

Daftar Pustaka

- [1] P. M. Jackaria, B. H. Hajan, dan A. R. H. Mastul, "A Comparative Analysis of the Rating of College Students' Essays by ChatGPT versus Human Raters," *International Journal of Learning, Teaching and Educational Research*, vol. 23, no. 2, pp. 478–492, Feb. 2024, doi: 10.26803/ijlter.23.2.23.
- [2] A. Mizumoto dan M. Eguchi, "Exploring the potential of using an AI language model for automated essay scoring," *Research Methods in Applied Linguistics*, vol. 2, no. 2, Aug. 2023, doi: 10.1016/j.rmal.2023.100050.
- [3] Y. Song, Q. Zhu, H. Wang, dan Q. Zheng, "Automated Essay Scoring and Revising Based on Open-Source Large Language Models," *IEEE Transactions on Learning Technologies*, vol. 17, pp. 1920–1930, 2024, doi: 10.1109/TLT.2024.3396873.
- [4] M. Kostic, H. F. Witschel, K. Hinkelmann, dan M. Spahic-Bogdanovic, "LLMs in Automated Essay Evaluation: A Case Study," 2024.
- [5] A. Pack, A. Barrett, dan J. Escalante, "Large language models and automated essay scoring of English language learner writing: Insights into validity and reliability," *Computers and Education: Artificial Intelligence*, vol. 6, Jun. 2024, doi: 10.1016/j.caeai.2024.100234.
- [6] W. Mansour, S. Albatarni, dan S. Eltanbouly, "Can Large Language Models Automatically Score Proficiency of Written Essays?," 2786. [Online]. Available: <https://www.kaggle.com/c/asap-aes>
- [7] K. Bouziane dan A. Bouziane, "AI versus human effectiveness in essay evaluation," *Discover Education*, vol. 3, no. 1, p. 201, Oct. 2024, doi: 10.1007/s44217-024- 00320-6.
- [8] E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," *Learn Individ Differ*, vol. 103, p. 102274, Apr. 2023, doi: 10.1016/J.LINDIF.2023.102274.
- [9] F. Yavuz, Ö. Çelik, dan G. Yavaş Çelik, "Utilizing large language models for EFL essay grading: An examination of reliability and validity in rubric-based assessments," *British Journal of Educational Technology*, 2024, doi: 10.1111/bjet.13494.
- [10] M. Stahl, L. Biermann, A. Nehring, dan H. Wachsmuth, "Exploring LLMS Prompting Strategies for Joint Essay Scoring and Umpan balik Generation," 2024. <https://github.com/webis-de/BEA-24>.
- [11] W. X. Zhao et al., "A Survey of Large Language Models," Mar. 2023, <http://arxiv.org/abs/2303.18223>.
- [12] Liew, P. Y., dan Tan, I. K. T. (2025). On automated essay grading using large language models. In *Proceedings of the 2024 8th International Conference on Computer Science and Artificial Intelligence (CSAI '24)* (pp. 204–211). Association for Computing Machinery.
- [13] Seßler, K., Fürstenberg, M., Bühler, B., dan Kasneci, E. (2025). Can AI grade your essays? A comparative analysis of large language models and teacher ratings in multidimensional essay scoring. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference (LAK '25)* (pp. 462–472). Association for Computing Machinery.
- [14] Xiao, C., Ma, W., Song, Q., Xu, S. X., Zhang, K., Wang, Y., dan Fu, Q. (2025). Human-AI collaborative essay scoring: A dual-process framework with LLMs. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference (LAK '25)* (pp. 293–305). Association for Computing Machinery.
- [15] Kundu, A., dan Barbosa, D. (2024). Are Large Language Models Good Essay Graders?. *arXiv preprint arXiv:2409.13120*.
- [16] Ishida, T., Liu, T., Wang, H., dan Cheung, W. K. (2024). Large language models as partners in student essay evaluation. *arXiv preprint arXiv:2405.18632*.
- [17] da Silva, W. A., dan de Araujo, C. C. Automated ENEM Essay Scoring and Umpan baliks: A Prompt-Driven LLMS Approach.
- [18] Flodén, J. (2025). Grading exams using large language models: A comparison between human and AI grading of exams in higher education using ChatGPT. *British Educational Research Journal*, 51, 201–224.

- [19] Lee, S., Cai, Y., Meng, D., Wang, Z., dan Wu, Y. (2024). Unleashing Large Language Models' Proficiency in Zero-shot Essay Scoring. arXiv preprint arXiv:2404.04941.
- [20] Tang, X., Chen, H., Lin, D., dan Li, K. (2024). Harnessing LLMs for multi-dimensional writing assessment: Reliability and alignment with human judgments. *Heliyon*, 10(14), e34262.
- [21] Harunasari, Siti. (2023). Examining the Effectiveness of AI-integrated Approach in EFL Writing: A Case of ChatGPT. *International Journal of Progressive Sciences and Technologies*. 39. 357-368. 10.52155/ijpsat.v39.2.5516.
- [22] J. . Zhao, Y. . Li, dan W. . Feng, "Investigating the Validity and Reliability of a Comprehensive Essay Evaluation Model of Integrating Manual Umpan balik and Intelligent Assistance", *Int. J. Emerg. Technol. Learn.*, vol. 18, no. 04, pp. pp. 248–261, Feb. 2023.
- [23] Zhao, X. (2022). Leveraging Artificial Intelligence (AI) Technology for English Writing: Introducing Wordtune as a Digital Writing Assistant for EFL Writers. *RELC Journal*, 54(3), 890-894. <https://doi.org/10.1177/00336882221094089>
- [24] J.Wisnu, "Workshop ChatGPT," <https://www.slideshare.net/slideshow/workshop-chatgpt-politeknik-masamipdf/262618640#15>