



AI-DRIVEN SUBJECTIVE ASSESSMENT IN ONLINE EDUCATION: CHALLENGES, SOLUTIONS, AND FUTURE DIRECTIONS

Ahmad Tajudin Baharin^{*1}, Nurul Azlin Iyana Mohamad Redzuan², Rozlini Mohamed³,
Deden Witarasyah⁴ and Bazen Jacques Cornelis⁵

^{1, 3, 4}Faculty of Computer Science and Information Technology, Universiti Tun Hussien Onn Malaysia.

²Faculty of Computing & Multimedia, Universiti Poly-Tech Malaysia.

⁵School of Business, Building & Technology, Saxion University of Applied Sciences, Enschede, Netherlands.

¹<https://orcid.org/0009-0009-8809-2192>, ²<https://orcid.org/0009-0005-8345-2540>, ³<http://orcid.org/0000-0002-6592-5632>

⁴<http://orcid.org/0000-0001-8102-0326>, ⁵<http://orcid.org/0000-0001-7779-7271>

Email: tajudin@uthm.edu.my, azlin@uptm.edu.my, rozlini@uthm.edu.my, deden@uthm.edu.my, j.c.bazen@saxion.nl.

ARTICLE INFO

Article History

Received: January 21, 2026

Reviewed: July 24, 2026

Accepted: August 24, 2026

Published: September 30, 2026

Keywords:

Artificial Intelligence,
automated assessment,
large language models,
online learning,
explainable AI,
educational technology.

ABSTRACT

The increasing reliance on online education has increased the need for efficient and fair assessment methods to assess students' subjective responses such as essays and reflections. Manual scoring of these tasks remains time-consuming, inconsistent, and difficult to scale. Artificial Intelligence (AI), particularly through Natural Language Processing (NLP) and Large-Scale Language Models (LSM) such as GPT and BERT, offers a potential alternative. This systematic review synthesizes findings from 30 peer-reviewed research articles (2020–2025) examining AI approaches to automated subjective assessment. Through thematic synthesis, three critical issues are identified: (1) transparency and explainability, (2) contextual understanding, and (3) lack of uniform implementation guidelines. The review identifies key challenges, synthesizes emerging solutions such as Explainable AI (XAI) and human-AI hybrid models, and outlines future research and policy directions for AI-driven subjective assessment. A conceptual framework is also proposed to map the relationship between AI technologies, pedagogical principles, and institutional trust. This study contributes by providing practical guidance to educators, researchers, and policymakers in designing AI-based assessment systems that are transparent, fair, and responsive to context.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Rapid advances in Artificial Intelligence (AI) have led to a paradigm shift from traditional algorithmic systems to data-driven, generative models such as GPT-based architectures, enabling more sophisticated interaction, reasoning, and content generation capabilities [1]. In the context of online learning, one of the most pressing challenges is the need to efficiently and fairly assess students' subjective responses, especially for tasks such as essays, reflections, and open-ended responses. This type of assessment is crucial for measuring higher order thinking skills such as critical reasoning, creativity, and conceptual understanding that cannot be effectively measured through objective questions alone. Traditionally [2], subjective assessments have relied entirely on human judgment, which has brought with it various limitations in terms of scalability, consistency, and time efficiency especially in large-scale educational platforms such as Massive Open Online Courses (MOOCs [3]). This situation has driven a growing interest in the use of AI, particularly through Natural Language Processing (NLP) and Large-Scale Language Models (LSM) technologies such as GPT and BERT [4-6], [2], to automate subjective assessments. These technologies have shown great potential in understanding, analyzing, and generating text that resembles human writing, making them suitable candidates for assessment purposes in education. However, despite the increasingly sophisticated technical capabilities of AI models, there are still several important issues that hinder their effective implementation in real-world educational settings. These issues include the transparency of AI decision-making processes, limitations in understanding diverse student contexts, and the absence of comprehensive implementation frameworks and ethical guidelines. Furthermore, empirical evidence supporting the pedagogical suitability of these systems is still insufficient, raising questions about the fairness and reliability of their use.

This systematic review aims to explore existing approaches to AI-driven automated subjective assessment, assess their effectiveness, and identify gaps that need to be addressed to ensure successful implementation in the context of online learning. Through a synthesis of recent literature findings, this study aims to offer useful guidance to educators, researchers, and policymakers in developing AI-based assessment systems that are fairer, more transparent, and aligned with pedagogical goals.

II. CONCEPTUAL FOUNDATIONS

II.1 AI IN EDUCATION

Artificial Intelligence (AI) is increasingly playing a significant role in the modern educational landscape. The use of AI in education encompasses a wide range of applications, including learning analytics, intelligent tutoring, recommendation systems, and most prominently – automated assessment. With the rapid advancements in the field of Natural Language Processing (NLP) [7] and the development of Large-Scale Language Models (LSM) such as GPT [2] and BERT [8], the ability of AI systems to understand, analyze, and generate human-like text has opened up new opportunities in the automation of subjective assessment. A study by [9] emphasized that models such as ChatGPT have shown great potential in NLP, but also bring issues related to accuracy and ethics of their use in educational contexts. In addition, [4] also described the current trend of LLM [10], [11] which shows improvement in terms of performance but has not yet fully addressed the issues of reliability and contextual understanding in education [12]. This technology is seen as very potential in building large-scale and adaptive assessment systems, especially in online learning environments. However, the use of AI in education also raises concerns in terms of fairness, transparency and accessibility. A study by [13] emphasizes that the ethical and effectiveness aspects of AI need to be carefully addressed through the implementation of policies and periodic audits so that the use of this technology truly supports the needs of holistic and inclusive education. By [14] reviewed foundation models in agriculture, while [15] introduced AI applications in orthopaedic research, and both emphasized the importance of data quality, model training, and responsible deployment for effective decision-support. These themes are also central to AI in education, where similar challenges must be addressed to ensure reliable, transparent, and impactful use of AI technologies in enhancing teaching and learning. By [16] explored emotion detection using deep learning techniques, which can be applied in educational contexts to monitor student engagement and provide adaptive feedback. Similarly, [17] conducted a comprehensive survey on personality trait detection with machine and deep learning, offering insights into how AI can personalize learning experiences by aligning teaching strategies with individual learner characteristics. Both studies highlight the potential of AI in education to foster more adaptive, learner-centered approaches.

II.1.1 Subjective Assessment in Online Learning

Subjective assessments, such as essays, reflections, and open-ended responses, are important elements in measuring students' higher-order skills such as critical thinking, deep analysis, and creative expression. In the context of online education, the main challenges are in terms of scale, consistency, and time efficiency of conducting manual assessments for each student. Therefore, the emergence of AI-based subjective assessment automation systems has been considered as a promising alternative. However, according to [3], the use of AI in subjective assessment still faces questions regarding accuracy, explainability, and reliability. They found that although models such as GPT-4 are able to extract information from text quickly, they are still prone to errors in understanding context. Recent systematic reviews within educational contexts indicate that while large language models demonstrate promising capabilities in automated grading and feedback generation, concerns remain regarding reliability, contextual interpretation and pedagogical alignment in real classroom environments [18]. Meanwhile [19] stated that automated scoring systems need to be reviewed from a pedagogical perspective, as not all models are suitable for curriculum approaches that require narrative interpretation. Recent studies examining the impact of large language models on higher education assessment suggest that generative AI challenges traditional grading paradigms and necessitates revised evaluation frameworks that account for authenticity, bias, and academic integrity [20]. By [21] investigated transformer-based models within the legal domain, demonstrating their capacity to interpret nuanced and complex textual inputs, which parallels the requirements of automated grading in subjective assessments such as essays and case analyses. Likewise, [22] examined the use of large language models for materials language processing, emphasizing their role in analyzing specialized, discipline-specific text. These works collectively underscore the potential of AI to enhance subjective assessment in online learning by enabling more accurate and context-aware evaluation of open-ended student responses. In a systematic review by [23], it was found that the effectiveness of NLP in the health field can be used as a reference for education, but adaptation to the learning context and subjective tasks requires a more holistic and ethical approach.

III. METHODOLOGY

To ensure rigor and transparency, this study adopted a systematic review approach guided by PRISMA principles. The methodology included four main stages: identification, screening, eligibility, and inclusion.

III.1 SEARCH STRATEGY

A comprehensive literature search was conducted across five major scholarly databases: Scopus, Web of Science, IEEE Xplore, ACM Digital Library, and ScienceDirect. The search was limited to studies published between January 2020 and March 2025 to capture the rapid evolution of AI and large language models in educational assessment.

Search strings combined keywords and Boolean operators, including:

("artificial intelligence" OR "AI" OR "machine learning" OR "large language model" OR "LLM" OR "natural language processing" OR "NLP")

AND

(“subjective assessment” OR “essay scoring” OR “open-ended response” OR “automated grading” OR “student reflections”)
AND
(“education” OR “online learning” OR “e-learning” OR “MOOC” OR “digital learning”).

III.2 INCLUSION AND EXCLUSION CRITERIA

To maintain focus, the following criteria were applied:

Inclusion criteria:

- Peer-reviewed journal articles or conference papers.
- Published in English between 2020–2025.
- Empirical studies, systematic reviews or conceptual papers addressing AI-based or NLP/LLM-driven subjective assessment in education.

Exclusion criteria:

- Grey literature (blogs, white papers, dissertations).
- Articles focusing solely on objective test automation (e.g., multiple-choice grading).
- Non-educational contexts (e.g., AI in healthcare or law, unless directly transferable to education).

III.3 SCREENING AND SELECTION PROCESS

The initial search yielded 432 records. After removing duplicates (n = 96), titles and abstracts were screened for relevance (n = 336). Of these, 210 articles were excluded for not meeting the inclusion criteria. A further 96 full-text articles were assessed, and 66 were excluded for reasons such as lack of educational context or focus on objective rather than subjective assessment. Finally, 30 articles met all criteria and were included in the review. Figure 1 presents the PRISMA flow diagram of the selection process.

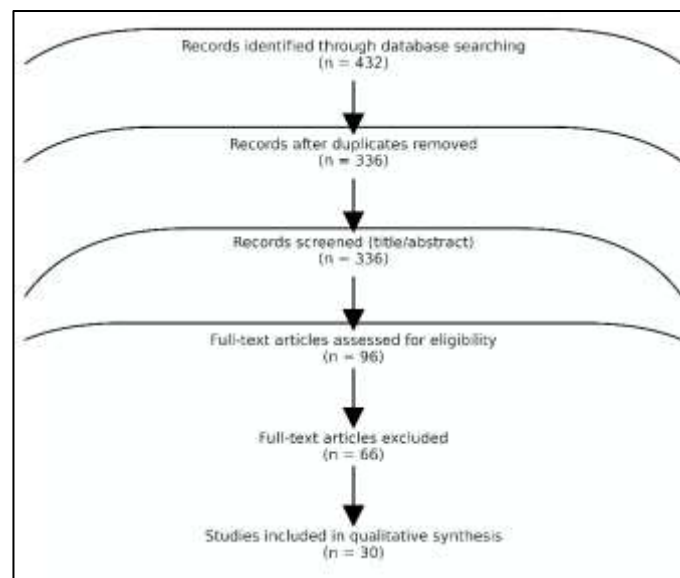


Figure 1: PRISMA flow diagram of study selection.

Source: Authors, (2026).

The figure shows the process used in this systematic review. From 432 identified records, 96 duplicates were removed, 336 records were screened, and 96 full-text articles were assessed. A total of 30 studies met all eligibility criteria and were included in the final synthesis.

III.4 DATA EXTRACTION AND ANALYSIS

One Data were systematically extracted using a coding matrix that captured:

- Bibliographic information (author, year, venue).
- Type of AI approach used (e.g., GPT, BERT, RoBERTa, hybrid).
- Educational context (higher education, K–12, MOOCs).
- Assessment focus (essays, reflections, short answers).
- Reported strengths, limitations, and proposed solutions.

Thematic synthesis was then employed to identify recurring issues and opportunities, which were organized into the three central themes:

- Transparency and Explainability
- Contextual Understanding
- Implementation Guidelines

III.5 THEME 1: TRANSPARENCY AND EXPLAINABILITY

One of the main concerns in using AI systems for assessment is the level of transparency and explainability of the models used. Most LLM models such as GPT and BERT function as "black box" systems that are difficult to understand by end users, especially lecturers and students. This raises questions about how a score or response is given by the system. A study by [24] highlighted that although sophisticated NLP models provide accurate results, justification for the results generated is often not available, thus hindering user confidence. By [23] also emphasized that the aspect of explainability is very important in critical domains such as health and should be applied in education. Meanwhile [25] propose the use of Explainable AI (XAI) in automated scoring systems to improve user understanding of the assessment process.

III.6 THEME 2: CONTEXTUAL UNDERSTANDING

AI models often fail to fully understand the context of students, including cultural background, language proficiency level, writing style, and the implicit meaning of students' responses. In the context of subjective assessment, these elements are very important because they reflect the students' deep thinking and originality. A study by [26] found that most LLMs have limitations in interpreting complex linguistic nuances and pedagogical values. In turn [27] also showed that GPT-4 has a tendency to make generic assessments without taking into account the uniqueness of the students' context. To overcome this issue [19] propose a hybrid model that incorporates human input to ensure more accurate context understanding.

III.7 THEME 3: IMPLEMENTATION GUIDELINES

While the use of AI in assessment is becoming more widespread, the lack of clear standard guidelines is a major challenge in terms of safety, ethics and implementation effectiveness. Each institution tends to develop its own system without reference to a national or global framework. A study by [13] emphasizes that educational AI systems should be accompanied by performance monitoring procedures, ethical impact assessments and periodic reporting. Meanwhile [26] suggest the need for a comprehensive educational policy framework to guide the implementation of AI in higher education systems. A study by [28] also states that standardization is a key prerequisite in ensuring the fairness and consistency of AI systems in assessment.

Table 1: Summarization of thematic synthesis literature.

Theme	Key Studies	Main Findings	Limitations	Validated and Emerging Solutions
Transparency & Explainability	[24], [25]	LLMs act as "black boxes"; lack justification	Low user trust	Use XAI, user-friendly interfaces
Contextual Understanding	[26], [27], [19]	Struggle with cultural/linguistic context	Generic scoring, bias	Hybrid AI-human models
Implementation Guidelines	[13], [28]	Absence of standards across institutions	Ethical and policy gaps	National/global frameworks, audits

Source: Authors, (2026).

IV. DEVELOPMENT OF CONCEPTUAL FRAMEWORK

The conceptual framework for this study was developed based on a thematic synthesis of 30 articles related to AI approaches in online subjective assessment. Three main themes Transparency and Explain ability, Contextual Understanding, and Implementation Guidelines were identified as critical factors influencing the effectiveness of AI-based assessment systems in higher education. In general, this framework establishes the relationship between AI technology inputs (such as LLM, NLP, and XAI) and student learning outcomes through an automated assessment mechanism that is influenced by the following three main components:

IV.1 AI APPROACH IN SUBJECTIVE ASSESSMENT

This element refers to the technologies and models used to automatically perform assessments, including GPT, BERT, and various machine learning techniques. It is the foundation of the entire system. The review identified 30 relevant articles. AI models commonly used include GPT-3/4, BERT, RoBERTa, and hybrid models. Three major themes were found:

- Transparency and Explainability: Studies highlighted the 'black-box' issue and proposed Explainable AI (XAI) as a solution.
- Contextual Understanding: LLMs struggle with understanding student-specific context, affecting grading fairness.
- Lack of Standardization: Absence of unified guidelines led to inconsistent implementation across platforms and institutions.

Traditional methods such as narrative feedback and letters of reference are subjective, unreliable, and biased by gender or demographics [29]. While NLP tools may uncover these latents e.g., differences in tone or word choice between groups they have the potential to make existing biases worse when trained on flawed historical data. This raises issues of fairness and transparency, as residents and teachers may not understand and accept algorithmic inferences about performance or risk.

The article highlights that without well-defined ethics and bias-reduction strategies, NLP use in employment during residency evaluation may unintentionally continue to reinforce inequities rather than eliminating them. Supported by studies: [30], [31].

IV.2 CRITICAL INFLUENCING FACTORS (THREE THEMES)

The conceptual framework in this study was refined to capture the complex interplay between technological, pedagogical, and ethical dimensions of AI-driven subjective assessment. Based on the thematic synthesis of 30 studies, three interdependent factors emerged as critical: Transparency and Explainability, Contextual Understanding, and Implementation Guidelines. Unlike a linear cause-effect relationship, these factors operate within a cyclical and iterative system that continuously shapes outcomes in online learning.

IV.3 TRANSPARENCY AND EXPLAINABILITY

Transparency enables both students and educators to understand how automated scoring decisions are derived. Beyond improving user trust, explainability also acts as a bridge to pedagogical validity, ensuring that assessment criteria are clearly aligned with intended learning outcomes. For example, integrating Explainable AI into essay scoring allows teachers to trace which textual features influenced the grade, reinforcing alignment with assessment-for-learning principles. Supported by studies: [24], [25].

IV.4 CONTEXTUAL UNDERSTANDING

Automated systems often struggle with nuanced student contexts such as cultural background, linguistic diversity and individual writing styles. Informed by constructivist learning theory, contextual understanding is essential because assessments should not merely measure correctness but reflect critical reasoning and creativity. A hybrid AI-human model provides a safeguard here, where AI handles scalability while human evaluators interpret originality and cultural nuances. Supported by studies: [26], [27].

IV.5 IMPLEMENTATION GUIDELINES

Institutional and policy frameworks act as moderators that regulate the use of AI in assessment. Clear guidelines covering ethics, bias audits, and privacy protection ensure equitable deployment across diverse educational settings. For example, national assessment authorities could mandate regular algorithmic audits, similar to accreditation audits, to safeguard fairness and accountability. Supported by studies: [13], [28].

IV.6 OUTCOME & CONCEPTUAL RELATIONSHIP SUMMARY

When the three critical factors of transparency and explainability, contextual understanding, and implementation guidelines are incorporated into system design, AI systems in education are expected to deliver outcomes that extend beyond efficiency. First, they improve the scalability and timeliness of assessment, making it feasible to provide rapid feedback even in large-scale online learning environments. Second, they enhance fairness and reliability, ensuring that assessment is consistent, unbiased, and aligned with pedagogical goals. Third, by embedding transparency and ethical safeguards, these systems build greater acceptance and trust among both educators and learners, increasing the likelihood of sustained adoption in real classrooms. The proposed framework can be summarized as a cyclical relationship between AI technologies, critical influencing factors and educational outcomes. AI-driven assessment models (such as LLMs and NLP-based systems) provide the technical foundation, but their effectiveness is mediated by the three pillars of transparency, contextual understanding and implementation guidelines. When these pillars are operationalized, they lead to fair and effective automated assessment practices. Over time, this results in high-quality learning outcomes and institutional trust, which in turn generate feedback that informs further system refinement and policy development. Thus, the framework emphasizes not only the linear flow from technology to outcomes but also the iterative feedback loop that drives continuous improvement.

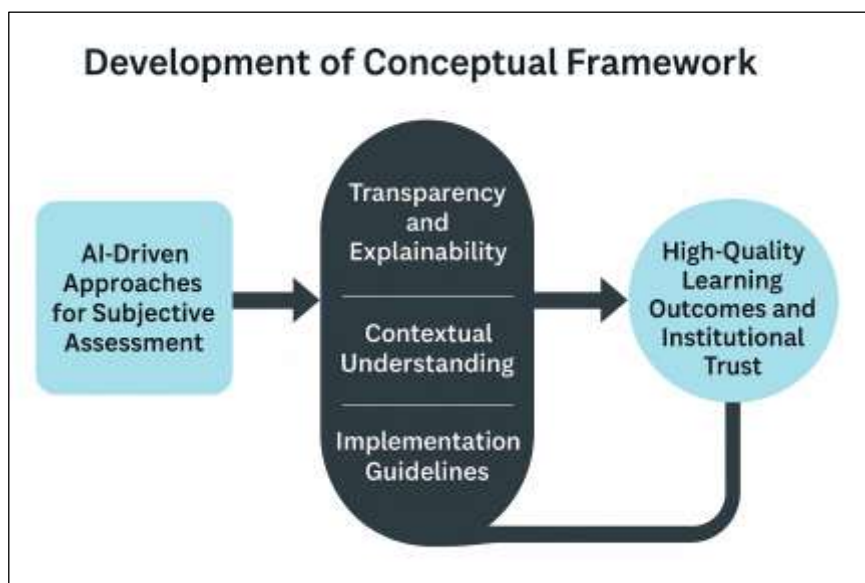


Figure 2: Development of Conceptual Framework.
Source: Authors, (2026).

V. IMPLICATION AND FUTURE DIRECTIONS

Building on the identified challenges and emerging solutions, this section outlines future research and policy directions for AI-driven subjective assessment in online education. The integration of AI in subjective assessment offers a unique opportunity to improve the efficiency, scalability, and personalization of feedback in online learning environments. However, its effectiveness depends greatly on the extent to which AI systems are aligned with pedagogical principles and values such as fairness, transparency, and inclusiveness. Educators need to be trained not only to use these tools, but also to understand their limitations, especially regarding context sensitivity and the comprehensibility of automated feedback [26]. Explainable AI (XAI) mechanisms need to be integrated into assessment platforms to enable lecturers and students to trace and understand how scores are generated. This is crucial in building trust, especially in formative and high-stakes assessments [24]. Educational institutions should also ensure that the use of AI does not marginalize students from diverse linguistic and cultural backgrounds, thereby reinforcing the need for systems that involve human intervention [27].

V.1 OUTCOME & CONCEPTUAL RELATIONSHIP SUMMARY

From a technological perspective, the future of AI-based subjective assessment lies in the development of hybrid models that combine the strengths of AI and human raters. This includes the use of domain-specific fine-tuning of large-scale language models (LLMs) and the integration of ethical reasoning frameworks in AI systems [19]. AI models also need to be able to handle subtle aspects of student responses such as creativity, ambiguity, and critical thinking, which are often poorly interpreted by existing models. Future systems should employ more adaptive architectures that can adjust assessment logic based on student level and learning objectives [23].

V.2 POLICY AND RESEARCH IMPLICATIONS

While the use of AI in education is becoming more widespread, there are still no global or institutional standard guidelines to regulate the use of automated assessment systems. Policy frameworks need to be developed collaboratively to ensure responsible implementation, data protection, and reduction of bias [13]. Future research should focus on the following areas:

- Development of cross-cultural and multilingual AI models for education
- Longitudinal studies of the impact of AI feedback on student learning
- Research on the effects of explainability on student motivation and trust
- Design of an audit system for AI scoring models in real classrooms

In addition, a participatory research approach involving educators and students is essential to ensure that the AI systems developed truly meet the needs of teaching and learning and remain accountable and ethical.

VI. CONCLUSIONS

This systematic review integrated evidence from 30 recent works (2020–2025) related to AI-driven subjective assessment in online learning, and achieved a meta-analysis, which filled the gap of evidence synthesized on the challenges and opportunities of state-of-the-art this area. Unlike previous reviews guided by predominately technical accuracy advances, this review combines pedagogical, technological and social/ethical viewpoints to provide a comprehensive overview of how AI can help enable fair and scalable assessments. The review identifies three significant and enduring barriers – transparency and explainability, contextual grasp and the absence of standardized guidelines for implementation – and notes some promising solutions (Explainable AI (XAI), human–AI hybrid models, policy frameworks for ethical governance). These results are synthesized into an enhanced conceptual framework, theoretically connecting AI capabilities to educational assessment theory and practically informing educational institutions on how to responsibly adopt AI. In the future, research could shift from technical evaluations towards longitudinal effects of AI feedback on student learning, cross-cultural and multilingual fairness in assessments, and policy-driven mechanisms for audit that bring accountability. Equally valuable is the participatory engagement of teachers and learners by co-designing AI systems that respond to the authentic utilities of schools. By organizing existing evidence around challenges, solutions, and future directions, this review provides a structured foundation for the responsible adoption of AI-driven subjective assessment in online education. In summary, AI has the potential to turn the subjective into something both scalable, transparent and pedagogically defensible. This is an accomplishment that can only be unlocked through the shared labour between researchers, policy-makers, and educators, ensuring that the systems of assessment powered by AI are not only novel but also trustworthy, inclusive, and ethically responsible.

VII. ACKNOWLEDMENT

This research was supported by Universiti Tun Hussien Onn Malaysia through Tier 1(vot H222).

VIII. REFERENCES

- [1] Al-Hasan, T. M., Sayed, A. N., Bensaali, F., Himeur, Y., Varlamis, I., & Dimitrakopoulos, G. (2024). From Traditional Recommender Systems to GPT-Based Chatbots: A Survey of Recent Developments and Future Directions. *Big Data and Cognitive Computing*, 8(4), 36. <https://doi.org/10.3390/bdcc8040036>.
- [2] S. Kasneci et al., “ChatGPT for good? On opportunities and challenges of large language models for education,” *Nature Machine Intelligence*, vol. 5, pp. 1004–1013, 2023. <https://doi.org/10.1038/s42256-023-00736-3>
- [3] Khraisha, Y., & Al-Hajj, M. (2024). Can Large Language Models Replace Humans in Systematic Reviews? *Research Synthesis Methods*. <https://doi.org/10.1002/jrsm.1715>
- [4] Mohammed, A., & Kora, R. (2025). A Comprehensive Overview and Analysis of Large Language Models: Trends and Challenges. *IEEE Access*, 13, 95850–95875. <https://doi.org/10.1109/ACCESS.2025.3573955>

- [5] Rawther, F. A., Rasaldeen, R., Fernandez, S. M., Jaison, I. R., & Mathews, R. M. (2025). A survey on automating answer-sheet evaluation using AI techniques. *International Journal on Emerging Research Areas (IJERA)*, 4(2), 72–76. <https://doi.org/10.5281/zenodo.14714351>
- [6] Yenduri, G., et al. (2024). GPT (Generative pre-trained transformer) A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access*, 12, 54608–54649. <https://doi.org/10.1109/ACCESS.2024.3389497>
- [7] Bharadiya, J. P. (2023). A comprehensive survey of deep learning techniques: Natural language processing. *European Journal of Technology*, 7(1), 58–66. <https://doi.org/10.47672/ejt.1473>
- [8] Dodia, S., Spoorthy, V., & Chandak, T. (2023). Machine learning-based automated system for subjective answer evaluation. In *2023 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)* (pp. 1–6). IEEE. <https://doi.org/10.1109/CONECCT57959.2023.10234818>
- [9] Alawida, M., Mejri, S., Mehmood, A., Chikhaoui, B., & Abiodun, O. I. (2023). A Comprehensive Study of ChatGPT: Advancements, Limitations, and Ethical Considerations in NLP and Cybersecurity. *Information*, 14(8), 462. <https://doi.org/10.3390/info14080462>
- [10] Khan, A., Khan, M. Z., Jamshed, S., Ahmad, S., Zainab, A., Khatib, K., Bibi, F., & Rehman, A. (2025). Advances in LLMs with focus on reasoning, adaptability, efficiency, and ethics. *arXiv*. <https://arxiv.org/abs/2506.12365>
- [11] Raiaan, M. A. K., Mukta, M. S. H., Fatema, K., Fahad, N. M., Sakib, S., Mim, M. M. J., Ahmad, J., Ali, M. E., & Azam, S. (2024). A review on large language models: Architectures, applications, taxonomies, open issues and challenges. *IEEE Access*, 12, 26839–26874. <https://doi.org/10.1109/ACCESS.2024.3365742>
- [12] Hemmat, A., Sharbaf, M., Kolahdouz Rahimi, S., Lano, K., & Yassipour Tehrani, S. (2025). Research directions for using LLM in software requirement engineering: A systematic review. *Frontiers in Computer Science*, 7, Article 1519437. <https://doi.org/10.3389/fcomp.2025.1519437>
- [13] Elbasi, E., Nadeem, M., Alzoubi, Y. I., Topcu, A. E., & Varghese, G. (2025). Machine Learning in Education: Innovations, Impacts, and Ethical Considerations. *IEEE Access*, 13, 128741–128770. <https://doi.org/10.1109/ACCESS.2025.3590134>
- [14] Yin, S., Xi, Y., Zhang, X., Sun, C., & Mao, Q. (2025). Foundation Models in Agriculture: A Comprehensive Review. *Agriculture*, 15(8), 847. <https://doi.org/10.3390/agriculture15080847>
- [15] Zsidai, B., Kaarre, J., Narup, E., Hamrin Senorski, E., Pareek, A., Grassi, A., Ley, C., Longo, U. G., Herbst, E., Hirschmann, M. T., Kopf, S., Seil, R., Tischer, T., Samuelsson, K., Feldt, R., & ESSKA Artificial Intelligence Working Group. (2024). A practical guide to the implementation of artificial intelligence in orthopaedic research Part 2: A technical introduction. *Journal of Experimental Orthopaedics*, 11(3), e12025. <https://doi.org/10.1002/jeo2.12025>
- [16] Chutia, T., & Baruah, N. (2024). A review on emotion detection by using deep learning techniques. *Artificial Intelligence Review*, 57, 203. <https://doi.org/10.1007/s10462-024-10831-1>
- [17] Naz, A., Khan, H. U., Bukhari, A., Anwar, S., Shujaat, S., & Muhammad, S. (2025). Machine and deep learning for personality traits detection: A comprehensive survey and open research challenges. *Artificial Intelligence Review*, 58, 239. <https://doi.org/10.1007/s10462-025-11245-3>
- [18] X. Zhai, "ChatGPT User Experience: Implications for Education," *SSRN Electron. J.*, 2022, doi:10.2139/ssrn.4312418.
- [19] Lu, J., Balasubramanian, B. K., Joy, M., & Xu, Q. (2025). Survey and Analysis for the Challenges in Computer Science to the Automation of Grading Systems. *ACM Computing Surveys*. <https://doi.org/10.1145/3748521>
- [20] J. Rudolph, S. Tan, and S. Tan, "ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?" *Journal of Applied Learning and Teaching*, vol. 6, no. 1, 2023.
- [21] Greco, C. M., & Tagarelli, A. (2024). Bringing order into the realm of Transformer-based language models for artificial intelligence and law. *Artificial Intelligence and Law*. Springer Nature. <https://doi.org/10.1007/s10506-023-09374-7>
- [22] Choi, J., & Lee, B. (2024). Accelerating materials language processing with large language models. *Communications Materials*, 5, 13. <https://doi.org/10.1038/s43246-024-00449-9>
- [23] Clay, B., Bergman, H. I., Salim, S., Pergola, G., Shalhoub, J., & Davies, A. H. (2025). Natural language processing techniques applied to the electronic health record in clinical research and practice. *Computers in Biology and Medicine*, 188, 109808. <https://doi.org/10.1016/j.combiomed.2025.109808>
- [24] Mohammadi, H., Bagheri, A., Giachanou, A., & Oberski, D. L. (2025). Explainability in Practice: A Survey of Explainable NLP Across Various Domains. *SAGE Journals*. <https://doi.org/10.1016/j.datak.2025.102470>
- [25] Ventura, J., Rodrigues, C., & Armando, N. (2025). Automatic Exam Correction System Involving XAI for Admission to Public Higher Education Institutions: Literature Review. *Proceedings*, 895–904. <https://doi.org/10.5220/0013429900003929>
- [26] Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90–112. <https://doi.org/10.1111/bjet.13370>
- [27] Rodrigues, L., Xavier, C., Costa, N., Gasevic, D., & Mello, R. F. (2025). Is GPT-4 fair? An empirical analysis in automatic short answer grading. *Computers and Education: Artificial Intelligence*, 8, 100428. <https://doi.org/10.1016/j.caeai.2025.100428>
- [28] Sharma, A. K., & Kunkel, J. (2025). A Review of Tools and Techniques for Optimization of Workload Mapping and Scheduling in Heterogeneous HPC System. <https://doi.org/10.1177>
- [29] Janumpally, R. K. (2025). The role of natural language processing in graduate medical education: A scoping review. *Cureus*, 17(3), e81078. <https://doi.org/10.7759/cureus.81078>
- [30] Kulsoom, U., Glavin, F. G., & Bendeche, M. (2025). Natural Language Processing and Machine Learning for Analysis of Radiology Reports: A Systematic Review. *IEEE Access*, 13, 112215–112254. <https://doi.org/10.1109/ACCESS.2025.3582728>
- [31] Patil, R., Boit, S., Gudivada, V., & Nandigam, J. (2023). A Survey of Text Representation and Embedding Techniques in NLP. *IEEE Access*, 11, 36120–36146. <https://doi.org/10.1109/ACCESS.2023.3266377>