Smart Assessment and Guided Education with Responsible AI

Alexander Mikroyannidis[1][0000-0002-9518-1443], John Domingue[1], Aisling Third[1], David Tarrant[2] and Tom Pieroni[2]

[1] Knowledge Media Institute, The Open University, Milton Keynes MK7 6AA, UK
[2] Open Data Institute, 90 York Way, London N1 9AG, UK

Abstract. This paper presents the vision of the SAGE-RAI (Smart Assessment and Guided Education with Responsible AI) project. Inspired by Bloom's seminal work on the efficacy of personalised learning, the SAGE-RAI project aims to leverage responsible Generative AI towards transforming teaching and learning practices for improved student outcomes. By exploring the integration of Generative AI into tutoring processes, we seek to provide scalable, personalised learning experiences for large cohorts of students. Our research focuses on harnessing Generative AI to offer tailored educational content and generate constructive feedback for students. By applying responsible AI practices, we aim to mitigate issues such as misinformation, copyright infringement, and bias.

Keywords: Generative AI, Responsible AI, Personalised Learning.

1 Introduction

The education community has been motivated for decades by Bloom's famous 1984 study [1]. As shown in Fig. 1, this study found that students taught 1-to-1 performed 2 standard deviations better than students taught in a standard classroom setting, with the average 1-to-1 student performing better than 98% of the students taught in the control group. Access to personalised teaching limited by socioeconomic issues is a problem for equality. AI offers the potential to unlock low-cost personalised teaching to massively assist both students and teachers and dramatically improve learning outcomes, while widening access for more, and more diverse, learners.

The advent of Generative AI (GenAI) has brought this dream closer and has captured the attention of many educators. For example, a recent study of more than 1,000 K-12 teachers and 1,000 students found that 51% of teachers reported using ChatGPT

- 40% weekly.[1] Interest and take-up is similar in the Higher Education Arena. For example, an analysis of eight Russell Group universities found that there were over 1 million visits to the ChatGPT site during December 2022 and January 2023.[2]

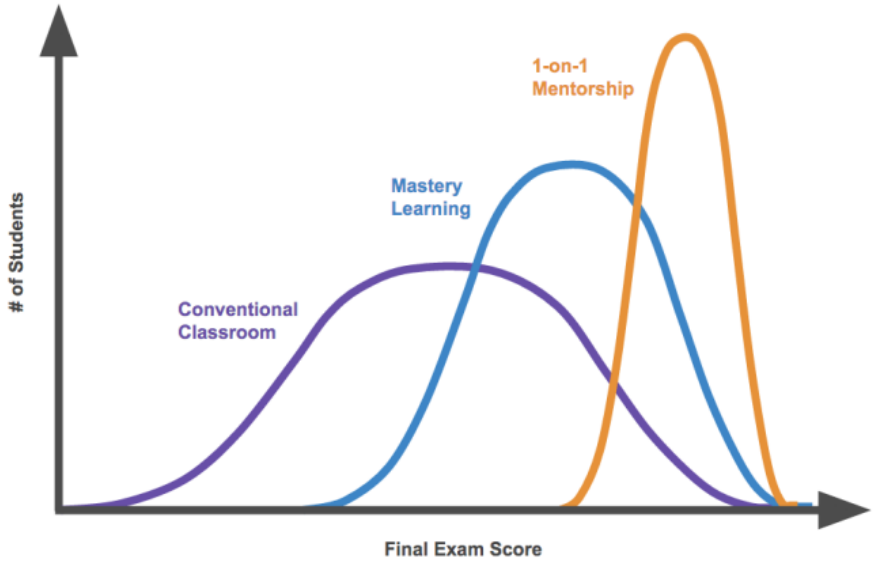

Fig. 1. Bloom's 2-sigma problem: achievement distribution of students.[3]

The above is based on a proprietary platform which has not been transparent on how the underlying Large Language Model (LLM) was trained and where there are several issues including:

- Misinformation - responses may not be truthful.
- Copyright - text may be generated which is similar to existing copyrighted content.
- Bias - bias in training data may mean that responses replicate existing biases and prejudices.

The combination of the impressive capabilities of LLMs, exemplified by ChatGPT, alongside these issues has stirred alarm as well as interest within the global education community [2-5]. On the assumption that their abilities are likely to continue to increase, it is therefore vital to understand how they can be used responsibly by educators

[1] https://www.edweek.org/technology/what-do-teachers-think-of-chatgpt-you-might-be-surprised/2023/03

[2] https://thetab.com/uk/2023/03/21/there-were-more-than-1-million-visits-to-chatgpt-website-at-universities-in-last-exam-season-299853

[3] https://techcrunch.com/2016/01/09/how-startups-are-solving-a-decades-old-problem-in-education

and learners and to ensure they enable more equality, diversity, inclusion, and access, not less.

The remainder of this paper is structured as follows. First, we present the objectives of the SAGE-RAI project, and we discuss the equality, diversity and inclusion issues that will be addressed by the project. We then analyse the educational use cases that fall within the scope of the project, specifically regarding the interactions of GenAI with teachers and students, as well as the benefits and drawbacks of these interactions. Finally, the paper is concluded.

2 Objectives

The SAGE-RAI (Smart Assessment and Guided Education with Responsible AI) project[4] seeks to address the pressing challenges in contemporary education by harnessing the potential of GenAI to provide tailored support, increase engagement, and ultimately lead to improved student outcomes. Through its objectives, the project endeavours to usher in a new era of personalised, efficient, and ethical education, shaping the future of learning for the better. More specifically, the objectives of the SAGE-RAI project closely align with its mission and vision as follows:

- **AI Digital Assistant Integration:** Building upon our ongoing work on developing and evaluating AI digital assistants for teachers and students [6, 7], we will develop plugins that facilitate the seamless integration of GenAI into existing online learning resources.
- **AI Digital Assistant Demonstrator:** We will deploy these plugins in two distinct settings:
    - Within the learning environment of the Open Data Institute (ODI), which provides a wide range of both free and paid courses to learners across the globe.[5]
    - Within OpenLearn, the Open University's highly visible and influential free learning resource.[6]
- **Educational Benefits Evaluation:** We will conduct a formal evaluation to assess the educational benefits of GenAI integration. The aim is to ensure that student learning is not only enhanced but that student satisfaction remains high throughout the process.
- **Business Case and Models:** We will develop a comprehensive business case and structural and process models. These will illustrate why and how educational providers can effectively leverage GenAI to enhance student outcomes without incurring undue resource implications.
- **Ethical Guidance:** Drawing upon the classification framework that underpins the EU AI Act, as well as the project's own evaluation study, we will develop ethical guidance on how to responsibly apply GenAI in educational settings.

By identifying issues specific to the application of GenAI in offering feedback to students, this project will proactively address potential challenges and uncertainties,

[4] https://sage-rai.kmi.open.ac.uk
[5] https://learning.theodi.org
[6] https://www.open.edu/openlearn/

ensuring that AI-driven support enhances the educational experience while mitigating negative outcomes. Rigorous fact-checking mechanisms, regular audits and ethical reviews will assess the AI's hallucinations, fairness, and repetitiveness, and address any detected biases.

While AI can offer prompt responses, we recognise the essential role of human tutors to address complex emotional or sensitive concerns, ensuring students receive well-rounded support. Human feedback is also important to mitigate any AI-based lack of contextual understanding or ability to recognise creative or unconventional work, so we will incorporate the facility to fallback to human tutors if feedback is deemed inadequate. We emphasise the importance of using AI-generated feedback as a supplement to, rather than a replacement for, independent learning and critical analysis.

3 Equality, Diversity & Inclusion

Addressing Bloom's 2-sigma problem with responsibly developed GenAI for lifelong learning is the overall goal of the SAGE-RAI project. Additionally, the mitigation of equality, diversity and inclusion (EDI) issues is itself one of the main topics being investigated within SAGE-RAI.

There are many sources of inequality and bias in GenAI. Existing societal inequality and biases mean privileged groups are overrepresented in training data, and even an unbiased dataset may only contain limited perspectives from smaller/intersectional groups, e.g., LGBTQ+ people or those with particular disabilities. Anti-bias strategies in, e.g., ChatGPT, are opaque, limited to very few people, and may be subject to unconscious bias. GenAI systems may create inaccessible resources or interactions and may not have the capability to tailor outputs appropriately.

The reliance on digital platforms for lifelong learning can exacerbate the digital divide, as individuals who lack access to technology or the necessary digital literacy skills are often excluded. This exclusion is particularly noticeable among disadvantaged communities, including those in rural areas, low-income households, and older adults who may not be as adept with modern technology. As a result, the benefits of lifelong learning are disproportionately enjoyed by those who already have the resources and skills to navigate digital environments, leaving behind those who could potentially benefit the most.

Financial barriers also play a significant role in limiting access to lifelong learning opportunities. The high cost of accredited courses can be prohibitive, particularly for individuals from lower socioeconomic backgrounds. This financial hurdle ensures that only those who can afford to pay for these courses can enhance their skills and qualifications, perpetuating existing socioeconomic disparities. Consequently, lifelong learning, which should ideally bridge gaps and provide equal opportunities for all, instead becomes another arena where inequality is reinforced.

Recognition of prior learning often shows bias towards traditional educational paths, marginalising individuals with non-traditional or informal educational experiences. People from marginalised communities, who may not have had the opportunity to pursue conventional education, find their skills and knowledge devalued in formal recognition processes. This bias undermines the inclusive spirit of lifelong learning by failing

to acknowledge the diverse ways in which individuals acquire valuable skills and knowledge.

Lifelong learning courses frequently lack inclusive pedagogical approaches that cater to diverse learning needs and abilities. Many lifelong learning programmes are designed with a one-size-fits-all mentality, failing to accommodate the unique requirements of learners with disabilities, or those who may need additional support. Without inclusive pedagogy, lifelong learning cannot truly be accessible to all, and many learners are left without the resources they need to succeed.

Cultural relevance is another significant EDI issue in lifelong learning. Educational content that fails to reflect the cultural backgrounds and experiences of diverse learners can alienate and disengage those from different cultural contexts. When the learning material is not relatable or relevant, it becomes difficult for learners to connect with and benefit from it. Ensuring that lifelong learning programmes incorporate diverse cultural perspectives is essential for making education meaningful and effective for all participants.

4 Educational Use Cases

The following sections present the educational use cases that will be explored within the SAGE-RAI project. Additionally, the various benefits and drawbacks associated with these use cases are discussed. The SAGE-RAI educational use cases define the range of interactions of GenAI with teachers and students. Based on these use cases, bespoke plugins will be developed for the integration of GenAI into online learning resources, while addressing the educational requirements of teachers and students.

4.1 Tutor-Controlled Interactions

In tutor-controlled interactions, course creators have the authority to determine the types of tasks and activities that the AI can perform. Use cases for tutor-controlled interactions include:

- **Providing Feedback:** AI is used to provide feedback to students on activities where multiple potential solutions exist, such as creating a data schema.
- **Creating Quizzes and Activities:** AI generates quizzes and activities based on the content within individual course modules, in order to enhance the learning experience.
- **Learning Needs Analysis:** AI offers diagnostic facilities to help students identify their strengths and weaknesses within a course, towards enabling personalised learning pathways.

As an example of a use case for tutor-controlled interactions, Fig. 2 shows a screenshot from a prototype GenAI tool that we are currently developing. This prototype is intended to help tutors create quizzes and activities for their students on a particular topic. In the example shown in Fig. 2, the GenAI tool has generated a set of flashcards for students to test their knowledge on the topic of language and culture.

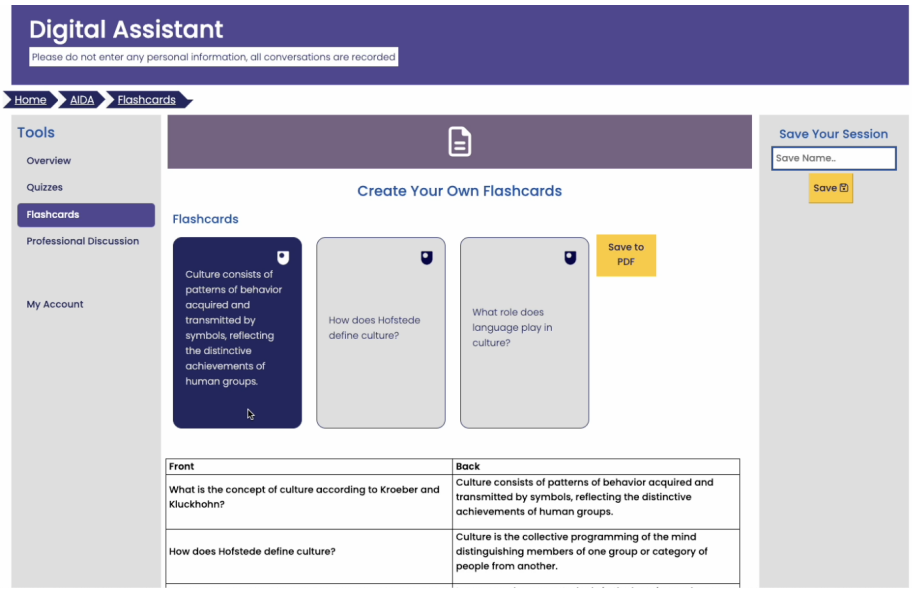

Fig. 2. Screenshot of a prototype GenAI tool for creating flashcards on a particular topic.

The above interactions encompass several benefits both for teachers and students. Firstly, curriculum alignment is ensured, as course creators dictate how GenAI engages with students and align activities with learning objectives. Secondly, scalability is enhanced as GenAI can provide consistent, immediate feedback, lightening the workload of human tutors and facilitating timely guidance. Thirdly, human-AI collaboration empowers human tutors to oversee AI integration, facilitating assessment of AI effectiveness and quality improvement. Notably, AI augments rather than replaces human tutors, allowing tutors to focus on facilitating discussions and providing insights beyond AI capabilities. Finally, personalisation of the learning experience can be achieved, as the AI tailors feedback and activities to individual student needs.

On the other hand, the tutor-controlled interactions present certain drawbacks that include limited creativity, as the AI's responses are bound by the instructions of course creators. This can hinder the AI's capacity for direct conversational engagement and adaptability to individual student needs and potentially impede learning. Additionally, the initial development effort required is a challenge, as crafting and refining AI-driven interactions demand substantial time and effort from course creators during the initial stages.

4.2 Student-Led Interactions

In student-led interactions, students have the opportunity to directly interact with the AI, which is trained on course content. This facilitates freeform discussions and questions. Use cases for student-led interactions include:

- **Content Connection:** Students can ask the AI to connect course content to their immediate needs, helping them discover the most relevant material.

- **Adaptive Learning:** Students' interactions with the AI may deliver customised learning pathways, resources, and activities to address the unique needs of each student. As the student progresses, the AI could continuously adapt the personalised learning pathway. If the student demonstrates mastery in certain areas, the AI may suggest more advanced topics or specialised subjects based on the student's progress.
- **Clarification:** The AI can assist in clarifying unclear concepts or content through responses to student questions. The AI can respond with detailed explanations, examples, and supplementary resources, catering to the student's pace and level of understanding.
- **Navigation Support:** The AI can help students navigate course materials and provide guidance on where to find specific information. Additionally, the AI can recommend supplementary materials, such as research papers [8], therefore broadening the student's understanding beyond the core curriculum.
- **Practice and Feedback:** The AI can generate and review multiple-choice questions (fixed questions in a machine-readable structure) for higher levels of learning (application/creation/reflection). The AI can also generate practice problems and assignments that align with the student's current learning objectives. After the student completes these tasks, the AI can provide automated feedback, pointing out errors and offering suggestions for improvement.
- **Self-Regulated Learning:** Beyond the mentioned use cases, the project could potentially explore additional ways in which student-led interactions can support learning, for example how AI can foster self-regulated learning [9] by empowering learners to become proactive and independent [10, 11].

Table 1 shows a set of pedagogically driven AI roles for student-led interactions [12, 13]. These roles include a Possibility Engine for broadening perspectives, a Socratic Opponent for engaging into argumentation, a Collaboration Coach for facilitating collaborative learning, a Personal Tutor for providing immediate feedback on student progress, a Co-designer for engaging into a collaborative design task, an Exploratorium for exploring, visualising, and interpreting a database or design space, a Study Buddy for helping the student reflect on learning material, as well as a Dynamic Assessor for providing the student with a profile of their current knowledge. Additionally, Table 1 includes pedagogically driven AI roles for either student-led or tutor-controlled interactions, specifically a Guide on the side for navigating physical and conceptual spaces, and a Motivator for extending learning via games and challenges. All these diverse roles leverage AI's potential to foster creativity, critical thinking, and collaborative skills among learners for a comprehensive and enriching learning experience.

As an example of a use case for student-led interactions based on these pedagogically driven AI roles, Fig. 3 shows a screenshot from a prototype GenAI tool that we are currently developing. This tool is intended to help students reflect on what they have learned on a particular topic, using the AI as a Socratic Opponent. Students can engage with the AI as an opponent in an argument, during which they reflect on the responses provided by the AI and challenge the AI to clarify or defend its position.

Table 1. Pedagogically driven AI roles for student-led interactions [12, 13].

Role	Description	Example of implementation
Possibility Engine	AI generates alternative ways of expressing an idea.	Students enter prompts into the AI and submit each prompt multiple times to examine alternative responses.
Socratic Opponent	AI acts as an opponent to develop and argument.	Students enter prompts into the AI following the structure of a conversation or debate.
Collaboration Coach	AI helps student groups to research and solve problems together.	Working in groups, students use the AI to find out information to complete tasks and assignments.
Personal Tutor	AI tutors each student and gives immediate feedback on progress.	The AI offers personalised feedback to students based on information provided by students or teachers (e.g., test scores).
Co-designer	AI assists throughout the design process.	Students ask the AI for ideas about designing or updating a website, or focus on specific goals (e.g., how to make the website more accessible).
Exploratorium	AI provides tools to play with, explore and interpret data.	Students use the AI to explore different ways to visualise and explain a large database, such as census data.
Study Buddy	AI helps the student reflect on learning material.	Students explain their current level of understanding to the AI and ask for ways to help them study the material. The AI can also be used to help students prepare for other tasks (e.g., job interviews).
Dynamic Assessor	AI provides the student with a profile of their current knowledge.	Students interact with the AI in a tutorial-type dialogue and then ask the AI to produce a summary of their current state of knowledge that can be shared with their teacher.
Guide on the side	AI acts as a guide to navigate physical and conceptual spaces.	Students or teachers use the AI to generate content for classes/courses (e.g., discussion questions) and advice on how to support students in learning specific concepts.
Motivator	AI offers games and challenges to extend learning.	Students or teachers ask the AI for ideas about how to extend students' learning after providing a summary of the current level of knowledge (e.g., quizzes, exercises).

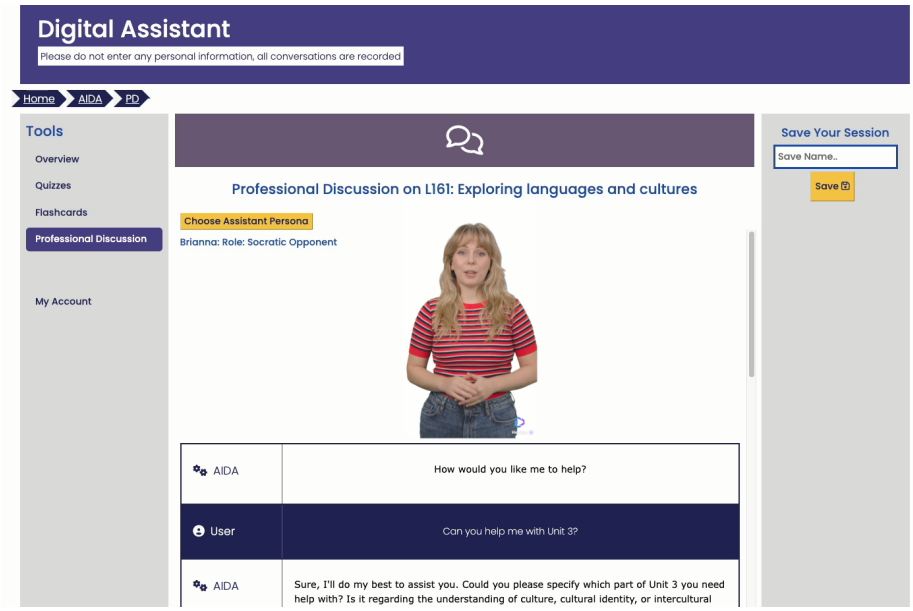

Fig. 3. Screenshot of a prototype GenAI tool acting as a Socratic Opponent for facilitating student reflection.

The benefits of the student-led interactions include the facilitation of personalised learning, as students can seek assistance and ask questions tailored to their individual needs and interests. Secondly, accessibility is improved as students can access educational support at any time, accommodating diverse learning paces, languages, and other accessibility needs. Thirdly, exploration and curiosity are encouraged as students can delve into topics beyond the curriculum, asking open-ended questions to deepen their understanding. Additionally, engagement is enhanced through conversational interaction with GenAI, boosting student motivation and enjoyment of the learning process [14]. Finally, self-regulation is promoted by allowing students to interact with the AI freely, therefore supporting metacognitive development and fostering motivational and behavioural involvement in learning [10, 11].

The drawbacks of student-led interactions include several concerns. Firstly, quality control poses a challenge as GenAI responses may vary in accuracy and reliability, potentially leading to misinformation and misleading students. Secondly, misuse or irrelevant queries may occur, with students posing inappropriate or irrelevant questions, thus disrupting the educational process. Thirdly, irrelevant responses from the AI could steer students away from intended learning outcomes. Additionally, privacy and security concerns arise due to the potential for students to share personal information during open-ended interactions. Lastly, dependence and reduced critical thinking can be a result from over-reliance on the AI, potentially diminishing the problem-solving skills of students as they depend on the AI for answers rather than engaging in independent critical thinking.

Allowing students unfiltered access to an AI model with the freedom to input any prompt raises several important issues, particularly in the context of potential offensive content. To mitigate these issues, several strategies can be adopted. Firstly, content moderation entails employing systems and personnel to review and filter user-generated content for offensive material. Secondly, user guidelines should be established, clearly outlining acceptable behaviour and consequences for rule violations. Thirdly, education and awareness initiatives are essential to inform students about the ethical use of AI and promote responsible behaviour. Fourthly, user reporting mechanisms can be implemented to allow swift action against offensive content. Lastly, AI safeguards involve ongoing efforts to enhance the AI model's ability to identify and reject prompts leading to inappropriate outputs, therefore ensuring alignment with educational content.

5 Conclusion

The SAGE-RAI project is dedicated to exploring the potential of responsible AI for transforming education. The project seeks to address the challenges posed by accommodating large cohorts of students by examining how responsible AI can enhance tutoring, offer tailored, personalised learning experiences, and generate valuable student feedback. The goal is to create an educational platform that supports assessment and student guidance while responsibly applying GenAI to address critical issues such as misinformation, copyright, and bias. A range of educational use cases will be implemented within SAGE-RAI to facilitate the interactions of teachers and students with GenAI. Through this endeavour, we aim to contribute to the responsible integration of GenAI towards achieving cost-effective and scalable personalised education.

Acknowledgement

This work has received funding from the Engineering and Physical Sciences Research Council (EPSRC), RAI UK Award Agreement Ref: EP/Y009800/1 (SAGE-RAI).

References

1. Bloom, B.S.: The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational researcher 13, 4-16 (1984).
2. Dai, W., Lin, J., Jin, F., Li, T., Tsai, Y.-S., Gasevic, D., Chen, G.: Can Large Language Models Provide Feedback to Students? A Case Study on ChatGPT. In: Proceedings of the 23rd IEEE International Conference on Advanced Learning Technologies (ICALT 2023). IEEE Computer Society, (2023).
3. Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J.M., Milligan, S., Selwyn, N., Gašević, D.: Assessment in the age of artificial intelligence. Computers and Education: Artificial Intelligence 3, 100075 (2022).
4. Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E.: ChatGPT for good? On opportunities and

challenges of large language models for education. Learning and Individual Differences 103, 102274 (2023).
5. Bozkurt, A.: Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. Asian Journal of Distance Education 18, (2023).
6. Rienties, B., Domingue, J., Duttaroy, S., Herodotou, C., Tessarolo, F., Whitelock, D.: What distance learning students want from an AI Digital Assistant. Distance Education (In press) (2024).
7. Mikroyannidis, A., Sharma, N., Ekuban, A., Domingue, J.: Using Generative AI and ChatGPT for improving the production of distance learning materials. In: 24th IEEE International Conference on Advanced Learning Technologies (ICALT). IEEE Computer Society, (2024).
8. Pride, D., Cancellieri, M., Knoth, P.: CORE-GPT: Combining open access research and large language models for credible, trustworthy question answering. In: International Conference on Theory and Practice of Digital Libraries, pp. 146-159. Springer, (2023).
9. Nussbaumer, A., Dahn, I., Kroop, S., Mikroyannidis, A., Albert, D.: Supporting Self-Regulated Learning. In: Kroop, S., Mikroyannidis, A., Wolpers, M. (eds.) Responsive Open Learning Environments, pp. 17-48. Springer Open Access (2015).
10. Jin, S.-H., Im, K., Yoo, M., Roll, I., Seo, K.: Supporting students' self-regulated learning in online learning using artificial intelligence applications. International Journal of Educational Technology in Higher Education 20, 37 (2023).
11. Chiu, T.K.: A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: a case of ChatGPT. Educational technology research and development 1-16 (2024).
12. Sharples, M.: Towards social generative AI for education: theory, practices and ethics. Learning: Research and Practice 9, 159-167 (2023).
13. Sabzalieva, E., Valentini, A.: ChatGPT and Artificial Intelligence in higher education: Quick start guide. UNESCO (2023).
14. Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., Farrow, R., Bond, M., Nerantzi, C., Honeychurch, S.: Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. Asian Journal of Distance Education 18, (2023).