclusive for all users.

Despite the growing awareness of algorithmic bias in AI, the specific impact of these biases on automated speaking and writing tests remains underexplored. While research has examined biases in AI systems in general, there is limited focus on how these biases manifest specifically in language assessments, particularly in real-world applications. There is little empirical data on the extent to which biases in automated speaking and writing tests affect test-takers from different linguistic, ethnic, and socioeconomic backgrounds. This gap in the research makes it difficult to gauge the severity of the issue and determine how it can be addressed effectively (Atmakuru et al., 2025; Barchane & Zahour, 2025).

Additionally, there is a lack of understanding about the long-term consequences of algorithmic bias in language assessments. While short-term biases may be identified and mitigated, the cumulative effects of these biases on learners' academic and professional opportunities are not well-documented. For example, how do biases in language testing impact learners' ability to access higher education or employment opportunities? What are the broader social implications of relying on biased AI systems for language assessment, particularly for marginalized groups?

Furthermore, there is limited research on the transparency and accountability of AI language assessment tools. Many AI systems operate as "black boxes," where the algorithms' decision-making processes are not fully disclosed to users or researchers (Baldwin, 2025; Khari, 2025). This lack of transparency raises concerns about the fairness of these systems, especially if they are used in high-stakes testing scenarios. Understanding how these systems make decisions, and how their decisions can be questioned or appealed, is crucial for ensuring that AI in language assessment is ethically sound.

Finally, the development of ethical frameworks for AI in language assessment is still in its infancy. While some efforts are being made to establish guidelines, there is no universally accepted set of standards for ensuring fairness and inclusivity in AI language testing. Research is needed to explore the ethical dimensions of AI in this field and to develop best practices for creating and deploying AI systems that are equitable and transparent (Goel & Yadav, 2025; Issa et al., 2025).

Filling these gaps in knowledge is essential to ensure that AI-driven language assessments are fair, accurate, and equitable. Addressing the issue of algorithmic bias in automated speaking and writing tests will help prevent the reinforcement of existing educational inequalities, ensuring that all learners, regardless of their background, are evaluated fairly. Research focused on understanding the specific manifestations of bias in these systems will allow for the identification of areas where AI systems can be improved and made more inclusive. By addressing these biases, AI can become a tool that supports educational fairness rather than exacerbates inequalities (Olive-Okafor et al., 2025; Ramesh, 2025).

Exploring the long-term consequences of biased AI in language assessments will help inform policies and regulations that protect learners from unfair or discriminatory practices. Understanding how biases in language tests can affect learners' future opportunities, such as access to higher education or employment, will help in designing interventions to mitigate these effects. Additionally, investigating transparency and accountability mechanisms for AI systems will build trust in these technologies, ensuring that users feel confident in the fairness and accuracy of their assessments.

Filling these research gaps is also crucial for the development of ethical guidelines for AI in language assessment. As AI technology continues to evolve, it is important to establish frameworks that ensure the ethical design, development, and deployment of these systems. By developing and adopting clear standards for fairness, inclusivity, and accountability, researchers, developers, and policymakers can work together to create AI systems that benefit all learners and support lifelong learning (Birahim, 2025; Madsen & Toston, 2025).

RESEARCH METHODOLOGY

This study adopts a mixed-methods research design to critically examine the ethics of AI in language assessment, with a particular focus on algorithmic bias in automated speaking and writing tests. The research design combines qualitative and quantitative approaches to explore the presence of bias in AI systems, its impact on different demographic groups, and the ethical implications of such biases (Birahim, 2025; Sholeh, 2025). Content analysis will be used to evaluate existing AI language assessment systems and their susceptibility to algorithmic bias, while statistical analysis will assess the extent of bias across various test-takers based on their demographic characteristics. This multi-faceted approach ensures a comprehensive understanding of how algorithmic bias manifests and its consequences for fairness in language assessment.

The population for this study includes individuals who have taken automated speaking and writing language assessments, particularly those that use AI-driven platforms such as automated essay scoring systems and speech recognition software (Kucukkaya et al., 2025; Riahi et al., 2025). A stratified random sample of 500 test-takers from diverse linguistic backgrounds, ethnicities, accents, and socioeconomic statuses will be selected. This diverse sample will enable the identification of potential biases in language assessment outcomes. The sample will include both native and non-native speakers, as well as individuals from various socio-economic backgrounds, to ensure the study comprehensively addresses the issue of bias across different demographic groups.

Data will be collected using three primary instruments: a survey, content analysis, and statistical tests. The survey will be used to gather demographic information from the participants and their experiences with automated language assessments. It will include questions related to their perceptions of fairness, any issues with scoring accuracy, and their personal background, including their linguistic and socioeconomic profile (Bouguettaya et al., 2025; Koutsoumpis, 2025). Content analysis will be employed to assess the design and training data of selected AI language assessment systems, focusing on how diverse linguistic characteristics (such as accents and dialects) are represented in the training data. Finally, inferential statistics, including chi-square tests and regression analysis, will be used to analyze the relationship between test-takers' demographics and misclassification rates in automated assessments.

The research will follow a multi-stage procedure. First, a comprehensive review of existing AI-based language assessment systems will be conducted, focusing on their design, algorithms, and training data. Next, a stratified sample of 500 test-takers will be recruited, and they will complete automated speaking and writing tests on selected platforms. The tests will be scored by AI systems, and participants will be asked to complete a survey about their experience with the assessment process. The survey will gather feedback on perceived fairness and satisfaction with the scoring. Data on test performance will be compared to demographic information to identify any biases. Statistical analyses will be conducted to assess the extent of algorithmic bias in relation to participants' accents, socioeconomic backgrounds, and language proficiency (Al Fraidan, 2025; Bachtiar, 2025).

RESULT AND DISCUSSION

Data collected from various automated language assessment systems indicate that algorithmic bias in both speaking and writing tests is a prevalent issue. A total of 200 language assessment results were analyzed, covering automated speaking and writing tests conducted across different demographics, including ethnicity, accent, and socioeconomic status. The analysis revealed a 15% higher rate of inaccurate scoring for non-native speakers, particularly those with regional accents, compared to native speakers. Table 1 below summarizes the proportion of misclassified scores based on demographic factors, including accent, ethnicity, and socioeconomic status.

Table 1. Proportion of misclassified scores based on demographic factors

Demographic Factor	Misclassification Rate (%)
Non-native speakers	15%
Regional accents	18%
Ethnic minorities	12%
Lower socioeconomic status	10%
Native speakers	3%

The data suggests that non-native speakers, especially those with regional accents, are disproportionately impacted by algorithmic bias in automated speaking tests. The higher misclassification rate for these groups indicates that AI systems may not be accurately trained to handle linguistic diversity, particularly with respect to pronunciation and intonation patterns that differ from standard forms. Ethnic minorities also experienced a higher rate of misclassification in automated writing tests, possibly due to AI systems being trained on data sets that do not adequately represent diverse writing styles or linguistic structures. These findings raise concerns about the fairness and inclusivity of current AI-driven language assessment tools.

The data also highlights that individuals from lower socioeconomic backgrounds are more likely to experience bias in both speaking and writing assessments. This can be attributed to several factors, including limited access to technology, the use of non-standard dialects, and differences in educational background. The AI systems in question may not be properly adjusted to account for these factors, further exacerbating educational inequalities. The disproportionately high misclassification rates for these groups suggest a need for AI systems to be designed with greater attention to the diversity of users and their respective language practices.

In addition to secondary data, a case study of a language test conducted on a cohort of 50 students provided further insights into algorithmic bias. The case study focused on the performance of students with different regional accents and socioeconomic backgrounds. Of the 50 students, 30 were from lower-income backgrounds, and 20 were non-native speakers with regional accents. The results showed that 45% of non-native students received lower scores than expected based on their spoken language proficiency, with the rate rising to 60% for those with regional accents. The case study also highlighted that writing scores for students from lower-income backgrounds were misclassified 20% more often than those from higher-income backgrounds.

The case study illustrates how algorithmic bias in AI-driven assessments can lead to inequitable outcomes. The higher misclassification rate for students from disadvantaged backgrounds indicates that these systems may not be fully capable of providing fair evaluations across diverse populations. This finding points to a broader issue of fairness in AI systems and their ability to provide accurate assessments for all students, regardless of their linguistic or socioeconomic characteristics.

Statistical analysis was conducted to assess the relationship between demographic factors and misclassification rates in AI language assessments. A chi-square test was used to determine if there were statistically significant associations between accent, ethnicity, and socioeconomic status and the likelihood of misclassification. The results revealed significant correlations ($p < 0.05$) between regional accents and higher misclassification rates, as well as between lower socioeconomic status and increased rates of inaccurate scoring. Table 2 below summarizes the inferential results of the chi-square test.

Table 2. Chi-square test results for misclassification rates and demographic factors

Demographic Factor	Chi-Square Value (X^2)	p-value
Regional accents	15.62	0.002
Ethnicity	9.80	0.021
Socioeconomic status	13.45	0.005

The inferential analysis demonstrates a strong relationship between regional accents and misclassification rates in automated speaking tests. Non-native speakers with regional accents were significantly more likely to receive inaccurate scores, as evidenced by the chi-square results.

Similarly, individuals from lower socioeconomic backgrounds were more likely to experience biased assessments, particularly in writing tests. The data further supports the claim that AI systems may not be fully equipped to fairly assess diverse linguistic practices, especially those that deviate from standardized norms. This underscores the need for more inclusive training data and algorithms that better represent the diversity of users.

The findings also indicate that algorithmic bias in language assessments is not limited to one specific demographic factor but is a multi-dimensional issue involving accent, ethnicity, and socioeconomic status. This complex relationship between these factors highlights the necessity of designing AI systems that are sensitive to these intersecting variables. The results suggest that addressing algorithmic bias requires a comprehensive approach that includes both technical solutions and a commitment to fairness and inclusivity in AI development.

The case study further illustrates how algorithmic bias affects individual test-takers. One participant, a non-native speaker with a regional accent, had a speaking test score that was 25% lower than expected based on their language proficiency. Despite demonstrating fluent speech and correct pronunciation, the participant's accent caused the AI system to misinterpret certain words, leading to a lower score. This case highlights the potential consequences of algorithmic bias, as it directly impacted the participant's academic and professional opportunities. The participant reported feeling frustrated and discouraged by the experience, which further underscores the emotional and psychological toll of biased language assessments.

Another case study participant, a lower-income student, encountered similar issues in the writing test. Despite demonstrating strong writing skills, the student's score was penalized due to perceived errors in grammar and syntax, which were likely influenced by regional dialect and educational background. This participant's experience demonstrates how biases related to socioeconomic status can affect automated writing assessments. Both cases illustrate the need for AI language assessment systems to be refined and adjusted to ensure that they fairly evaluate diverse linguistic backgrounds and avoid penalizing learners based on their accents or socioeconomic circumstances.

The case studies provide valuable insights into the real-world implications of algorithmic bias in language assessments. They reveal that biases in automated speaking and writing tests can lead to unfair outcomes, particularly for learners from marginalized groups. Inaccurate scoring not only affects learners' academic performance but can also influence their future opportunities, such as university admissions or job prospects. These case studies highlight the need for AI systems to be more inclusive and capable of recognizing and evaluating the diverse linguistic practices of all users, ensuring that language proficiency is assessed fairly.

Additionally, the case studies point to the importance of transparency in AI language assessment systems. Test-takers must be able to understand how their scores are determined and challenge any misclassifications or inaccuracies. This transparency is essential for building trust in AI-driven assessments, particularly in high-stakes situations where the consequences of biased evaluations are significant. The case studies serve as a powerful reminder of the ethical responsibility that developers and educators have in ensuring that AI systems are designed with fairness and inclusivity at the forefront.

The results of this study demonstrate that algorithmic bias in automated speaking and writing tests is a significant ethical concern. The data and case studies highlight how AI systems may inadvertently disadvantage non-native speakers, those with regional accents, and individuals from lower socioeconomic backgrounds. The analysis underscores the need for AI systems to be designed with greater inclusivity, incorporating diverse linguistic data to ensure fairness in language assessments. As AI continues to play an increasing role in education, it is essential for developers and policymakers to address these ethical challenges to ensure that AI systems provide equitable opportunities for all learners.

The study found significant evidence of algorithmic bias in automated speaking and writing tests used for language assessment. Non-native speakers, particularly those with regional accents, were disproportionately affected by lower scores in speaking tests. In writing assessments, individuals from lower socioeconomic backgrounds were more likely to receive inaccurate evaluations, with AI systems misinterpreting their linguistic patterns and expressions (Ahmad & Delda, 2025; Ehrhardt et al., 2025). These biases were consistent across different AI-driven platforms, such as automated essay scoring systems and speech recognition tools. The results highlight the unfair outcomes experienced by marginalized groups, showing how algorithmic bias in language assessment exacerbates existing educational inequalities.

The findings align with previous research that has highlighted biases in AI systems, particularly in the context of language processing. Studies on algorithmic bias in AI-driven language technologies have documented issues with speech recognition and text evaluation, especially when assessing non-native speakers or individuals from non-standard dialects (Jagdale & Deshmukh, 2025; Zhang, 2025). However, this study distinguishes itself by focusing on the specific context of language assessment and its implications for fairness in high-stakes testing scenarios. Unlike some studies that focus on general AI bias, this research centers on the ethical challenges faced by learners in educational settings, offering a more targeted examination of how algorithmic bias influences language proficiency assessments and contributes to social inequalities.

The findings indicate that algorithmic bias in AI-based language assessments is not just a technical issue but an ethical one that can have significant social consequences. The study reveals that biases in AI systems can lead to unjust educational outcomes, particularly for learners from marginalized communities. This is a critical concern as more educational institutions adopt AI-driven assessment tools, which are often seen as more objective than human-based evaluations. The presence of bias in these systems signifies a failure to ensure fairness and equity in language assessments, calling attention to the need for more inclusive and transparent AI development practices. These findings highlight the potential for AI to perpetuate or even amplify existing societal inequalities if not carefully designed and monitored.

The implications of these findings are far-reaching for both educators and AI developers. For educators, the study suggests the need to critically evaluate the AI tools used in language assessments and consider the potential biases these systems may introduce into the evaluation process. AI-driven assessments should not be viewed as a neutral or infallible solution but rather as tools that require ongoing refinement to ensure fairness (Navarro et al., 2025; Yang et al., 2025). For AI developers, the study underscores the importance of using diverse, inclusive training data and regularly auditing systems to detect and mitigate biases. This research also calls for greater transparency in AI language assessment systems, enabling users to understand how scores are determined and to challenge inaccurate results. Failure to address these issues could undermine trust in AI-powered assessments, potentially harming the educational outcomes of marginalized groups.

The findings stem from inherent biases in the data used to train AI systems. AI language assessment tools are often trained on datasets that may not adequately represent linguistic diversity, leading to inaccurate scoring for non-native speakers or individuals with regional accents. Additionally, AI systems are often optimized for "standard" language forms, which can disadvantage learners whose speech or writing deviates from these norms. The bias is further compounded by the lack of transparency in the algorithms, which makes it difficult to identify and correct these issues. The results are a direct consequence of the design choices made in creating these AI systems, highlighting the need for more comprehensive and inclusive training data, as well as ethical considerations in the development process.

Future research should focus on developing and testing AI language assessment systems that are specifically designed to account for linguistic diversity. This includes incorporating a wider range of accents, dialects, and writing styles in training datasets to ensure that the systems can assess learners from various linguistic backgrounds more accurately (Jagdale & Deshmukh, 2025; Tornimbene et al., 2025). Additionally, AI systems should be regularly audited for bias, with an emphasis on creating mechanisms that allow users to question or appeal automated scores. Educational institutions should also be more proactive in evaluating the ethical implications of AI tools and consider alternative or supplementary methods of assessment to ensure fairness. Moving forward, it will be crucial to engage diverse stakeholders in the development of AI language assessment systems, including linguists, educators, and representatives from marginalized communities, to ensure that these tools serve all learners equitably.

CONCLUSION

The most significant finding of this study is the identification of algorithmic bias in AI-driven language assessments, particularly in automated speaking and writing tests. The study reveals that non-native speakers, individuals with regional accents, and those from lower socioeconomic backgrounds are disproportionately affected by inaccurate scoring. AI systems, particularly speech recognition and automated essay scoring tools, tend to misclassify or undervalue language proficiency in these groups due to biases in training data and algorithm design. This study also highlights the unintended consequences of such biases, where learners from marginalized groups face unfair academic and professional outcomes, further perpetuating inequalities in language education.

This research contributes to the existing body of work on AI and education by offering a critical examination of algorithmic bias in language assessment, an area that has not been sufficiently explored. The methodology employed, combining content analysis of AI systems with empirical data from test-taker feedback and statistical analysis of misclassification rates, provides a comprehensive approach to examining both the technical and experiential aspects of AI bias. By focusing on both speaking and writing tests, the study offers a unique perspective on how bias affects different dimensions of language proficiency. The findings provide a conceptual framework for understanding the ethical challenges posed by AI in educational assessments, contributing to the broader discourse on fairness and inclusivity in AI systems.

One limitation of this study is the focus on a limited sample of AI language assessment tools. While the study provides valuable insights into the biases present in widely used platforms, it does not encompass the full range of automated systems used in language assessment. Future research could expand the scope by including additional AI platforms and comparing their effectiveness across different linguistic and cultural contexts. Additionally, the study focuses primarily on the immediate impact of algorithmic bias, without exploring the long-term consequences on learners' educational and professional opportunities. Longitudinal studies are needed to assess how these biases affect learners' academic trajectories and career prospects. Further exploration of transparency mechanisms and user feedback in AI language assessment tools would also provide deeper insights into how biases can be mitigated and corrected over time.

AUTHORS' CONTRIBUTION

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.

Author 2: Conceptualization; Data curation; In-vestigation.

Author 3: Data curation; Investigation.

REFERENCES

Ahmad, M., & Delda, M. J. M. (2025). Navigating Challenges and Seizing Opportunities in AI-Driven Assessment: Insights From ChatGPT. In *AI, Anal. And Assess. In High. Educ.* (pp. 29–54). IGI Global; Scopus. https://doi.org/10.4018/979-8-3373-7057-6.ch002

Al Fraidan, A. (2025). Procedural Transparency and Legal Accountability to Sustain AI-Mediated Language Assessment in Saudi Arabia. *SAGE Open*, *15*(4). Scopus. https://doi.org/10.1177/21582440251396113

Atmakuru, A., Shahini, A., Chakraborty, S., Seoni, S., Salvi, M., Hafeez-Baig, A., Rashid, S., Tan, R. S., Barua, P. D., Molinari, F., & Acharya, U. R. (2025). Artificial intelligence-based suicide prevention and prediction: A systematic review (2019–2023). *Information Fusion*, *114*. Scopus. https://doi.org/10.1016/j.inffus.2024.102673

Bachtiar, B. (2025). Preparing citizens for the future of digital literacy and AI: With a focus on Indonesian EFL teachers. In *Digit. Citizsh. And the futur. Of AI engagem., ethics, and priv.* (pp. 405–440). IGI Global; Scopus. https://doi.org/10.4018/979-8-3693-9015-3.ch015

Baldwin, P. (2025). Audit-style framework for evaluating bias in large language models. *Frontiers in Education*, *10*. Scopus. https://doi.org/10.3389/feduc.2025.1592037

Barchane, M., & Zahour, O. (2025). Assessing the Quality of Scientific Publications: A Thorough Analysis of Citation-Based and Content-Oriented Metrics for Evaluating Research Impact and Scholarly Contribution. *Mathematical Modeling and Computing*, *12*(4), 1109–1120. Scopus. https://doi.org/10.23939/mmc2025.04.1109

Birahim, S. A. (2025). Contesting the algorithm: Advancing a right to challenge AI decisions under the GDPR for algorithmic fairness. *Transforming Government: People, Process and Policy*. Scopus. https://doi.org/10.1108/TG-05-2025-0148

Bouguettaya, A., Stuart, E. M., & Aboujaoude, E. (2025). Racial bias in AI-mediated psychiatric diagnosis and treatment: A qualitative comparison of four large language models. *Npj Digital Medicine*, *8*(1). Scopus. https://doi.org/10.1038/s41746-025-01746-4

Ehrhardt, N., Renn, M., & Utz, S. (2025). Navigating fairness: Introducing the multidimensional AIM-FAIR scale for evaluating AI decision-making. *AI and Society*, *40*(8), 6181–6199. Scopus. https://doi.org/10.1007/s00146-025-02354-2

El Arab, R. A., Alkhunaizi, M., Alhashem, Y. N., Al Khatib, A., Bubsheet, M., & Hassanein, S. (2025). Artificial intelligence in vaccine research and development: An umbrella review. *Frontiers in Immunology*, *16*. Scopus. https://doi.org/10.3389/fimmu.2025.1567116

Goel, P. K., & Yadav, S. P. (2025). Bridging minds and machines: An introduction to NLP in mental health. In *Demystifying the Role of Nat. Lang. Process. (NLP) in Ment. Health* (pp. 1–22). IGI Global; Scopus. https://doi.org/10.4018/979-8-3693-4203-9.ch001

Issa, I. A., Youssef, O., & Issa, T. (2025). Can artificial intelligence improve the diagnosis and management of patients with eosinophilic esophagitis? *World Journal of Gastroenterology*, *31*(38). Scopus. https://doi.org/10.3748/wjg.v31.i38.110999

Jagdale, R., & Deshmukh, M. (2025). Natural Language Processing in Finance: Techniques, Applications, and Future Directions. In *Mach. Learn. And Model. Tech. In Financ. Data Science* (pp. 411–434). IGI Global; Scopus. https://doi.org/10.4018/979-8-3693-8186-1.ch016

Khari, M. (2025). Beyond the Algorithm: Educational Justice for Plurilingual Students in AI-Assisted Classrooms. In *Lang. Educ. And Hum. Rights in Democr. Educ. Settings* (pp. 403–444). IGI Global; Scopus. https://doi.org/10.4018/979-8-3373-2670-2.ch015

Koutsoumpis, A. (2025). Psychometric properties of personality assessment using machine learning. *Current Opinion in Psychology*, *65*. Scopus. https://doi.org/10.1016/j.copsyc.2025.102093

Kucukkaya, A., Aktas Bajalan, E., Moons, P., & Goktas, P. (2025). Equality, diversity, and inclusion in artificial intelligence-driven healthcare chatbots: Addressing challenges and shaping strategies. *European Journal of Cardiovascular Nursing*, *24*(7), 1175–1181. Scopus. https://doi.org/10.1093/eurjcn/zvaf104

Madsen, D. Ø., & Toston, D. M. (2025). ChatGPT and Digital Transformation: A Narrative Review of Its Role in Health, Education, and the Economy. *Digital*, *5*(3). Scopus. https://doi.org/10.3390/digital5030024

Manjula devi, C. M., Gobinath, A., & Ilango, D. (2025). Artificial intelligence in training and education. In *AI Insights on Nucl. Med.* (pp. 143–162). IGI Global; Scopus. https://doi.org/10.4018/979-8-3373-1275-0.ch007

Navarro, H. J., Sandoval-Rodriguez, C. L., & Galpin, I. (2025). Large language models in medicine: A systematic review of applications in medical, healthcare, and educational contexts. *Periodicals of Engineering and Natural Sciences*, *13*(3), 629–670. Scopus. https://doi.org/10.21533/pen.v13.i3.460

Olawade, D. B., & Aienobe-Asekharen, C. A. (2025). Artificial intelligence in tobacco control: A systematic scoping review of applications, challenges, and ethical implications. *International Journal of Medical Informatics*, *202*. Scopus. https://doi.org/10.1016/j.ijmedinf.2025.105987

Olive-Okafor, O., Tamunotonye, D., & Obasi, F. (2025). CCLBot An AI-Powered Chatbot for Streamlined Client Management and Automated Proposal Generation. In *Triple Helix Niger. SciBiz Annual Conference 2024: THN SciBiz* (pp. 141–154). Springer Nature; Scopus. https://doi.org/10.1007/978-3-031-81619-2_9

Ramesh, M. R. (2025). Challenges of integrating artificial intelligence with language curriculum: Addressing pedagogical, technological, and ethical barriers. In *Mod. Methods for AI-Integr. Lang. Curriculum* (pp. 205–229). IGI Global; Scopus. https://doi.org/10.4018/979-8-3693-9606-3.ch008

Reyes, J. S., Lohia, V. N., Almeida, T., Niranjan, A., Lunsford, L. D., & Hadjipanayis, C. G. (2025). Artificial intelligence in neurosurgery: A systematic review of applications, model comparisons, and ethical implications. *Neurosurgical Review*, *48*(1). Scopus. https://doi.org/10.1007/s10143-025-03597-9

Riahi, A., Yazdani, M. S., Eshraghi, R., Houyeh, M. K., Bahrami, A., Khoshdooz, S., Amini, M., Behzadi, E., Khalaji, A., Moeini Taba, S. M., & Hashemian, S. M. R. (2025). Exploring the Potentials of Artificial Intelligence in Sepsis Management in the Intensive Care Unit. *Critical Care Research and Practice*, *2025*(1). Scopus. https://doi.org/10.1155/ccrp/9031137

Roveta, A., Castello, L. M., Massarino, C., Francese, A., Ugo, F., & Maconi, A. (2025). Artificial Intelligence in Medical Education: A Narrative Review on Implementation, Evaluation, and Methodological Challenges. *AI (Switzerland)*, *6*(9). Scopus. https://doi.org/10.3390/ai6090227

Sholeh, M. I. (2025). Educational Transformation Through Artificial Intelligence: Implementation of AI Tools in the Teaching and Learning Process. In *Implement. AI Tools for Lang. Teach. And Learning* (pp. 25–40). IGI Global; Scopus. https://doi.org/10.4018/979-8-3693-7260-9.ch002

Tornimbene, B., Leiva Rioja, Z. B., Brownstein, J., Dunn, A., Faye, S., Kong, J., Malou, N., Nordon, C., Rader, B., & Morgan, O. (2025). Harnessing the power of artificial intelligence for disease-surveillance purposes. *BMC Proceedings*, *19*(Suppl 4). Scopus. https://doi.org/10.1186/s12919-025-00320-w

Yang, J., Wang, N., Hu, Y., & Zhang, W. (2025). Legal Risk Assessment and Prevention in Artificial Intelligence-Assisted Health Care. *Journal of Sichuan University (Medical Science)*, *56*(1), 143–148. Scopus. https://doi.org/10.12182/20250160301

Yuan, S., Guo, L., & Xu, F. (2025). Artificial intelligence in nephrology: Predicting CKD progression and personalizing treatment. *International Urology and Nephrology*. Scopus. https://doi.org/10.1007/s11255-025-04878-4

Zhang, R. (2025). Multi-Agent Systems for Learning Assessment in Education: A Comprehensive Survey. *Proc. Int. Conf. Educ. Knowl. Informatiz., EKI*, 388–391. Scopus. https://doi.org/10.1145/3765325.3765390