Jurnal Inovasi dan Teknologi Pembelajaran
Vol 11, No 3, (2024), page 137-148
https://doi.org/10.17977/um031v11i32024p137
P-ISSN: 2406-8780 E-ISSN: 2654-7953
Open access: http://journal2.um.ac.id/index.php/jinotep/index

AI-driven feedback system: Implementing advanced NLP and openAI for online learning

Liberius Sabinus Koe, Cecep Kustandi, Eveline Siregar
Master of Educational Technology, State University of Jakarta, Jakarta, Indonesia
Jl. Rawamangun Muka Raya No.11, Rawamangun, Kec. Pulo Gadung, East Jakarta City, Special Capital Region of Jakarta 13220
**Corresponding author, e-mail: lsbryanteacher@gmail.com*

ARTICLE INFO

Article history:
Received: 25-09-2024
Revised: 26-11-2024
Accepted: 30-11-2024

Keywords:
Artificial intelligence; NLP; OpenAI; Automatic feedback; Online learning

Keywords:
Artificial intelligence; NLP; OpenAI; Automatic feedback; Online learning

ABSTRACT

Penelitian ini bertujuan untuk mengembangkan umpan balik otomatis berbasis Artificial Intelligence (AI) dengan teknologi Natural Language Processing (NLP) dan GPT OpenAI dalam pembelajaran online. Jenis penelitian ini adalah penelitian pengembangan atau Research and Development (R&D) dengan model pengembangan Integrative Learning Design Framework (ILDF) sebagai acuan untuk merancang, memproduksi, serta menguji efektivitas produk. Produk yang dikembangkan berperan untuk menganalisis respons siswa secara otomatis, memberikan umpan balik yang cepat, relevan, serta menawarkan saran perbaikan secara real-time yang mencakup fitur-fitur utama seperti sentiment score, entities detection, syntax & grammar, correction, improvement suggestions, hingga relevance score. Pengembangan produk ini mencakup perancangan sistem hingga pengujian awal, namun tidak melibatkan evaluasi para ahli atau uji coba skala besar. Fokus penelitian adalah memastikan bahwa produk dapat berfungsi sesuai dengan rancangan teknis dan memenuhi kebutuhan awal pengguna. Hasil dari tahap pengembangan diharapkan dapat menjadi dasar bagi penelitian lanjutan dan membuka peluang baru untuk inovasi dalam teknologi pendidikan di masa depan.

ABSTRACT

This research aims to develop Artificial Intelligence (AI)-based automatic feedback with Natural Language Processing (NLP) technology and OpenAI GPT in online learning. The type of research is Research and Development (R&D) with the Integrative Learning Design Framework (ILDF) development model as a reference for designing, producing, and testing product effectiveness. The developed product plays a role in automatically analysing student responses, providing quick and relevant feedback, and offering real-time improvement suggestions, including key features such as sentiment score, entity detection, syntax & grammar, correction, improvement suggestions, and relevance score. The product development included system design to initial testing but did not involve expert evaluation or large-scale trials. The research focuses on ensuring that the product can function according to the technical design and fulfil the initial user needs. The results of the development phase are expected to form the basis for further research and open up new opportunities for innovation in educational technology in the future.

INTRODUCTION

Online learning is rapidly growing and serves as a significant alternative to traditional education worldwide (Farrell, 2020;Bayrak et al., 2020). The transition from in-person classes to online formats is now permanent, reflecting the increasing demand from students of all backgrounds (Yu & Jee, 2021). This learning allows the delivery of material mediated by the internet, intranets, extranets, satellite broadcasts, audio or video recordings, CDs, video conferencing, and computer-based training (Castro & Tumibay, 2021). One of the advantages of online learning is the flexibility that allows students to learn anywhere and anytime, overcoming geographical and time constraints that are often a barrier in traditional education (Zimmerman et al., 2020;Lakhal et al., 2021). It is evident that technology has affected the traditional education system, especially in educational environments with a large number of participants enrolled in a learning programme.

Technology not only acts as a tool but also a medium capable of changing the traditional educational paradigm. The field is evolving every day, thus requiring the ability to choose the right tools and follow a holistic perspective to improve teaching (Ugur et al., 2021). Technological innovation continues to lead to the development of artificial intelligence (AI), which is now also penetrating the education sector, potentially revolutionising education further and offering new methods for more personalised and adaptive learning (González-Calatayud, 2021). The growth of AI has driven an urgent need to understand how educators can best utilise these techniques for academic success. The use of various intelligence platforms and tools has enabled the improvement of teacher effectiveness and efficiency, resulting in richer or better teaching quality. Many of these AI models can then be utilised for learner profiling which enables predictions in offering timely support or providing feedback with guidance in the learning process (Zawacki-Richter et al., 2019). Vai and Sosulski (2011) view automated feedback as excellent for self-assessment. Such a system is claimed to help clarify, reinforce, and extend a learned topic. Automated feedback helps self-directed learning to be more effective so that it focuses on performance that can improve student learning. Thus, instructors and teachers can carry out their administrative functions, such as grading and providing feedback to students in a more effective way (Chen et al., 2020). Researchers see such methods as crucial to the future of education to reduce the burden of student performance assessment or feedback in online learning as well as in courses (del Gobbo et al., 2023).

With the development of artificial intelligence (AI) technology and the increasing adoption of online learning, there is a need to create more efficient and standardised feedback systems. The advent of automated feedback emerged as a solution to the heavy instructor workload of supporting a large number of learners enrolled in a course (Cavalcanti et al., 2021). Apart from workload, other conventional evaluation and feedback issues that need to be improved are due to the subjectivity of assessment, time consumption, ineffective peer assessment and feedback, and the traditional collective feedback style (Hussein et al., 2019; Nandini & Uma Maheswari, 2020; Pham et al., 2023). Although there have been many implementations of automated feedback in recent years, they still focus on technical model development as most of the development dimensions are still dominated by the computer science domain. As a result, the application of AI prediction models that consider pedagogical aspects to support teaching and learning procedures is still suboptimal (Zhai et al., 2021; González-Calatayud, 2021; Ouyang, 2023). This indicates a gap between technical developments and pedagogical needs, especially in the context of implementing feedback in education, which has driven the main reason for this research.

Along with the need to address such issues, this research also sought to explore the challenges faced in the practice of feedback in real educational settings. A survey conducted at SMA Negeri 2 Bajawa highlighted several challenges associated with conventional feedback from both teachers' and students' perspectives. Time constraints frequently inhibit teachers from providing detailed feedback, particularly in larger classes. Moreover, subjective judgment can result in inconsistencies; teachers' personal perceptions and emotional states may influence their evaluations, leading to disparities in feedback for students with similar abilities. Overall, the limitations of time and resources significantly hinder the ability to deliver thorough and timely evaluations for each student.

Conventional feedback can also present challenges for students. Many students have difficulty distinguishing between constructive feedback and less useful ones. The guidance in the feedback is often unclear, making it difficult for students to understand which aspect of their work needs to be improved. Another difficulty students face is the inability to identify their particular mistakes, often due to vague information, negative language, and lack of concrete examples. Dawson et al. (2019) highlighted students' weaknesses in identifying the strengths and weaknesses of their own work. Feedback becomes meaningless as students are unable to find areas where improvements are needed to improve their work. The same was also explored by Van der Kleij, (2019) that feedback becomes ineffective due to a lack of student engagement; students perceive feedback as useless; they do not pay attention to it, do not have time, and are unwilling or unable to use it.

The above problems and challenges of conventional feedback have prompted the need to develop an innovative solution to provide feedback that is standardised, fast, accurate, and personalised according to individual needs. The developed system is not only expected to ease the teacher's workload but also provide automatic feedback on student responses in a structured and in-depth manner. This research aims to develop *automatic feedback* based on artificial intelligence (AI) by utilising *Natural Language Processing* (NLP) technology and advanced models from OpenAI to analyse student responses automatically. The system is developed to provide fast, accurate, and specific feedback, which is expected to improve learning effectiveness by giving students a clearer guide to the learning process and understanding the strengths and areas of improvement in their work to drive overall learning quality improvement.

METHOD

This research uses the R&D development model with the process of analyzing, planning, producing, and testing products (Sugiyono, 2022). The selection of R&D was to adjust to the development of AI in the learning environment. Thus, in the AI development process, researchers used the Integrative Learning Design Framework (ILDF) model as shown in Figure 1. The use of ILDF is in line with current learning conditions that lead to the practical implementation of AI (Bannan-Ritland, 2003).

The ILDF framework consists of three primary phases: exploration, enactment, and evaluation. Each phase includes specific steps designed to ensure a systematic development process that addresses user needs. The exploration phase is the initial stage of the ILDF development paradigm. In this phase, the steps taken are needs analysis, literature survey, theory development and user character analysis. The exploration phase also involves analysing relevant AI technologies to determine how they can be integrated into the system. In this phase, the initial technical specifications and conceptual design of the product to be developed, including the main features and workflow of the system, were formulated.

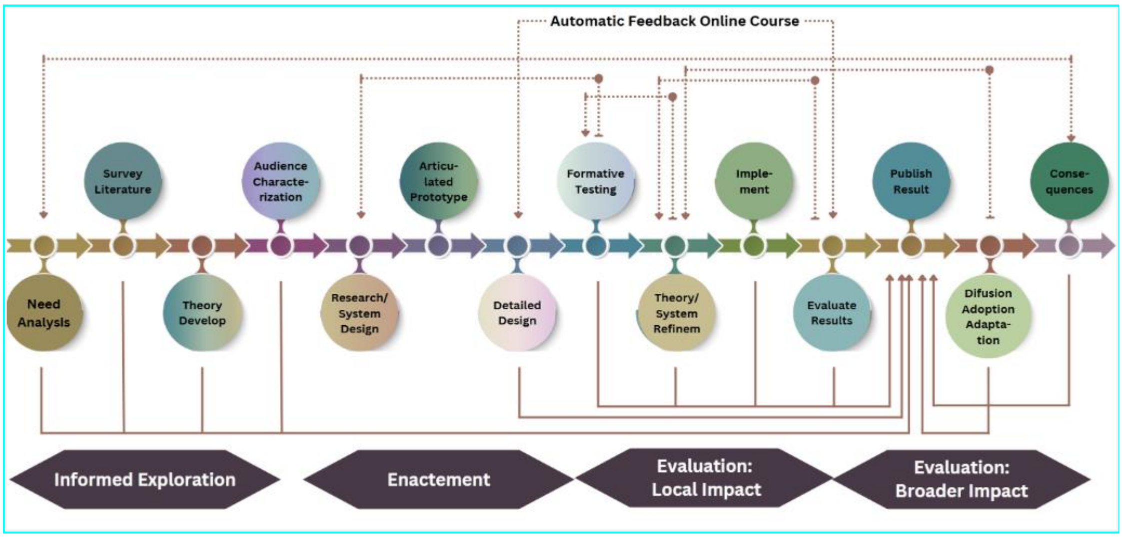

Figure 1. ILDF model (Shelton & Scoresby, 2011)

The second phase is the implementation or enactment stage. This phase consists of several steps: system design, prototype articulation, and detailed product design. This implementation phase includes the creation of the initial prototype of the automatic feedback product. Development began with building the system following the design, including developing an NLP module that could analyse student responses and provide automatic feedback. The resulting prototype was tested internally to ensure each function was working properly and by the specified specifications.

The third phase is the evaluation stage. In this phase, the steps taken are formative evaluations such as material expert, media expert, and learning expert tests to ensure that the content, media, and learning design used in automatic feedback products follow the required standards. In addition to expert testing, in this phase, one-to-one and small-group tests were conducted to collect feedback from direct users, namely students and teachers. The input from these two tests became the material for product improvement, which was later ready to be tested in a wider context. This evaluation phase ensures that the automatic feedback product developed is not only accurate in the analysis and feedback provided but also easy to use and by user needs.

The development of automatic feedback follows the development flow of the ILDF model but is only limited to the initial testing stage without involving direct input from experts or large-scale trials. The focus of the research was to ensure that the developed product functions according to the technical design and specifications and fulfils the initial needs of providing automatic feedback to students. Initial testing was conducted to assess the functionality of the product, including the key features required. The initial development of this product is part of a long-term development plan, so the study is the initial stage of a broader set of research. A vital objective of this study was to develop a product that could form the basis for further research and development in the future.

RESULT

The development of this automatic feedback system was conducted in two main stages, namely the exploration stage and the design implementation stage. However, it did not continue to the third phase, which is the evaluation phase. The following is a detailed explanation of the two phases:

Exploration Phase

This exploratory phase aims to identify the real needs in the learning environment and examine the challenges faced by students and teachers regarding conventional feedback practices. Through needs analysis and a literature survey, an in-depth understanding of the existing problems, such as time constraints, subjectivity of assessment, and lack of clarity in the feedback guidelines provided to students, was obtained. The user characteristics analysis also helped to understand how students process and utilise the feedback they receive, which is crucial in designing an effective automated feedback system as shown in Figure 2.

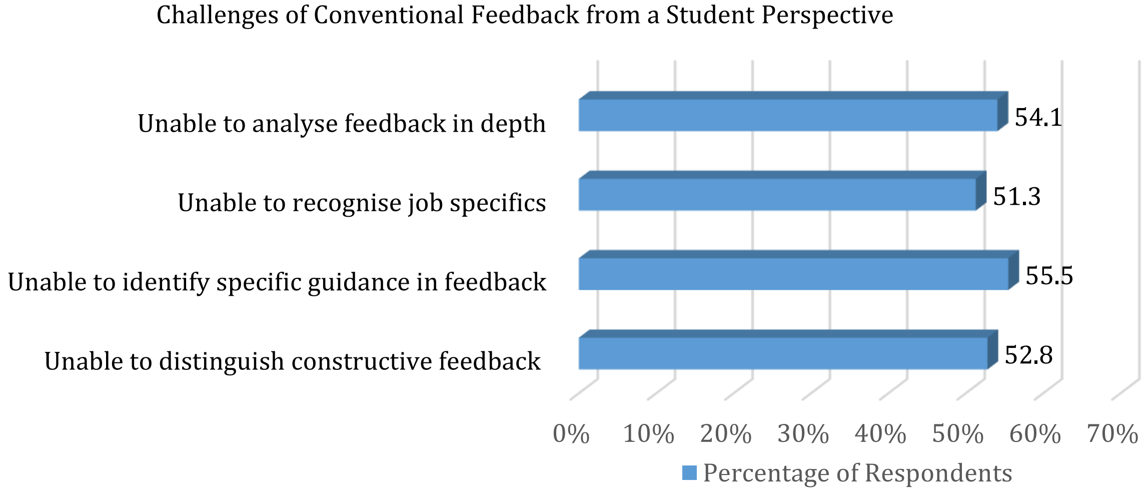

Figure 2. Conventional feedback challenge

The challenges of conventional feedback from the students' perspective are faced as learning problems that need to be addressed. A survey conducted on students at SMA Negeri 2 Bajawa with a total sample of 72 respondents in class X and class XI, revealed several challenges that occur in conventional feedback including: (1) 52.8% of the total respondents admitted that they were unable to differentiate between constructive feedback and less useful feedback; (2) some students (55.5%) reported that they were also unable to clearly identify guidelines in the feedback they received from teachers for the improvement of their work; (3) as many as 51.3% of learners revealed that although they received feedback from teachers, they were also unable to recognise specific errors in their work due to things such as the vagueness of information, the use of negative language, the lack of concrete examples, as well as the mismatch between the feedback provided and the expected assessment standards; (4) another finding from the survey was the admission of 54.1% of the total respondents who evaluated their ability after receiving feedback, that they were unable to analyse in depth the information from the various feedback they received in the learning process.

Implementation Phase

The resulting product is an *automatic feedback* system as shown in Figure 3. The AI technology used is *Natural Language Processing* (NLP) by utilising the main *library from* Google Cloud Platform (GCP). Meanwhile, the OpenAI models chosen are GPT-3.5 Turbo, GPT-4, and GPT 4 (0613). Several GPT models were selected to compare the analysis results provided by each model in terms of speed, depth of analysis, and relevance of feedback. The system can give accurate, timely, and learning-appropriate input because of the combination of OpenAI's NLP and GPT models. The automatic feedback product was incorporated into a learning management system (LMS) that uses Moodle.

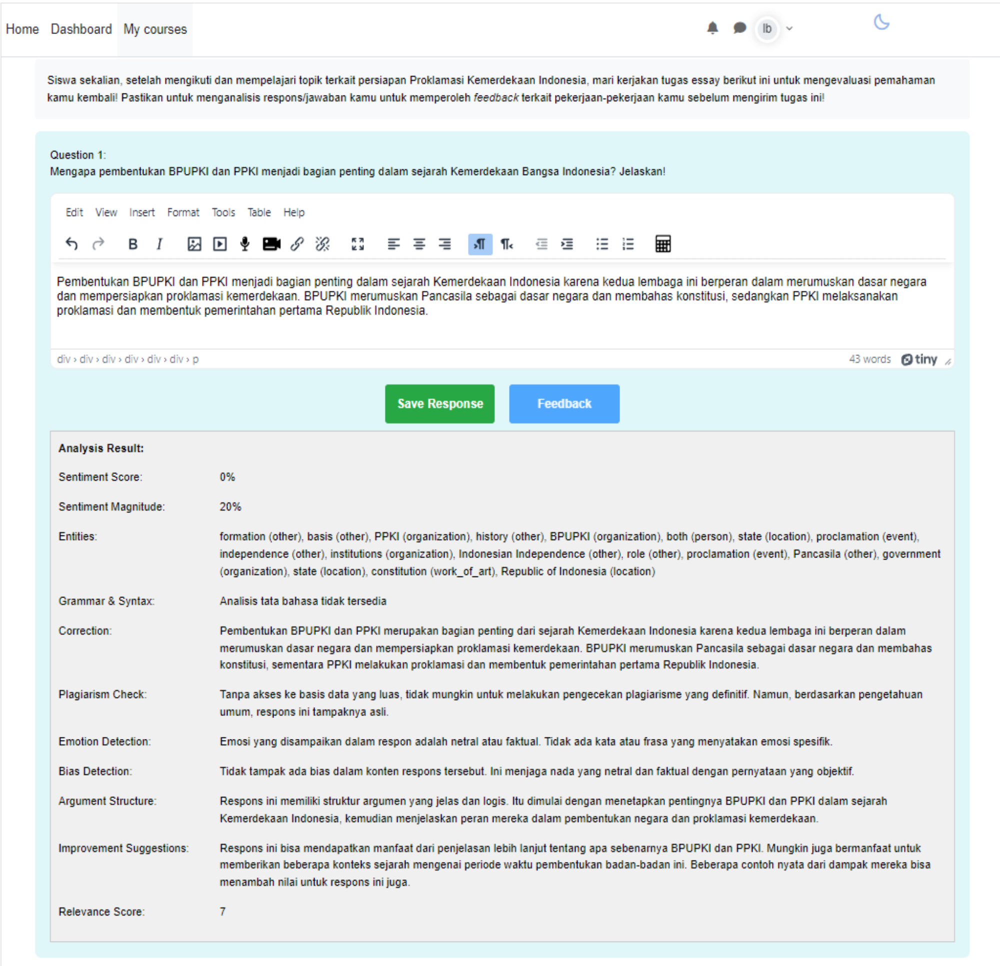

Figure 3. Automatic feedback analysis result interface

Table 1. Tested features and functions

No.	Features tested	Function description	Testing status
1	Sentiment score	Analyse student responses to show positive, negative or neutral sentiments.	Successful
2	Sentiment magnitude	Determines how strong or intense the sentiment is in the student's response.	Successful
3	Entities	Detect key entities such as name, location, relevant events in the response.	Successful
4	Grammar & Syntax	Checking grammar and sentence structure in student responses.	Successful
5	Correction	Provide corrections for detected grammar and sentence structure errors.	Successful
6	Plagiarism check	Checking for possible plagiarism in student responses.	Need further development
7	Emotion detection	Analyse the emotions expressed in the student response text.	Successful
8	Bias detection	Identifying biases that may be contained in student responses.	Successful
9	Argument structure	Evaluate the logical structure and flow of arguments in student responses.	Successful
10	Improvement suggestions	Provide suggestions for improvement related to strengthening arguments and writing.	Successful
11	Relevance score	Determine the relevance of the student's response to the given question.	Successful

According to the findings of the product's internal test (Table 1), depending on the number of characters analysed, the system could deliver feedback for each student response in an average of two to three minutes. The analysis finished in less than two minutes if the student responses were brief; the longer the responses, the longer the time. The intricacy of the text being analysed, including the number of entities identified, the intricacy of the syntax, and the breadth of the remedial recommendations made will affect this processing time. This internal test focused on feedback accuracy and speed. Although there were some areas for improvement, the algorithm did a good job of detecting grammatical problems in nearly all of the replies that were analysed. The corrections provided can also include improvements to sentence structure, punctuation errors, and the correspondence between the subject and predicate in the sentence. The system successfully detects some key elements, such as event entities corresponding to the given topic, ensuring that student responses remain relevant to the context of the question. The system can also detect plagiarism, although this feature still needs further development.

In initial testing, the product showed the potential to detect phrases or sentences that are similar to other relevant sources. However, analyses relevant to this still require integration with larger databases for this feature to function optimally in a wider learning environment. One of the key features tested was the system's ability to provide detailed improvement suggestions, both in terms of grammar and argument strengthening. The advice provided not only focuses on errors but also on how to improve argument structure, provide stronger supporting evidence, and clarify the logic of reasoning in student responses. This feature is crucial to encourage the improvement of learning quality, where data accuracy and argumentation skills are required in the learning process.

DISCUSSION

The automatic feedback product development started with an exploratory phase through a needs analysis to identify the main challenges in providing feedback in a conventional learning environment. The results obtained from the needs analysis imply that conventional feedback, in terms of time, quality, and personalisation, has not been able to meet students' needs optimally. Teachers' limited time to provide detailed evaluations and the subjectivity of judgment have led to inconsistencies in the quality of feedback. In addition, the feedback delivered tends to be less targeted and inadequate in providing specific guidance for improvement. Conventional feedback is not enough to help students significantly improve the quality of their work because the information is unclear and sometimes irrelevant to the assessment standards.

This finding underscores the provision of conventional feedback that has not fully reflected the principles of behaviorism as proposed by Skinner that learning will take place very effectively if learners are immediately given *feedback* on the accuracy of their learning (Prabawa et al., 2019). This learning theory sees feedback as *reinforcement* to strengthen more effective and correct learning behaviour (Slamet, 2020). Meanwhile, Gagne (1990) views feedback as reinforcement for the work given to students as one part of the nine learning processes or *instructional events* that can foster good learning activities and cognitive processes (Al-Mahiroh & Suyadi, 2020). Gagne sees feedback as an essential part of reinforcing and directing students' learning process, including helping them revise their understanding to achieve better learning goals.

Based on the principles of behaviourism theory, an innovative approach is needed to provide feedback that is more adaptive to the needs of learners. The results of this study show that audience character is required when conducting the feedback development process. In this study, the intended audience characters are students and teachers. The product development carried out by researchers not only focuses on teachers but also students who are the ultimate goal of the product. Thus, the feedback developed must be relevant, personalised, and have good timeliness. This condition is supported by technological advances in the digital era. Students have become accustomed to using technology and interacting quickly. Therefore, the feedback in this product must be able to facilitate the student's access to learning promptly. In addition, guidance is one of the significant aspects of product development. Although students in today's digital era are able to access information independently, they also need special guidance to make it easier to understand the material in the developed product. With the guide, they can also foster independent learning.

On the other hand, Open AI, developed by the researchers, also focuses on the teacher audience. Teachers become feedback givers whose characteristics are different from students, given that teachers need a system that can facilitate them in conducting evaluations. Moreover, the number of students handled by teachers is not low. So, to make time efficient, NLP and OpenAI are needed to maintain the quality of feedback given to students. This product development can also help teachers, making the evaluation much better than before and providing easy access. This condition is in line with previous research where the use of AI significantly impacts the world of education. Therefore, NLP and OpenAI are solutions that can assist teacher performance in the evaluation process. This is because NLP is commonly used in online learning that can support interactive learning, such as automatically grading exams (Süzen et al., 2020;Vo et al., 2022;Zhai et al., 2021). In addition, NLP can also classify the topics discussed and execute performance assessments of the words written as a response quality evaluation tool (Urrutia & Araya, 2023). Thus, NLP is suitable as a feedback programme. In contrast, OpenAI is with GPT (Generative Pre-trained Transformer) models such as GPT-3.5 and GPT-4. OpenAI is familiar to users in the digital age in processing and showing natural language results similar to human understanding. Students today are familiar with the use of ChatGPT in the learning process. This is because ChatGPT is an artificial intelligence that can be used as a source of information to process reasonable or creative responses to a question (French et al., 2023). Therefore, ChatGPT OpenAI is an alternative system that can be a tool for students to perform language translation and classify data-based recommendations (Fitria, 2023).

NLP and OpenAI are technologies in the world of education with the potential to improve the quality of feedback in online learning. Thus, users can perform language analysis and also provide personalised recommendations. In addition, NLP and OpenAI can make online learning capable of producing accurate and relevant responses. Thus, NLP and OpenAI can be solutions for teachers and students needed in online learning. The combination of NLP and GPT OpenAI can encourage the development of automatic feedback to accelerate the evaluation process. Thus, students, as users, get timely and relevant feedback. Therefore, this automated AI product can be applied systematically according to the needs of teacher and student audiences. Several components enable the NLP and GPT OpenAI systems to work efficiently and responsively. Firstly, the development of the user interface was through direct integration into the *Learning Management System* (LMS), specifically choosing Moodle as the primary platform. Moodle was selected due to its high flexibility, *open-source* nature, and wide use in various educational institutions across the world (Makruf et al., 2022). Moodle also supports API integration for the development of additional features and has an active user community, which enables support and further development. Moodle is the best *platform* to achieve educational effectiveness (Rodriguez et al., 2023). This system integration permits the *automatic feedback* feature to operate well within the Moodle ecosystem. A course activity module was developed to allow students to submit their essay responses directly through the platform.

Secondly, NLP integration is done with the Google Cloud Natural Language API as the main library to perform text analysis functions, especially in grammar, sentence structure, entity detection, and sentiment analysis. This API also analyses sentiment and emotion in writing and the feedback is technical and contextual. For example, errors in writing are identified, suggestions for improvement are provided, and the emotion and tone of the writing are evaluated to help students understand the impact of how they convey ideas. The findings of this research illustrate that Google Cloud Natural Language API faces challenges in handling questions from the Indonesian language. This is because the NLP is designed to handle prominent world languages. As, indeed, the Indonesian language is limited, for instance, in terms of the use of compound words, idioms, and regional variations, which are often difficult to process with precision by NLP algorithms, additional customisations are made so that the system can provide relevant and meaningful feedback despite the limitations of Indonesian language analysis. This challenge is an opportunity for future developers and researchers to improve the analysis of the Indonesian language by training specialised NLP models that can handle variations in the language better.

Thirdly, the integration of the automatic feedback system with OpenAI utilises several available GPT models. The system can choose which model to use, such as the GPT-3.5 Turbo model, GPT-4, or the newer model, to provide feedback to students. These two models have significant differences based on the feedback provided. The GPT-3.5 Turbo model can be used for quick analyses, but the results tend to lack consistency and depth. For example, in terms of providing corrective feedback, it may only provide basic guidance on grammatical or sentence structure errors but does not always highlight more complex flaws or the relevance of the answer to the question. In contrast, GPT-4 provides more in-depth and consistent feedback. It not only corrects basic errors but also helps students understand the strengths and weaknesses of their arguments, provides more relevant suggestions for improvement, and evaluates answers more accurately according to the context of the question. However, it should be noted that GPT-4 is resource-intensive, both in terms of analysis time and cost, so its use may be more appropriate for analyses that require greater depth and rigour. Users need to consider their needs and available resources before deciding which OpenAI GPT model to use.

The initial test results show that the developed *automatic feedback* product can provide fast and relevant feedback on student responses. It includes several vital elements, including (1) *sentiment score* and *sentiment magnitude* that show students' responses that contain positive, negative, or neutral emotions. This analysis is particularly crucial for some learning cases that require a deep understanding of students' emotional responses, such as in language learning, where teachers need to gauge whether students understand the emotional nuances in their essays or reflections. In social learning, *sentiment score* can also help identify how students respond to meaningful events with positive, negative, or neutral feelings, which could indicate their understanding or engagement with the material. (2) *Entities* that can identify responses such as names, organisations, places, or events. This analysis is very appropriate for learning processes involving factual material and in-depth analysis, such as history learning that identifies names of figures and events or geography education that identifies names of places, cities, or influential landmarks. In science lessons, entity identification can be very practical for recognising scientific concepts and names of chemical compounds or organisms. Thus, the system's ability to detect entities is very helpful in ensuring that students mention accurate and relevant information according to the field of study being discussed. (3) *Grammar & syntax* to detect gr*ammar* and sentence structure errors in students' responses, which is very substantial in language learning and social studies and science. (4) *Correction* to support immediate correction of detected errors, helping students understand specific errors in their answers and providing concrete guidance to correct them. In the learning process, *correction* is very invaluable for students to improve their work in *real time.* (5) *Plagiarism check* is used to detect potential plagiarism in student responses. Although this feature still requires further development due to technological limitations, *plagiarism check* has shown significant potential in detecting similarities with other available sources. This feature is vital in maintaining academic integrity, especially in tasks involving writing essays, scientific reports, or research. (6) *Emotion detection* and bias *detection* to assess student responses that contain certain unintended emotions or biases. This feature is especially beneficial in some learning, such as social sciences or languages, where students are to objectively and fairly respond to events or narratives. (7) *Argument structure* evaluates how students structure their arguments, including the logic and interconnectedness of ideas presented. This is particularly necessary in areas that require critical thinking skills and crafting clear and coherent arguments. This feature helps identify weaknesses in students' logical flow and provides suggestions to strengthen their argument structure. (8) *Improvement suggestions* can provide specific guidance to improve the quality of students' responses. These can be suggestions for reorganising arguments, adding supporting evidence, or correcting unclear passages. This feature is very useful in all subject areas, especially those that involve writing or organising ideas systematically. (9) *Relevance score* is used to assess the extent to which students' responses are relevant to the given question or topic. This analysis is crucial in ensuring that students do not just answer the question in general terms but also focus on the core of what is asked in the question. The *relevance score* feature helps measure how well students understand and respond to questions with appropriate and related information and ensures that they truly capture the essence of the topic at hand.

The features generated in the successfully developed *automatic feedback* have supported various forms of expected feedback. Feedback can include varied analyses, such as correct statements or statements that need improvement for detailed explanations. This finding is in line with experts' views that feedback in the learning process should be in the form of incorrect or correct statements, providing the correct answer, explanation, and additional teaching or concepts for reinforcement (Sofyatiningrum et al., 2019). This comprehensive feedback approach ensures that students recognise their mistakes and receive the necessary guidance that improves their performance effectively. Moreover, the system can provide prompt and contextually relevant feedback, following the principles outlined in Skinner's behaviourist learning theory, where reinforcing stimuli enhances responses (Chyung, 2008). The system's ability to provide timely and detailed feedback underscores its alignment with existing educational theories while enhancing its utility in diverse learning environments.

Other features such as *sentiment score, grammar correction, and relevance score* can help students understand their performance, guiding them on what went wrong and how to improve it. These features also uphold the principle of feedback expressed by Nicol & Macfarlane-Dick (2006), stating that good feedback should clarify performance criteria, stimulate reflection, and uplift students. Feedback should be a method to reinforce reflective activities by providing information to the learner about the state of learning they are doing (Wong et al., 2019). Feedback should also be considered a decisive form of communication that gives the communicator or the informer an idea of the results of his communication; in this case, the teacher can understand how students respond and process the learning materials (Murniarti, 2019). Features such as *correction* and *improvement* suggestions can support continuous teaching by providing concrete suggestions for improvement while allowing teachers to monitor the gap between current student performance and expected targets, ultimately strengthening the adaptive and responsive learning process. Thus, the developed automated feedback is not only an evaluation tool but also a means of effective communication between students and teachers in the learning process.

CONCLUSION

The development of *artificial intelligence* (AI)-based *automatic feedback* system with *Natural Language Processing* (NLP) and OpenAI technology is still at an early stage of development, but the results have shown promising potential to be further developed and utilised in the online learning process. Features such as *sentiment score, entities detection, and relevance score* allow students to understand their mistakes, provide concrete guidance for improvement, and improve comprehension and learning quality. In learning contexts involving arguments or essay writing, the system can facilitate a more structured evaluation with remedial suggestions appropriate to the context of the student's response. Nonetheless, some challenges were encountered, including limitations in the analysis of more complex Indonesian language, especially in *syntax* detection and *grammar correction*, and the need to extend the *plagiarism check* feature to provide a more in-depth evaluation of the originality of student work. Along with the advancement of artificial intelligence (AI) technology that increasingly enables large-scale data processing, automated feedback systems have great potential to be a superior solution in improving the quality of online learning sustainably. The developed system can serve as a foundation for future research in the development of more adaptive and contextual features, such as personalisation of feedback based on student profiles, improvement of text analysis capabilities in various disciplines, and optimisation of the use of AI to detect more in-depth elements in student responses. Further research could focus on integrating the system with project-based or collaborative learning methods and large-scale testing to ensure consistency, accuracy, and effectiveness of this *automatic feedback* in a wider learning context.

Author contributions

The authors contributed significantly to the conception and design of the study. They were responsible for the data analysis, interpretation, and discussion of the results. The final manuscript was reviewed and approved by the authors.

Funding

Grants for Master's Thesis Basic Research Scheme Directorate of Research, Technology and Community Service Ministry of Education, Culture, Research and Technology Year 2024**.**

Conflict of interest

The authors declare that there is no potential conflict of interest.

Data availability statement

All data are available from the authors.

REFERENCES

Al-Mahiroh, R. S., & Suyadi, S. (2020). Kontribusi Teori Kognitif Robert M. Gagne Dalam Pembelajaran Pendidikan Agama Islam. *QALAMUNA: Jurnal Pendidikan, Sosial, dan Agama*, *12*(2), 117-126. https://doi.org/10.37680/qalamuna.v12i2.353

Bannan-Ritland, B. (2003). The Role of Design in Research: The Integrative Learning Design Framework. *Educational Researcher*, *32*(1), 21-24. https://doi.org/10.3102/0013189X032001021

Bayrak, D. F., Moanes, D. M., & Altun, D. A. (2020). Development of Online Course Satisfaction Scale. *Turkish Online Journal of Distance Education*, *21*(4), 110-122. https://doi.org/10.17718/TOJDE.803378

Castro, M. D. B., & Tumibay, G. M. (2021). A literature review: efficacy of online learning courses for higher education institutions using meta-analysis. *Education and Information Technologies*, *26*(2), 1367-1385. https://doi.org/10.1007/s10639-019-10027-z

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y. S., Gašević, D., & Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, *2*, 100027. https://doi.org/10.1016/j.caeai.2021.100027

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, *8*, 75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510

Chyung, S. Y. (2008). *Foundations of Instructional and Performance Technology* (Bay Suzanne & Farnham

Sally (eds.)). HRD Press, Inc.

Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2019). What makes for effective feedback: Staff and student perspectives. *Assessment & Evaluation in Higher Education*, *44*(1), 25-36. https://doi.org/10.1080/02602938.2018.1467877

del Gobbo, E., Guarino, A., Cafarelli, B., Grilli, L., & Limone, P. (2023). Automatic evaluation of open-ended questions for online learning. A systematic mapping. *Studies in Educational Evaluation*, *77*, 101258. https://doi.org/10.1016/j.stueduc.2023.101258

Farrell, O. (2020). A balancing act: a window into online student engagement experiences. *International Journal of Educational Technology in Higher Education*, *17*(1). https://doi.org/10.1186/s41239-020-00199-x

Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *ELT Forum: Journal of English Language Teaching*, *12*(1), 44-58. https://doi.org/10.15294/elt.v12i1.64069

French, F., Levi, D., Maczo, C., Simonaityte, A., Triantafyllidis, S., & Varda, G. (2023). Creative use of OpenAI in education: case studies from game development. *Multimodal Technologies and Interaction*, *7*(8), 81. https://doi.org/10.3390/mti7080081

González-Calatayud, V. (2021). Artificial intelligence for student assessment: A systematic review. *Applied Sciences (Switzerland)*, *11*(12). https://doi.org/10.3390/app11125467

Hussein, M. A., Hassan, H., & Nassef, M. (2019). Automated language essay scoring systems: A literature review. *PeerJ Computer Science*, *2019*(8). https://doi.org/10.7717/peerj-cs.208

Lakhal, S., Khechine, H., & Mukamurera, J. (2021). Explaining persistence in online courses in higher education: a difference-in-differences analysis. *International Journal of Educational Technology in Higher Education*, *18*(1), 1-32. https://doi.org/10.1186/s41239-021-00251-4

Makruf, I., Rifa'i, A. A., & Triana, Y. (2022). Moodle-based online learning management in higher education. *International Journal of Instruction*, *15*(1), 135-152. https://doi.org/10.29333/iji.2022.1518a

Murniarti, E. (2019). *Komunikator, pesan, pedia/saluran, komunikan, efek/hasil, dan umpan balik.* Faculty of Teacher Training and Education, Universitas Kristen Indonesia Jakarta.

Nandini, V., & Uma Maheswari, P. (2020). Automatic assessment of descriptive answers in online examination system using semantic relational features. Journal of Supercomputing, 76(6), 4430-4448. https://doi.org/10.1007/s11227-018-2

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199-218. https://doi.org/10.1080/03075070600572090

Ouyang, F. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering courses. *International Journal of Educational Technology in Higher Education*, *20*(1), 4. https://doi.org/10.1186/s41239-022-00372-4

Pham, M., Singh, K., & Jahnke, I. (2023). Socio-technical-pedagogical usability of online courses for older adult learners. *Interactive Learning Environments*, *31*(5), 2855-2871. https://doi.org/10.1080/10494820.2021.1912784

Prabawa, I. P. A. E., Kanca, I. N., & Wijaya, I. M. A. (2019). Pengaruh metode pembelajaran reciprocal berbantuan feedback visual terhadap hasil belajar lompat jauh pada peserta didik kelas VIII SMP Negeri 1 Mengwi tahun pelajaran 2018/2019. *Jurnal Pendidikan Jasmani, Olahraga dan Kesehatan Undiksha*, *7*(2), 45-52.

Rodriguez, C. A., Valderrama, S., Vargas, D., Eliseo, M. A., Fracchia, C. C., & Roa, K. (2023). Quizzes via Augmented Reality on Learning Management System: A Case Study of Moodle. *Journal of Educators Online*, *20*(1), n1.

Shelton, B. E., & Scoresby, J. (2011). Aligning game activity with educational goals: Following a constrained design approach to instructional computer games. *Educational Technology Research and Development*, *59*(1), 113-138. https://doi.org/10.1007/s11423-010-9175-0

Slamet, S. S. (2020). Hubungan strategi umpan balik (feedback), motivasi berprestasi dan hasil belajar dalam pembelajaran PPKn di SMK. *PINUS: Jurnal Penelitian Inovasi Pembelajaran*, *5*(2), 39-56. https://doi.org/10.29407/pn.v5i2.14539

Sofyatiningrum, E., Ulumudin, I., & Perwitasari, F. (2019). Kajian umpan balik guru terhadap hasil belajar siswa. *Indonesian Journal of Educational Assesment*, *2*(2), 56-65.

Sugiyono. (2022). Research & Development Methods: for Education, Management, Social and Engineering. *Bandung: Alfabeta Publisher*, 28.

Süzen, N., Gorban, A. N., Levesley, J., & Mirkes, E. M. (2020). Automatic short answer grading and feedback using text mining methods. *Procedia Computer Science*, *169*, 726-743. https://doi.org/10.1016/j.procs.2020.02.171

Ugur, B., Arkun Kocadere, S., Nuhoglu Kibar, P., & Bayrak, F. (2021). An open online course to enhance digital competences of teachers. *Turkish Online Journal of Distance Education*, *22*(4), 24-42.

Urrutia, F., & Araya, R. (2023). Automatically Detecting Incoherent Written Math Answers of Fourth-Graders. *Systems*, *11*(7). https://doi.org/10.3390/systems11070353

Vai, M., & Sosulski, K. (2011). *Essentials of online course design: A standards-based guide*. Routledge.

van der Kleij, F. M. (2019). Comparison of teacher and student perceptions of formative assessment feedback practices and associations with individual student characteristics. *Teaching and Teacher Education*, *85*, 175-189. https://doi.org/10.1016/j.tate.2019.06.010

Vo, N. N. Y., Vu, Q. T., Vu, N. H., Vu, T. A., Mach, B. D., & Xu, G. (2022). Domain-specific NLP system to support learning path and curriculum design at tech universities. *Computers and Education: Artificial Intelligence*, *3.* https://doi.org/10.1016/j.caeai.2021.100042

Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G. J., & Paas, F. (2019). Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review. *International Journal of Human-Computer Interaction*, *35*(4-5), 356-373. https://doi.org/10.1080/10447318.2018.1543084

Yu, J., & Jee, Y. (2021). Analysis of online classes in physical education during the covid-19 pandemic. *Education Sciences*, *11*(1), 1-14. https://doi.org/10.3390/EDUCSCI11010003

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. *International Journal of Educational Technology in Higher Education*, *16*(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J. B., Yuan, J., & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, *2021.* https://doi.org/10.1155/2021/8812542

Zimmerman, W., Altman, B., Simunich, B., Shattuck, K., & Burch, B. (2020). Evaluating online course quality: A study on implementation of course quality standards. *Online Learning Journal*, *24*(4), 147-163. https://doi.org/10.24059/olj.v24i4.2325