

<https://doi.org/10.1038/s41539-025-00301-w>

# On opportunities and challenges of large multimodal foundation models in education



**Stefan Küchemann**<sup>1</sup> ✉, **Karina E. Avila**<sup>1</sup>, **Yavuz Dinc**<sup>1</sup>, **Chiara Hortmann**<sup>1</sup>, **Natalia Revenga**<sup>1</sup>, **Verena Ruf**<sup>1</sup>, **Niklas Stausberg**<sup>1</sup>, **Steffen Steinert**<sup>1</sup>, **Frank Fischer**<sup>2</sup>, **Martin Fischer**<sup>3</sup>, **Enkelejda Kasneci**<sup>4</sup>, **Gjergji Kasneci**<sup>5</sup>, **Thomas Kuhr**<sup>6</sup>, **Gitta Kutyniok**<sup>7,8,9</sup>, **Sarah Malone**<sup>10</sup>, **Michael Sailer**<sup>11</sup>, **Albrecht Schmidt**<sup>12</sup>, **Matthias Stadler**<sup>3</sup>, **Jochen Weller**<sup>6</sup> & **Jochen Kuhn**<sup>1</sup>

Recently, the option to use large language models as a middleware connecting various AI tools and other large language models led to the development of so-called large multimodal foundation models, which have the power to process spoken text, music, images and videos. In this overview, we explain a new set of opportunities and challenges that arise from the integration of large multimodal foundation models in education.

The advent of large language models (LLMs) like ChatGPT-3.5, developed by OpenAI, marked a pivotal moment in the potential of artificial intelligence (AI) in education. This manuscript builds upon prior overviews of LLMs in education, which outlined the transformative potential and inherent challenges of integrating LLMs into educational environments<sup>1–4</sup>. With the introduction of ChatGPT-3.5, it appeared as though we had reached the pinnacle of LLMs. This was a momentous advancement in the application of AI. However, the landscape of AI has proven to be ever-evolving with increasingly short innovation cycles. The recent emergence of large multimodal foundation models (LMFMs), such as ChatGPT-4-Turbo by OpenAI or Gemini by Google, has reshaped our previous perceptions, bringing to the table a fresh set of opportunities and challenges for education. This overview aims to delve deeper into the opportunities and challenges presented by these advanced LMFMs, such as ChatGPT-4-Turbo and Gemini, shedding light on the future of AI in education.

Historically, the integration of AI in education has been a gradual but impactful process. From the early adoption of systems that adapt the sequence of tasks and learning content to learners' understanding with early intelligent tutoring systems<sup>5</sup> to more learner-centered systems that include natural language processing<sup>6</sup>, the educational landscape has been reshaped<sup>7</sup>. The development of the Generative Pretrained Transformer (GPT) and Bidirectional Encoder Representations from Transformers (BERT) marked a significant leap forward in the processing of natural language texts and – by

implication – the processing of learners' and teachers' inputs and consequent responses. These models, trained on extensive text data, showcased an unprecedented ability to generate human-like text, answer complex questions, classify teachers' written reflections and facilitate very high learner-centered interactive learning experiences<sup>8,9</sup>. The release of ChatGPT-3.5 especially sparked numerous empirical studies demonstrating the wide variety of opportunities for education<sup>3</sup>. For instance, it can be used to support teachers in generating course material<sup>10,11</sup>, to detect students' errors during experimentation<sup>12</sup>, to enhance peer feedback processes<sup>13</sup>, to augment empirical data in education<sup>14</sup> and to provide various types of formative feedback to enhance self-regulated learning<sup>15</sup>. At the same time, there are also numerous challenges and concerns, such as the risk of over-reliance by students on the output of an LLM without critically reflecting on it, the difficulty for teachers to identify texts that were created by an LLM rather than by students. There are concerns about the reliability and accuracy of the output of the LLM, potential inherent biases, and the protection of students' and teachers' personal data<sup>1</sup>.

A few months after the release of ChatGPT, it became available via an application programming interface (API). This allowed for its integration into custom-built applications with specific preprompting that allow the assignment of tasks and roles to the LLM<sup>15</sup>. As an important step, Shen and colleagues demonstrated that large language models can be effectively used as a middleware to communicate between different AI models by using

<sup>1</sup>Chair of Physics Education, Faculty of Physics, Ludwig-Maximilians-Universität München, Munich, Germany. <sup>2</sup>Chair of Education and Educational Psychology, Ludwig-Maximilians-Universität München, Munich, Germany. <sup>3</sup>Institute of Medical Education, University Hospital, Ludwig-Maximilians-Universität München, Munich, Germany. <sup>4</sup>Chair for Human-Centered Technologies for Learning, Technical University of Munich (TUM), Munich, Germany. <sup>5</sup>Chair for Responsible Data Science, Technical University of Munich (TUM), Munich, Germany. <sup>6</sup>Faculty of Physics, Ludwig-Maximilians-Universität München, Munich, Germany. <sup>7</sup>Chair of Mathematical Foundations of Artificial Intelligence, Ludwig-Maximilians-Universität München, Munich, Germany. <sup>8</sup>Department of Physics and Technology, The Arctic University of Norway, Tromsø, Norway. <sup>9</sup>DLR - German Aerospace Center, Oberpfaffenhofen, Germany. <sup>10</sup>Department of Education, Campus A 4.2., Saarland University, Saarbrücken, Germany. <sup>11</sup>Chair of Learning Analytics and Educational Data Mining, University of Augsburg, Augsburg, Germany. <sup>12</sup>Chair of Human-Centered Ubiquitous Media, Ludwig-Maximilians-Universität München, Munich, Germany. ✉ e-mail: [s.kuechemann@lmu.de](mailto:s.kuechemann@lmu.de)

language as the universal interface of every AI model<sup>16</sup>. In this way, the vast amount of developed AI tools can be effectively integrated into advanced LLMs, provided the LLM is fine-tuned to the specifications of the AI tools to generate correct inputs. This effective communication between a LLM and different AI tools led to the development of LMFMs, which not only allow an input and output via written text, but also via spoken text, images or videos<sup>17</sup>. Among these advanced LMFMs are ChatGPT-4-Turbo and Gemini. In comparison to the first version of ChatGPT-3.5, these models represent a significant advancement in this field due to its multimodality, boosted accuracy, faster response times, and more nuanced understanding of context and subtleties in language.

The potential opportunities of LMFMs in education are manifold. Their advanced capabilities can revolutionize personalized learning, enabling tailor-made educational experiences that adapt to individual student needs and learning preferences. The enhanced understanding and response mechanisms can aid in creating more engaging and interactive multimodal content, thus potentially increasing student motivation and participation in learning activities. Moreover, their ability to process and generate vast amounts of information in real time can serve as a powerful tool for research on learning and teaching processes.

However, with these opportunities come additional significant challenges. The reliance on AI-supported education raises questions about the accuracy and reliability of the information provided by these systems. The potential for inherent biases in the AI algorithms also remains a critical issue with LMFMs, and this issue may be intensified due to the complexities arising from processing multiple input modalities. Furthermore, the integration of such advanced technologies in educational settings necessitates a reevaluation of teaching methods and curricula. It also requires the development of new competencies among educators and students alike because these tools now allow for much more interactive learning and formative feedback, thus allowing a shift from a focus on declarative knowledge to facilitating complex skills. This is a crucial step towards the democratization of education, ensuring that learning opportunities are accessible and equitable, making it more likely now for illiterate people to use AI without even being able to read and write. Yet, the issue of the digital divide and accessibility remains a concern, as not all educational institutions may have the resources to implement and maintain such advanced technology effectively<sup>1</sup>.

The following sections of this manuscript will explore how LMFMs can enhance various aspects of education, such as personalized learning, student engagement, and educational content and application creation, and they will critically examine the potential drawbacks and limitations of LMFMs in education, including issues related to accuracy, bias, and the need for human oversight. Finally, the 'Mitigation Strategies' section will propose strategies to overcome these challenges, ensuring that the integration of LMFMs in educational settings is done responsibly, ethically, and effectively. This work aims to provide a comprehensive overview of the landscape of LMFMs in education, with a specific focus on the latest advancements represented by LMFMs, thereby contributing to the ongoing discussion about the role of AI in shaping the future of education.

## Opportunities

### Opportunities for learners

Personalized learning assistants can have multiple benefits for learners. They have the potential to enhance students' productivity and foster motivation<sup>18</sup>. ChatGPT, for instance, has been favorably received by students<sup>19,20</sup>. Furthermore, the advanced context understanding of LMFMs could potentially alleviate concerns regarding the accuracy of LLM outputs, due to their enhanced precision. Additionally, the integration of LLMs in an API allows efficient preprompting including the provision of an example solution. In this way, LLMs are aware of the correct solution and can guide the learner towards it without generating hallucinations<sup>15</sup>. In the context of LLMs, "hallucination" refers to the generation of information that appears plausible but is factually incorrect. LMFMs offer a range of options for learners, as recent research shows. Fauzi and colleagues discuss the role of

GPT in stimulating contextual learning strategies, thereby enhancing metacognitive abilities<sup>18</sup>. This aspect is important for the development of self-regulated learning skills. In addition, the opportunity to build personal learning assistants using LMFMs promotes generative learning<sup>21</sup>, where learners actively participate in the design of their learning tools, which may lead to higher engagement, better understanding, and retention. Furthermore, educators can now create customized LFMF applications with pre-defined prompts, design content based on their needs, the subject and teaching style by providing relevant domain-specific content to the fine-tuned LMFMs which can expedite learning through such personalized learning assistants. In this way, the specially created multimodal learning assistants can be individualized and offered on a large scale in the classroom or school setting. Learners can therefore benefit from using teacher-recommended Chatbots to boost their learning process.

Particularly, LMFMs have the capacity to facilitate a range of learning mechanisms and provide support to students across a diverse array of activities, as outlined in the following:

- **Multimedia-effect:** The capabilities of LMFMs, especially their abilities to process and generate images suggest a more comprehensive learning experience. It is already known from multimedia research that text in combination with images leads to better understanding than text alone<sup>22</sup>. This is particularly beneficial for complex or abstract concepts, where visual representations can provide additional clarity. In this line, LMFMs may enhance formative feedback by providing visual representation along with written feedback to better illustrate learners' difficulties<sup>23</sup>.
- **Barrier-free interaction:** Furthermore, LMFMs can play a crucial role in supporting students with disabilities by allowing a barrier-free interaction via speech-to-text, image-to-text, text-to-image, and text-to-speech functionalities. These features make educational content more accessible, especially for students with visual or auditory impairments or with difficulties in writing or reading. Similarly, students can also write or speak in their mother tongue if they are unfamiliar with the local language, and the text is translated into the local language. These features can also be beneficial in language learning, where the model can act as a conversational partner, offering real-time interaction and feedback<sup>24</sup>. It could help learners develop their language skills in a more natural and immersive environment, simulating real-life conversations. Similarly, LMFMs may enrich history lessons by enabling role-playing<sup>25</sup> and acting as conversational partners such as historically important characters like Napoleon, Abraham Lincoln, or others. These features enable a more inclusive learning environment where all students have equal opportunities to engage with and benefit from the educational materials. Additionally, the adaptability of GPTs in customizing learning experiences can help address individual learning needs in a diverse student population with varying abilities and learning preferences<sup>26</sup>.
- **Learning support during the generation of images:** Another significant aspect is the integration of visual transformer models such as DALL-E<sup>27</sup>, which enables the creation of visual outputs based on text inputs. LMFMs have also demonstrated the ability to understand the content of images, thereby offering the opportunity for students to receive learning support during generation through real-time scaffolding and feedback on their drawings. This may prove useful in providing multi-faceted feedback and in subjects where visual representation is key. For instance, in arts and creative subjects, it can be used to teach design and reflection, allowing students to explore different visual representations and interpretations of their ideas<sup>28</sup>. In several other fields of education, the generation of visual representations, such as graphs or schemas, is a common exercise<sup>29</sup>. Here, LMFMs may provide real-time feedback to students while generating graphs.
- **Learning support during the generation of handwritten text, mathematical solutions, and physics equations:** LMFMs have demonstrated the ability to even process images or plots that contain handwritten text, experimental results, function graphs or

mathematical equations to provide feedback to students in real-time<sup>30</sup>. In this way, they may not only support students in solving maths problems but also provide feedback during the solution of physics problems<sup>31</sup>.

- **Representational competence training:** Several authors point out that the integration of visual representations in the learning material can enhance learning<sup>32,33</sup>. To effectively learn from visual representations, Edelsbrunner and colleagues could show that students need to possess representational competence that allows them to extract information from visual representations and interpret them<sup>34</sup>. However, Rau points out that students often face a representational dilemma, which means that they do not possess the necessary representational competence during learning, so they need to acquire it during the learning of domain-specific content<sup>33</sup>. State-of-the-art LMFMs allow the explanation of images, translation of data to other representations and can also explain how to extract information from representations to students. In this way, they may enable an effective representational competence training. In this regard, Polverini and Gregorcic showed that ChatGPT-4-Turbo is able to solve tasks with line graphs in physics with a similar overall score as high school students<sup>35</sup>.
- **Animation of phenomena:** Animations can help learners to understand a number of phenomena if they follow certain principles<sup>36</sup>. With effective preprompting, these animation principles could be integrated into learning environments and provide learners with customized animations for their needs.
- **Targeted support during experimentation:** Performing experiments and collecting experimental data can be tedious for learners, often with limited learning gains<sup>37</sup>. Here, LMFMs perform tasks that are not part of the learning goal, for instance to generate artificial measurement data and visualize it at the same time to offer learners a tangible intuition of factual relationships and to test physical laws. Similar to other digital media, they are able to take over certain learning-relevant cognitive activities during experimentation to effectively reduce extraneous cognitive load and free up cognitive resources to focus on achieving the learning goals<sup>38</sup>. Moreover, LMFMs can analyze images of experimental setups and provide personalized support as students are setting up the experiment.
- **Targeted support during code generation:** The advent of APIs has facilitated the development of tools like CodeHelp, which leverage the power of LLMs as discussed by Liffiton et al.<sup>39</sup>. The capabilities of LMFMs further enhance this process. LMFMs enable the visualization of class and method dependencies or flow diagrams in a program. Hence, LMFMs can help create more accessible and engaging learning environments, especially in complex areas such as programming languages. These functionalities are not only for secondary school students, but the multimodality also allows more advanced students in higher education to access learning content with greater ease due to the multimodal nature of these models.
- **Creation of customized LFMF-based learning tools:** In addition, students are now able to create their own customized learning tools based on LMFMs just by verbally describing the tasks of the tool without the need for programming. Such activities, like creating personal learning assistants using LMFMs, promote generative learning, where learners actively participate in the design of their learning tools and digital learning environments, which may lead to a higher engagement, better understanding, and retention.
- **Interpretation of behavioral and physiological data:** As mentioned above, LMFMs understand a variety of inputs including physiological and behavioral data. Therefore, in comparison to conventional LLMs that can provide scaffolding support and feedback based on students' inputs, LMFMs are able to analyze, visualize, and interpret students' behavioral and physiological data and provide actionable feedback and guidance to learners. For instance, they can visualize these data in a dashboard for students to support self-reflection and identify the most efficient approaches for learning.

- **Support during collaborative learning:** With advanced context understanding, LMFMs can facilitate collaborative learning through peer feedback models. Bauer and Greisel et al.<sup>13</sup> suggested a model which uses natural language processing to support peer feedback in digital learning environments. This model describes learners' activities and textual products, and introduces a scheme to foster the peer feedback process. LMFMs can thus enhance understanding, promote active engagement, and lead to a more enriching learning experience, enhancing metacognitive abilities and self-regulated learning during collaborative learning.
- **Enhancing explainable user interfaces:** Explainable AI in education is employed to make system decisions understandable and verifiable, highlighting, for instance, the involvement and significance of specific features leading to actionable insights<sup>40</sup>. Explainable User Interfaces (XUIs) aim to bring this level of transparency directly to users, presenting system decisions in a clear and interactive manner to enhance users' agency and understanding. While XUIs hold substantial promise, particularly for applications like explainable learning analytics, the education domain is yet to fully adopt and optimize these interfaces. LMFMs provide potential for enhancing XUIs to better accommodate a wide spectrum of teachers and learners. Traditionally designed with adult users in mind (e.g.,<sup>41</sup>), XUIs must evolve to meet the diverse cognitive and developmental needs of users. The integration of LMFMs could potentially enable more adaptable and varied forms of information representation. Exploring the synergy between LMFMs and XUIs represent a promising direction for making AI in education more inclusive and effective.
- **Educational support in developing countries:** In areas in the world where expert teachers and high-quality education are scarce, applications based on LMFMs may contribute to closing the educational gap, enhancing the overall educational level and providing a diverse multimodal learning experience.

In sum, *LMFMs could make complex concepts more accessible, adapt to different learning preferences, and improve the approach learners use to tackle a problem or task. Combining personalization, contextual support, creative engagement and multimodal content, this multi-faceted approach illustrates the potential of LMFMs to improve the educational landscape.*

### Opportunities for teachers and educators

The use and integration of LMFMs in the classroom presents various new opportunities for educators. Similar to students, teachers may benefit from the seamless interaction with an LFMF and its user-friendliness, which enhances accessibility and ease of use. Combined with the option to create customized applications based on LMFMs that include multi-modal input and output, educators from diverse fields are now able to tailor and employ generative AI tools according to their specific instructional needs. Consequently, this advancement may offer benefits to teachers inside and outside of the classroom.

**For lesson planning and classroom activities.** LMFMs present multiple opportunities. With the ability to create customized tools based on LMFMs and visualize content, educators could have a tool at their disposal to conceptualize and structure their teaching strategies more effectively. This capacity for lesson planning integrates with learning task design, where up-to-date database and multi-modal task design capabilities offer a nuanced approach to educational content creation. The LMFMs' up-to-date database and access to the internet are beneficial in subjects like social sciences and politics, providing educators with current and historical context. Simultaneously, the ability to design multimodal tasks complements these contextual advantages. By employing an LFMF's capability to create learning activities that, for instance, include visual elements or symbolic representations, educators may make educational content more accessible and stimulating for learners<sup>42,43</sup>.

**Exploration of new teaching methods.** In addition, the technology may empower educators who are eager to explore new teaching strategies within traditional classroom settings. The introduction of executable code allows teachers to act as developers, who are likely to create bespoke learning experiences that cater to individual student needs, potentially making education more personalized and effective<sup>44,45</sup>. One example is learning environments, such as flipped classroom models and blended learning scenarios, which more strongly incorporate elements of self-regulated learning compared to traditional classroom settings<sup>46,47</sup>. Here, it is plausible that LMFMs will provide resources and support that complement students' independent learning phases, subsequently enriching the in-class teaching experience<sup>48</sup>. Such environments may be particularly relevant to the development of competencies for higher-order thinking skills. These often demand contextualized problem sets, open-ended problem formulation and discussion<sup>49</sup>, requiring teachers to use highly differentiated instructions. In this way, the use of LMFMs to design tasks that stimulate critical thinking and provide individualized feedback could be a valuable asset in reducing teacher workload (cf.<sup>50</sup>).

**For extracurricular work and professional development.** LMFMs may serve as a versatile tool for educators. In extracurricular contexts, they may act as personalized assistants, streamlining administrative tasks such as scheduling, email management, and document preparation. By automating these tasks with personalized GPTs, educators may reallocate their time towards more critical aspects of teaching and curriculum development. Furthermore, it is likely that LMFMs aid in curriculum design by suggesting tailored resources and visualizations, activities, and assessment methods, aligned with specific learning objectives and student needs. Moreover, LMFMs may provide feedback on teaching strategies based on current educational trends.

**Creation of simulations of common classroom practices.** Chernikova and colleagues have demonstrated in a meta-analysis that simulations of professional situations, such as interactions in a classroom for teachers, can be a beneficial tool during teacher training as they expose teachers to authentic problems in their fields<sup>51</sup>. Similarly, LLMs have demonstrated that they convincingly act as students with specific difficulties in physics<sup>14</sup>. Therefore, LMFMs have the potential to authentically simulate multimodal interactions with students with specific difficulties. Additionally, they may systematically support the teachers in simulation environments by automatically selecting and modifying representations of practice, called representational scaffolds<sup>52</sup>.

**Support of multimodal assessment creation and grading.** Prior research demonstrated that LLMs can effectively support teachers during the creation of assessment tasks in physics of comparable quality to textbooks<sup>10</sup> or during the assessment of experimentation errors in chemistry education<sup>12</sup>. By extension, LMFMs may also support teachers in creating multimodal assessment tasks and provide an initial correction of students' solutions in the form of handwritten text, handwritten mathematical solutions, or students' drawings.

**Creation of teacher dashboards.** As mentioned above, LMFMs can analyze student performance data and behavioral data of teachers. LMFMs may create data visualizations and create dashboards for teachers to provide an overview of class performance and give effective feedback to teachers<sup>53,54</sup>. Based on this analysis, teachers may adapt their teaching strategy and identify the most efficient approaches used in the classroom.

#### **Opportunities for (educational) researchers**

The transition to LMFMs represents a breakthrough not only for educational research, but also across diverse academic disciplines.

**Advanced data analysis.** LMFMs introduce advanced features for data analysis, such as data cleaning and data processing<sup>55,56</sup>, and corresponding

visualizations. These features significantly streamline the data analysis process, allowing for more experiments or studies to be conducted in the same timeframe. Here, the capabilities go beyond basic analysis, handling complex statistical tasks and processing large data sets. Furthermore, the conversational nature of ChatGPT-4 transforms the data analysis tool into more than just a mechanism for cleaning, evaluating, and presenting data. It serves as a guide, engaging in a dialogue with the researcher, elucidating the coding behind each analysis, and rationalizing every decision made. This interaction enables the tool to adapt to the researcher's criteria and preferences, fostering a collaborative environment. These enhancements pave the way for more nuanced, accurate, and efficient research.

**Literature reviews.** LMFMs often represent a remarkable advance in contextual understanding and language generation. They outperform conventional LLMs, such as ChatGPT-3.5, in terms of accuracy and depth<sup>57</sup>, which allows for more efficient processing of large volumes of textual data. Additionally, the combination of direct access to internet resources and enhanced contextual understanding facilitates comprehensive literature reviews and in-depth thematic analysis with unprecedented efficiency. Moreover, by elevating the accuracy of content generation, advanced LMFMs significantly contribute to the formulation of research texts, aiding in the creation of draft papers and proposals.

**Advanced graphical data interpretation.** As mentioned above, LMFMs exhibit the potential to decipher and interpret the context and components within figures<sup>35,57,58</sup>, assisting researchers in comprehending the underlying narrative and conveying the key message encapsulated in visual representations. Additionally, they have the capacity to provide guidance on optimizing figure design for clarity and comprehension by audiences. This multifaceted assistance signifies a promising trajectory in which AI can augment researchers' abilities to interpret and communicate complex information across diverse modalities.

**Creation of multimodal assessment and learning material.** LMFMs do not only benefit teachers by aiding in the creation of multimodal assessment and learning materials, but they also may prove useful to researchers. These materials can be utilized in empirical research studies. Additionally, they can assist researchers in categorizing and grading handwritten text or math solutions, thereby facilitating the analysis of studies.

**Advanced research opportunities.** Prior to the release of LMFMs, the development of intelligent tutoring systems was cost- and time-intensive, and required skilled personnel. Therefore, the development of tutoring systems was not common, and research in this area was not widely possible. Now, the opportunity to develop customized LFMF apps ('GPTs'<sup>59</sup> in case of ChatGPT-4-Turbo) empowers researchers to create and test customized intelligent tutors, which can also be readily shared with a global audience. This grants teachers and students swift access to state-of-the-art research tools, fostering an environment where cutting-edge advancements seamlessly integrate into educational practices. Additionally, its robust speech-to-text and text-to-speech processing capabilities enable studies where learners can interact with computers in a manner that more closely mirrors the dynamics and interactivity of a traditional classroom setting.

**Support for diverse, multilingual and interdisciplinary research teams.** LMFMs, designed for multilingual tasks, are likely to provide valuable assistance to diverse, interdisciplinary, and multilingual research teams by promoting effective communication within diverse teams. Special terms or methodological approaches in certain disciplines may be rephrased and explained in simple terms to researchers from other disciplines. Similarly, LMFMs can provide a real-time translation from spoken text to another preferred language enabling collaboration between research teams that do not speak a common language. This

ability to bridge language and professional barriers may significantly enhance collaboration and productivity in research teams.

Within the realm of educational research specifically, the new possibilities of LMFMs are manifold. Their proficiency in understanding multiple languages and representations offers a promising avenue for bridging the communication gap between researchers and students from different parts of the world.

### Opportunities for developers of educational applications

**Development of customized intelligent learning applications.** Since several years, intelligent tutoring systems (ITS) have been developed and effectively integrated into learning environments to provide enriched personalized learning experiences, for instance, by offering tailored feedback and guidance<sup>60,61</sup>. However, their development and integration may be cost-intensive and time-consuming, leading to a more localized and discipline-specific implementation and hindering a widespread use. The emergence of LMFMs could potentially revolutionize this field with their ability to comprehend and generate human-like text with a broad range of contexts and across multiple modalities, and LMFMs could significantly augment the capabilities of ITS, including cognitive tutors. For instance, their advanced contextual awareness could make cognitive tutors more effective in assisting with ill-defined problems.

Moreover, the possibility to easily generate pre-prompted chats and applications with LMFMs and to share these simple programs has opened up remarkable opportunities for both teachers and students, turning them into developers of specialized, user-friendly software applications. For example, recently OpenAI introduced a feature for their ChatGPT Plus users, termed GPTs<sup>59</sup>. These custom GPTs enable educators to rapidly create tools tailored for specific lessons, while students can develop applications to assist in their studies, for instance, for exam preparation. The ability to upload files with exercises and solutions ensures the precision of the GPT's responses, a critical feature in educational environments<sup>15</sup>. The utilization of LMFMs in software development can potentially boost students' programming self-efficacy and motivation, especially in specific scenarios. An example of this enhancement was observed during the incorporation of LMFMs into weekly programming exercises within programming education, as highlighted in a study by Yilmaz (2023)<sup>62</sup>. As a result, some scholars have recognized ChatGPT as an effective tool for imparting computational thinking skills<sup>62</sup>. Prior to the advent of LLMs, the development of software through low-code or no-code platforms was primarily embraced by individuals with some interest in computer science because of the significant time required to learn such a tool<sup>63</sup>. However, tools like LMFMs have a significantly reduced learning curve<sup>64</sup>. Therefore, this approach has the potential to expand its reach to individuals with less inclination towards programming.

**Intelligent human-centered design process.** Furthermore, LLM-powered platforms enable the creation of functional software without the necessity of explicitly specifying every intricate detail or meticulously planning the execution process. In this way, uncertain details and features of software that are often specified in a human-centered design process can be initially defined and later adapted by the LMFMs. This reduction in cognitive effort paves the way for a widespread utilization across various domains, including recreational applications. Despite the reduced cognitive load compared to traditional programming, it is likely that the use of LMFMs to create software applications could enhance computational thinking skills. For instance, a user might initiate the creation of an application based on an LMFMs without specifying every detail initially. During the initial use of the tool, they may encounter unexpected behaviors and subsequently refine and specify the previously overlooked details. It is possible that this iterative process of use could facilitate the development of a more precise ability to articulate intentions.

**Collaborative development of complex software.** Another significant opportunity lies in the use of advanced LLMs for complex software development. With the expanded context length of models like GPT-4

Turbo, LLMs can now consider and modify extensive amounts of source code, potentially encompassing 128k tokens which are more than 300 pages of text<sup>65</sup>. By sustaining a more extended context, users can develop intricate software, thereby augmenting the functionality and depth of the content they generate<sup>64</sup>. This feature may also be a boon for educational researchers who sometimes require sophisticated software platforms for their studies. Even with minimal programming knowledge, they can collaborate with LMFMs to express their needs (e.g., through drawings or speech) and develop complex platforms and user interfaces, a task that typically would require extensive funding and expertise. A notable example of collaborative programming involves the concept of Human-Bot Collaborative Architecting, with a specific emphasis on the integration of ChatGPT into human decision-making processes to streamline aspects of Architecture-Centric Software Engineering (ACSE). An initial case study within this domain explored the potential synergies between ChatGPT and architects' decision-making, aiming to automate various facets of ACSE, as documented by Ahmad<sup>66</sup>. The authors also uncovered ChatGPT's capacity to identify and address ethical, governance, and socio-technical considerations within this context<sup>66</sup>.

**Code optimization.** Furthermore, there are solutions that try to leverage the potential of simpler coding through LLMs. One example is CoPrompt, a system that assists programmers with prompt engineering in a collaborative context<sup>67</sup>. CoPrompt aids in prompt comprehension through detailed explanations, helping users understand and track the development of their programming tasks over time<sup>67</sup>.

Thus, *the integration of LMFMs in educational settings has huge potential to support teaching by offering personalized, up-to-date, and multimodal educational tools, enhancing lesson planning, enabling innovative teaching strategies, streamlining both extracurricular and administrative tasks, supporting professional development, and facilitating more effective assessments and feedback.*

### Challenges

#### Challenges for learners

Students working with LMFMs face several challenges, each presenting unique complexities and considerations:

- **Output verification:** A critical challenge is verifying the accuracy of information generated by LLMs<sup>68</sup>. These models respond based on data patterns rather than genuine understanding, leading to potential inaccuracies or misleading information. LMFMs extend the capabilities of traditional LLMs by generating speech, images, such as diagrams, and videos, which introduces additional complexity in the output verification. For example, learning with LMFMs may require a learner to critically reflect on a visual representation that they may have difficulty understanding and extracting information from<sup>33</sup>. In this way, there is a risk that students might overly rely on the outcomes provided by LMFMs.
- **Verification of sources:** Another challenge arises from the difficulty of verifying the sources from which information is retrieved<sup>69</sup>. LLMs amalgamate vast amounts of data, making it hard to trace back the origin of specific pieces of information, which is crucial for reliability. For LMFMs, this issue is even more severe as the connected AI Models typically lack information of the training data set and how the output is generated by each model.
- **Maintaining an overview of AI models:** The EU's General Data Protection Regulation gives users the right to know the basis of an algorithmic decision that significantly affects them<sup>70</sup>. The regulation thus enables the possibility of increased transparency in algorithmic decisions, among other things to avoid a "black-box society"<sup>71</sup> and discrimination by algorithms<sup>72</sup>. However, this transparency is hindered by a lack of technical understanding of AI models<sup>73</sup> and the challenge to maintain an overview of which AI models are involved in creating outputs of LMFMs. This awareness is key to understanding the limitations and capabilities of the responses they receive.

- **Potential segregation in society due to access costs:** Access to LMFMs can be expensive, potentially leading to segregation in society. This financial barrier might create exclusive learning groups that have access to LMFMs tools, which may influence educational outcomes and further widening the learning gap between different socioeconomic groups.
- **Ethical issues and inherent biases:** LLMs and other AI tools may have a number of biases that could arise from various factors, such as unbalanced training datasets<sup>74-77</sup>. LMFMs raise significant ethical concerns, particularly regarding fair use. The intricacy and interconnectedness of these AI models can make it challenging to identify and rectify inherent biases. These models have the potential to mirror and amplify biases present in their training data, leading to the propagation of stereotypes or prejudiced viewpoints. This is a matter of considerable concern, particularly in educational settings.
- **Dependency on LMFMs:** Anderson and colleagues argue that artificial intelligence may erode human abilities, thus creating a dependence<sup>78</sup>. The ease of using LMFMs might lead to underdevelopment of certain skills. Students may become dependent on these tools for tasks they should be able to accomplish independently, thus reducing their effort and engagement in the learning process. Thus, there is a need for teachers and educational leaders to decide which human skills should remain part of the curriculum, and which can be dropped or reduced. It is necessary to identify which new skills are required and should be incorporated into the curriculum.
- **Overuse in learning situations:** There is a risk of LMFMs being overused in scenarios where students should be developing their own learning skills. This overuse may impede the development of critical thinking, research skills, and independent problem-solving abilities.
- **Severity of personal data leaks:** Concerns about personal data leaks are paramount<sup>79</sup>. Students may not be fully aware of which services process and store their data, what personal information is retained, who has access to it, and how it is protected.
- **Unwilling and unknowing involvement in plagiarism:** Although using an LFMF does not constitute plagiarism in itself, there is a risk that the content produced may inadvertently include plagiarized material. This raises concerns about academic integrity and the need for students to be vigilant about the originality of the content they submit.

Each of these challenges necessitates a careful and informed approach to using LMFMs in educational settings. Students and educators alike must be aware of these issues to effectively integrate these powerful tools into the learning process while mitigating potential risks. If learners are aware of the challenges and risks, the overarching distribution of applications based on LMFMs in the social environment of learners may lead to hesitation and fear of using them, as it may be difficult to discriminate between tools that are trustworthy and incorporate ethical standards and those that do not.

### Challenges for teachers and educators

Several key challenges for the educational community arise with the introduction of LMFMs in education:

- **Digital divide:** The implementation of LMFMs into teaching curricula carries the risk of widening the knowledge gap among learners due to an accessibility gap<sup>80-82</sup>. Not all educational institutions possess the financial and intellectual resources necessary to adopt and integrate such technologies. This could result in disparities in educational quality and opportunities, perpetuating existing inequalities among learners.
- **Perpetuating knowledge base:** The proliferation of use of LMFMs poses a risk of perpetuating existing knowledge bases and social biases<sup>83,84</sup>. LMFMs now draw on web- and training data released past the introduction of ChatGPT, which may also contain GPT-generated content, thus limiting exposure to genuinely new and diverse information. This potentially hinders the evolution of educational content and curricular materials.
- **Erosion of human relations:** The emphasis on AI-driven verbal communication tools may inadvertently contribute to the erosion of nonverbal communication competencies and social skills among learners and educators<sup>2,85,86</sup>. Overreliance on technology for communication could detract from the human interactions that foster social understanding and emotional skills. These skills include the ability to recognize emotions in others within the learning environment and to assist others in managing their emotions.
- **Reliability of information:** The integration of LMFMs makes it more challenging to assess the reliability of information, particularly for non-specialists and educators<sup>87,88</sup>. The ease of accessibility may obscure the need for critical evaluation, potentially leading to the dissemination of inaccurate or biased information in educational settings<sup>83,84,89</sup>. For instance, a teacher working outside their specialization may struggle to evaluate the credibility of the information provided by the sources they use to prepare for teaching. This could affect their ability to teach the subject effectively and confidently. Likewise, it may become more difficult for teachers to fairly assess the quality of student assignments.
- **Need for AI literacy:** The ease of access to LMFMs may create a distraction from the essential requirement to not only learn with but also about AI. There is a risk that the convenience of AI usage might overshadow the importance of developing a comprehensive understanding of its capabilities, limitations, and ethical considerations<sup>90,91</sup>.
- **Curriculum design and educational policies:** Educational institutions face the challenge of keeping pace with the rapid changes in LLM capabilities, necessitating constant adaptation of curricula and educational policies to ensure relevance and effectiveness<sup>92,93</sup>.

### Challenges for (educational) researchers

As AI tools such as LMFMs become more effective, accurate and multimodal, offering exciting opportunities for educational research and research in general, their widespread adoption must be approached with caution. It is essential to balance the use of AI with the preservation of rigorous, ethical, and insightful academic practices.

There is a risk of developing an overreliance on LMFMs, creating a dependency that could be problematic, such as:

- **Vulnerability in research methodologies:** Researchers could find themselves ill-equipped to adapt or continue their work without the help of AI, leading to vulnerability in their research methodologies.
- **Originality and authorship:** If researchers begin to rely predominantly on AI to write papers or proposals, the line between the researcher's original ideas and AI-generated content may blur, raising questions about originality and authorship.
- **Transparency:** Even with the conversational nature of the data analysis tool, understanding how LMFMs arrive at certain conclusions or suggestions may be challenging. This lack of transparency can be a significant hurdle in research settings. Researchers may even be unable to explain the exact method of data analysis being used.
- **Over-trust:** It is likely that there is a risk of over-trusting the results and deductions drawn from LMFMs data analysis tools, which could lead to incorrect conclusions.
- **Bias in AI:** Despite advanced methods and benchmarks to mitigate bias in LMFMs and improve their alignment to social norms and goals<sup>94-98</sup>, bias in AI continues to be an issue<sup>99,100</sup>. AI algorithms could inadvertently perpetuate biases present in their training data, leading to biased research outcomes that further reinforce existing prejudices. A better understanding of the needs of LMFMs users in educational contexts, as well as human-centered evaluations of the underlying models, are both a necessity and a challenge, ensuring that such tools are both ethically aligned and transparent in their operations<sup>101,102</sup>. This is particularly concerning if researchers heavily rely on and trust AI, such as LMFMs.
- **Adaptability and limitations:** While LMFMs may offer improved adaptability due to, for example, their ability to handle extended context lengths<sup>103,104</sup> or their improved vision capabilities<sup>30</sup>, there might still

be limitations in fully meeting the diverse and specific needs of different research tasks and contexts. For example, during a systematic review, it is necessary to derive a balanced perspective on topics that may have been diversely displayed and discussed in multiple sources. Comparing and drawing conclusions from data presented in different documents, modalities and contexts may still be too complex for state-of-the-art LMFMs.

- **Ethical considerations:** Ensuring the ethical use of LMFMs-driven data analysis tools, especially when dealing with sensitive educational data, is crucial. Researchers must navigate data privacy laws and ethical considerations. For example, under the EU AI Act, the use of generative AI in education would be subject to rigorous ethical and regulatory standards, focusing on risk assessment, transparency, data protection, and fairness<sup>105</sup>.
- **Resource requirements:** Utilizing advanced LMFMs often requires substantial computational and economical resources, which might not be readily available to all educational researchers.

Overall, the art of academic writing, critical thinking, and synthesis of information are at the core of research work. AI should be viewed by researchers as a supportive tool that enhances their work and empowers researchers to leverage its strengths, rather than a substitute for the indispensable skills they contribute.

### Challenges for developers in education

Despite the evident opportunities, the use of LMFMs in educational software development also presents several challenges, especially for novices, who may need more guidance and support to use them effectively. Some of the challenges include:

- **Complexity and accuracy:** ChatGPT-4.0, which can be used for Android app development, can handle simple tasks well, but may struggle with longer descriptions or more complicated functions. Developers often need to ask additional follow-up questions<sup>106</sup>. When given long descriptions, ChatGPT-4.0 sometimes generates incorrect or incomplete code. Although step-by-step guidance is beneficial for simpler tasks, it is not always sufficient for complete app development<sup>106</sup>. Novice developers frequently require extra guidance to resolve programming errors and learn additional steps that LMFMs may not provide, such as code placement and package imports<sup>106</sup>. An effective strategy for application development with ChatGPT involves breaking down the app into smaller functions and handling them separately, which necessitates more follow-up questions as complexity rises<sup>106</sup>. For novices, it is recommended to first learn basic coding through tutorials before using ChatGPT and to utilize specific commands for more accurate responses<sup>106</sup>. This approach underscores the potential for further research into ChatGPT's application in more complex areas of app development<sup>106</sup>.
- **Adaptations for educational tool developers:** The rapid pace of development in LMFMs may present challenges for developers of educational tools. Keeping up with advancements, integrating them into tools, and understanding their potential roles in education are key issues. Technical challenges and resource-intensity add to the complexity.
- **Fine-tuning in education:** A significant challenge in leveraging LLMs for educational purposes lies in the process of fine-tuning. Fine-tuning these models to suit specific educational needs and contexts requires a substantial amount of relevant and high-quality data. However, obtaining such data in the educational field is often a hurdle, primarily due to privacy concerns, limited availability of diverse educational content, and the proprietary nature of many educational resources. Hence, scarcity of data can hinder the ability to tailor these AI models effectively for educational applications. It also raises questions about the generalizability and adaptability of the models to diverse educational settings and learner needs. Overcoming these challenges requires not only innovative approaches to data collection

and sharing in the educational domain, but also the employment of methods for data augmentation and generation<sup>107-109</sup>.

- **Consistency and ethics:** Studies on Architecture-Centric Software Engineering (ACSE) using ChatGPT highlight several concerns<sup>66</sup>. The varied responses from ChatGPT can affect the consistency in architecting, and the content generated may raise legal issues regarding intellectual property as well as ethical concerns<sup>66</sup>. Biases in ChatGPT's training data might also influence the quality of outputs<sup>66</sup>. To address these challenges, further empirical studies are necessary to validate the productivity and efficacy of ChatGPT in various contexts and teams<sup>66</sup>.
- **Access:** An additional challenge is the potential restriction of access to LMFMs, which may be limited to those who can afford them, especially if no viable alternatives from large companies exist. This can be seen in the restricted access to GPTs generated by OpenAI's users. The reliance on proprietary platforms raises concerns about inclusivity and equitable access to advanced educational technology, potentially creating a divide between those who can and cannot afford these tools<sup>91</sup>. Ensuring broad and equitable access to these transformative technologies is crucial to avoid exacerbating existing inequalities in education.

### Mitigation strategies

To effectively integrate LMFMs in education, it is critical for the educational community to adopt multifaceted risk mitigation strategies. Emphasizing AI as a supplementary tool in teaching is paramount, ensuring it complements rather than replaces human teaching.

Educators and researchers alike must navigate these advancements responsibly, upholding ethical standards and maintaining rigorous scholarly practices. Researchers need to engage in ongoing education regarding LMFMs advancements, rigorously uphold ethical standards, and transparently document AI utilization, fostering responsible integration within scholarly pursuits. Educators need comprehensive training in AI literacy, ethical usage, data privacy, and fact-checking to critically assess AI-generated content. Likewise, engaging parents as part of the educational community in understanding LMFMs' educational impact is essential to create responsible LMFMs-integrated learning environments and to foster their long-term acceptance.

For researchers delving into the integration of LMFMs into academic inquiry, several proactive strategies emerge as pivotal:

- **Systematic evaluation of AI limitations:** Instituting a regime of systematic and field-specific assessments to ascertain the bounds and constraints of the knowledge of LMFMs, which can enable researchers to discern where human expertise is indispensable and where AI augmentation is beneficial.
- **Continuous engagement with LMFMs advancements:** Maintaining an ongoing commitment to understanding LMFMs developments ensures researchers remain well-versed in the evolving landscape of these technologies.
- **Transparent documentation of LMFMs utilization:** Transparently documenting the use of LMFMs tools throughout the research process fosters transparency and integrity in scholarly pursuits.
- **Critical assessment and cross-referencing:** Embracing a critical mindset entails consistently comparing LMFMs-generated outputs with established resources, upholding the integrity of scholarly claims.
- **Diversified LMFMs training data and bias mitigation:** Advocating for diversified training data and employing bias detection and mitigation techniques helps in fostering fairer and more inclusive research outcomes.

Simultaneously, within educational frameworks, strategies to integrate AI tools responsibly among teachers and students are fundamental. In the design of curricula incorporating AI in general, and LMFMs in particular, a focus on AI literacy is essential. Making it a core part of the curriculum is necessary for both teachers and students, providing a clear understanding of AI's and LMFMs' capabilities and limitations. Additionally, curricula should include modules on fact checking and ethical considerations of AI, aiming to

develop students' critical evaluation skills of AI-generated content. Finally, creating assessment strategies that align with learning objectives in LMFMs-enhanced environments is crucial, ensuring that these strategies effectively gauge student understanding and the application of critical thinking in the context of LMFMs.

In educational policies and structures, mitigating the challenges posed by LMFMs involves several key strategies. Designing teacher training systems that adapt to rapid changes in educational technology is crucial. This includes prolonged phases of continual further qualification, enabling educators to stay informed about the evolving AI landscape. Addressing the digital divide is also essential; formulating policies and prioritizing resource allocation are necessary steps to ensure equitable access to AI technologies for all students, thereby creating a level playing field. Monitoring AI developments for biases is important to maintain fairness in educational applications, which requires regular policy reviews and updates. Additionally, promoting research into the pedagogical implementation of LMFMs in educational settings is vital. This helps in developing a knowledge base that informs effective teaching practices with LMFMs, contributing to the creation of evidence-based educational strategies.

### Open-source vs. proprietary LMFMs

Open-source vs. proprietary LMFMs: In the long term, the choice between open-source and proprietary LMFMs will be critical, as open-source LMFMs are more accessible and cost-effective, offering customization, transparency, and community support, which are beneficial for collaborative learning and innovation. Open-source models will allow educators and researchers to tailor models to specific needs and ensure ethical alignment, but can pose challenges in data security. However, this will also require more advanced technical skills around AI literacy, model fine-tuning, and quality assurance. With further technological advancements, the choice between open-source and proprietary LMFMs may not only impact the future of learning but also pave the way for a new era of digitally empowered, inclusive education.

### Conclusion

The emergence of LMFMs like ChatGPT-4-Turbo and Gemini represents a significant shift in the landscape of education, offering unparalleled opportunities for personalized learning, enhanced engagement, and innovative educational methodologies. In our perspective, it is important to understand LMFMs as tools with the potential to transform how we learn, teach, and conduct educational research by enabling more interactive, tailored, and efficient educational environments that teachers and students can personalize without the need for programming skills. However, it is important to understand the new challenges that arise with this technological revolution and not to replace human teachers, developers, and researchers as they are central to understanding, communicating, and mitigating the risks from LMFMs. Therefore, it is also relevant to keep the human in the loop, for instance, as a responsible teacher and/or an increasingly self-directed learner.

For learners, LMFMs present opportunities for enhanced learning experiences through personalization, barrier-free interaction, and support in various learning activities. Yet, there are challenges such as verifying the accuracy of LLM outputs, ethical issues, potential biases in AI models, and the risk of overreliance, which could hinder critical thinking and independent learning.

Educators stand to benefit from LMFMs through innovative teaching strategies, lesson planning, and professional development. However, they also face challenges like the digital divide, erosion of human relations, reliability of information, and the need for AI literacy. These challenges necessitate a reevaluation of teaching methodologies and the development of new competencies.

Educational researchers are presented with advanced data analysis capabilities and tools for creating multimodal assessment and learning material. However, they must exercise caution regarding overreliance on LMFMs, maintain transparency, and address ethical considerations.

Developers in education can now more easily create intelligent learning applications and engage in human-centered design processes. Nevertheless, they face issues related to the complexity and accuracy of LMFMs, consistency in outputs, ethical concerns, and access limitations.

To mitigate these challenges, a multifaceted approach is needed. For the educational community, emphasizing AI as a supplementary tool, providing comprehensive AI literacy training for educators and learners, and engaging parents in understanding the impact of LMFMs are essential. Researchers should systematically evaluate AI limitations, engage continuously with advancements, document LFMF utilization transparently, and employ diversified training data for bias mitigation. In educational policy and curriculum design, focusing on AI literacy, developing critical evaluation skills, and formulating equitable access policies are crucial. Moreover, choosing open-source over proprietary LMFMs can be crucial for a more flexible and community-centered future of education; however, the advantages related to accessibility, independent customization, and ethics require higher levels of technical skills and regulatory expertise.

Ultimately, while LMFMs offer a revolutionary path forward in education, balancing their benefits with the inherent challenges requires careful consideration, ethical awareness, and a commitment to ongoing adaptation and learning from all involved groups, i.e., teachers, students, educational researchers, and developers. The joint efforts of all involved groups will ensure that the integration of LMFMs in educational settings is done responsibly, ethically, and effectively, contributing to a future where technology enhances the educational experience.

Received: 22 January 2024; Accepted: 27 January 2025;

Published online: 26 February 2025

### References

1. Kasneci, E. et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individ. Differ.* **103**, 102274 (2023).
2. Tlili, A. et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learn. Environ.* **10**, 15 (2023).
3. Lo, C. K. What is the impact of ChatGPT on education? A rapid review of the literature. *Educ. Sci.* **13**, 410 (2023).
4. Ray, P. P. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet Things Cyber-Phys. Syst.* **3**, 121–154 (2023).
5. Shute, V. J. SMART: Student modeling approach for responsive tutoring. *User Model. User Adap. Inter.* **5**, 1–44 (1995).
6. Wang, Y., Sun, Y. & Chen, Y. Design and research of intelligent tutor system based on natural language processing. In: *2019 IEEE International Conference on Computer Science and Educational Informatization (CSEI)*, 33–36 (IEEE, 2019).
7. Ouyang, F. & Jiao, P. Artificial intelligence in education: The three paradigms. *Comput. Educ. Artif. Intell.* **2**, 100020 (2021).
8. Wulff, P., Mientus, L., Nowak, A. & Borowski, A. Utilizing a pretrained language model (BERT) to classify preservice physics teachers' written reflections. *Int. J. Artif. Intell. Educ.* **33**, 439–466 (2023).
9. Seßler, K., Xiang, T., Bogenrieder, L. & Kasneci, E. PEER: Empowering Writing with Large Language Models. *Eur. Conf. Technol. Enhanc. Learn. Springer*, **14200**, 755–761 (2023).
10. Küchemann, S. et al. Can ChatGPT support prospective teachers in physics task development? *Phys. Rev. Phys. Educ. Res.* **19**, 020128 (2023).
11. Stadler, M., Horrer, A., & Fischer, M. Crafting Medical MCQs with Generative AI: A How-To Guide on Leveraging ChatGPT. <https://doi.org/10.13140/RG.2.2.33494.22081> (2023).
12. Bewersdorff, A., Seßler, K., Baur, A., Kasneci, E. & Nerdel, C. Assessing student errors in experimentation using artificial intelligence and large language models: A comparative study with human raters. *Comput. Educ. Artif. Intell.* **5**, 100177 (2023).



13. Bauer, E. et al. Using natural language processing to support peer-feedback in the age of artificial intelligence: A cross-disciplinary framework and a research agenda. *Br. J. Educ. Technol.* **54**, 1222–1245 (2023).
14. Kieser, F., Wulff, P., Kuhn, J. & Küchemann, S. Educational data augmentation in physics education research using ChatGPT. *Phys. Rev. Phys. Educ. Res.* **19**, 020150 (2023).
15. Steinert, S., Avila, K. E., Ruzika, S., Kuhn, J. & Küchemann, S. Harnessing large language models to develop research-based learning assistants for formative feedback. *Smart Learn. Environ* **11**, 62 (2024).
16. Shen, Y., Song, K., Tan, X., Li, D., Lu, W., & Zhuang, Y. Hugginggpt: Solving ai tasks with chatgpt and its friends in hugging face. *Adv. Neural. Inf. Process. Sys.* **36**, (2024).
17. Wu, S., Fei, H., Qu, L., Ji, W. & Chua, T.-S. Next-gpt: Any-to-any multimodal llm. Preprint at <https://arxiv.org/abs/2309.05519> (2023).
18. Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A. M. A. & Hatta, H. R. Analysing the role of ChatGPT in improving student productivity in higher education. *J. Educ.* **5**, 14886–14891 (2023).
19. Romero Rodríguez, J. M. et al. *Use of ChatGPT at university as a tool for complex thinking: Students' perceived usefulness* (Universidad de Alicante, 2023).
20. Singh, H., Tayarani-Najaran, M.-H. & Yaqoob, M. Exploring Computer Science Students' Perception of ChatGPT in Higher Education: A Descriptive and Correlation Study. *Educ. Sci.* **13**, 924 (2023).
21. Brod, G. Generative learning: Which strategies for what age? *Educ. Psychol. Rev.* **33**, 1295–1318 (2021).
22. Mayer, R. E. Cognitive theory of multimedia learning. *Cambridge Handb. Multimedia Learn.* **41**, 31–48 (2005).
23. Bertlanga, A. J., Van Rosmalen, P., Boshuizen, H. P. A. & Sloep, P. B. Exploring formative feedback on textual assignments with the help of automatically created visual representations. *J. Comput. Assist. Learn.* **28**, 146–160 (2012).
24. Kohnke, L., Moorhouse, B. L. & Zou, D. ChatGPT for language teaching and learning. *RELC J.* (2023).
25. Shanahan, M., McDonell, K. & Reynolds, L. Role play with large language models. *Nature* **623**, 493–498 (2023).
26. Tavangarian, D. et al. Is e-Learning the Solution for Individual Learning? *Electron. J. E learn.* **2**, 265–272 (2004).
27. Betker, J. et al. Improving image generation with better captions. *Comput. Sci.* <https://cdn.openai.com/papers/dall-e-3.pdf> (2023).
28. Hutson, J. & Robertson, B. Exploring the Educational Potential of AI Generative Art in 3D Design Fundamentals: A Case Study on Prompt Engineering and Creative Workflows. *Glob. J. Hum.-Soc. Sci.* **23**, 485 (2023).
29. Fiorella, L. & Mayer, R. E. Eight ways to promote generative learning. *Educ. Psychol. Rev.* **28**, 717–741 (2016).
30. Google. Gemini: Explaining reasoning in math and physics. <https://www.youtube.com/watch?v=K4pX1VAxaAI> (2023).
31. Kortemeyer, G. Toward AI grading of student problem solutions in introductory physics: A feasibility study. *Phys. Rev. Phys. Educ. Res.* **19**, 020163 (2023).
32. Ainsworth, S. DeFT: A conceptual framework for considering learning with multiple representations. *Learn. Instr.* **16**, 183–198 (2006).
33. Rau, M. A. Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educ. Psychol. Rev.* **29**, 717–761 (2017).
34. Edelsbrunner, P. A. et al. The relation of representational competence and conceptual knowledge in female and male undergraduates. *Int. J. Stem Educ.* **10**, 1–19 (2023).
35. Polverini, G. & Gregorcic, B. Performance of ChatGPT on the test of understanding graphs in kinematics. *Phys. Rev. Phys. Educ. Res.* **20**, 010109 (2024).
36. Mayer, R. E. & Moreno, R. Animation as an aid to multimedia learning. *Educ. Psychol. Rev.* **14**, 87–99 (2002).
37. Finkelstein, N. D. et al. When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Phys. Rev. Spec. Top. Phys. Educ. Res.* **1**, 010103 (2005).
38. Becker, S., Klein, P., Gößling, A. & Kuhn, J. Using mobile devices to enhance inquiry-based learning processes. *Learn. Instr.* **69**, 101350 (2020).
39. Liffiton, M., Sheese, B. E., Savelka, J., & Denny, P. Codehelp: Using large language models with guardrails for scalable support in programming classes. In *Proceedings of the 23rd Koli Calling International Conference on Computing Education Research*, 1–11 (2023).
40. Khosravi, H. et al. Explainable artificial intelligence in education. *Comput. Educ. Artif. Intell.* **3**, 100074 (2022).
41. Meske, C. & Bunde, E. Design principles for user interfaces in ai-based decision support systems: The case of explainable hate speech detection. *Inf. Syst. Front.* **25**, 743–773 (2023).
42. Fadel, C. & Lemke, C. *Multimodal learning through media: What the research says*, 1–24 (Cisco Systems, 2008).
43. Sankey, M., Birch, D. & Gardiner, M. W. Engaging students through multimodal learning environments: The journey continues. In: *Australian Society for Computers in Learning in Tertiary Education Annual Conference*, 852–863 (Australasian Society for Computers in Learning in Tertiary Education, 2010).
44. Dockterman, D. Insights from 200+ years of personalized learning. *npj Sci. Learn.* **3**, 15 (2018).
45. Bloom, B. S. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educ. Res.* **13**, 4–16 (1984).
46. Azevedo, R., Johnson, A., Chauncey, A. & Burkett, C. Self-regulated learning with metatutor: Advancing the science of learning with metacognitive tools. In *New Science of Learning: Cognition, Computers and Collaboration in Education*, 225–247 (Springer, New York, NY, 2010).
47. Greene, J. A. & Azevedo, R. A theoretical review of winne and hadwin's model of self-regulated learning: New perspectives and directions. *Rev. Educ. Res.* **77**, 334–372 (2007).
48. Meyer, B., Haywood, N., Sachdev, D. & Faraday, S. *Independent learning. Learning and Skills Network* (Departement For Children, 2008).
49. Miri, B., David, B.-C. & Uri, Z. Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking. *Res. Sci. Educ.* **37**, 353–369 (2007).
50. Higton, J. et al. *Teacher workload survey 2016* (Department for Education London, 2017).
51. Chernikova, O. et al. Simulation-based learning in higher education: A meta-analysis. *Rev. Educ. Res.* **90**, 499–541 (2020).
52. Fischer, F. et al. Representational scaffolding in digital simulations-learning professional practices in higher education. *Inf. Learn. Sci.* **123**, 645–665 (2022).
53. Valle, N., Antonenko, P., Dawson, K. & Huggins-Manley, A. C. Staying on target: A systematic literature review on learner-facing learning analytics dashboards. *Br. J. Educ. Technol.* **52**, 1724–1748 (2021).
54. Molenaar, I. & Knoop-van Campen, C. Teacher dashboards in practice: Usage and impact. In *Data Driven Approaches in Digital Education: 12th European Conference on Technology Enhanced Learning, EC-TEL 2017*, Tallinn, Estonia, September 12–15, 2017, Proceedings **12**, 125–138 (Springer, 2017).
55. OpenAI. Introducing ChatGPT Enterprise. Available at <https://openai.com/blog/introducing-chatgpt-enterprise> (2023).
56. Open AI. Code Interpreter. Section: Code Interpreter. Available at <https://openai.com/blog/chatgpt-plugins> (2023).

57. OpenAI. GPT-4 Technical Report. *arXiv*, <https://arxiv.org/abs/2303.08774> (2023).
58. OpenAI. GPT-4 (Vision) System Card. Published on September 25, 2023, at <https://cdn.openai.com/papers/gpt-4-system-card.pdf> (2023).
59. OpenAI. Introducing GPTs. Retrieved on November 24, 2023, from <https://openai.com/blog/introducing-gpts> (2023).
60. Guo, L. et al. Evolution and trends in intelligent tutoring systems research: a multidisciplinary and scientometric view. *Asia Pac. Educ. Rev.* **22**, 441–461 (2021).
61. VanLehn, K. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educ. Psychol.* **46**, 197–221 (2011).
62. Yilmaz, R. & Karaoglan Yilmaz, F. G. The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Comput. Educ. Artif. Intell.* **4**, 100147 (2023).
63. Sahay, A., Indamutsa, A., Di Ruscio, D. & Pierantonio, A. Supporting the understanding and comparison of low-code development platforms. In *2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, Portoroz, Slovenia, 171–178 (IEEE, 2020).
64. Spano, L. D., Schmidt, A., Santoro, C. & Stumpf, S. eds. In *End-User Development: 9th International Symposium, IS-EUD 2023, Cagliari, Italy, June 6–8, 2023, Proceedings. Lecture Notes in Computer Science* Vol. 13917 (Springer Nature Switzerland, 2023).
65. OpenAI. New models and developer products announced at DevDay. Available at <https://openai.com/blog/new-models-and-developer-products-announced-at-devday> (2023).
66. Ahmad, A. et al. Towards human-bot collaborative software architecting with chatgpt. In *Proceedings of 27th International Conference on Evaluation and Assessment in Software Engineering*, 279–285 (ACM, 2023).
67. Feng, F. L. et al. CoPrompt: Supporting Prompt Sharing and Referring in Collaborative Natural Language Programming. *arXiv*, <https://arxiv.org/abs/2310.09235> (2023).
68. Kuchnik, M., Smith, V. & Amvrosiadis, G. Validating large language models with relm. In *Proceedings of Machine Learning and Systems*, **5** (MLSYS, 2023).
69. Krupp, L. et al. Unreflected acceptance—investigating the negative consequences of chatgpt-assisted problem solving in physics education. *arXiv*, <https://arxiv.org/abs/2309.03087> (2023).
70. Goodman, B. & Flaxman, S. European union regulations on algorithmic decision-making and a “right to explanation”. *AI Mag.* **38**, 50–57 (2017).
71. Pasquale, F. *The black box society: The secret algorithms that control money and information* (Harvard University Press, 2015).
72. Berendt, B. & Preibusch, S. Exploring discrimination: A user-centric evaluation of discrimination-aware data mining. In *2012 IEEE 12th International Conference on Data Mining Workshops*, 344–351 (IEEE, 2012).
73. Burrell, J. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data Soc.* **3**, 2053951715622512 (2016).
74. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. & Galstyan, A. A survey on bias and fairness in machine learning. *ACM Comput. Surv.* **54**, 1–35 (2021).
75. Abid, A., Farooqi, M. & Zou, J. Persistent anti-muslim bias in large language models. In *Proceedings of 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 298–306 (ACM, 2021).
76. Ferrara, E. Should chatgpt be biased? challenges and risks of bias in large language models. *arXiv*, <https://arxiv.org/abs/2304.03738> (2023).
77. Liang, P. P., Wu, C., Morency, L.-P. & Salakhutdinov, R. Towards understanding and mitigating social biases in language models. In *International Conference on Machine Learning*, 6565–6576 (PMLR, 2021).
78. Anderson, J., Rainie, L. & Luchsinger, A. Artificial intelligence and the future of humans. *Pew Res. Center* **10**, 12 (2018).
79. Sebastian, G. *Privacy and data protection in chatgpt and other ai chatbots: Strategies for securing user information*. SSRN 4454761 (SSRN, 2023).
80. Khowaja, S. A., Khuwaja, P. & Dev, K. Chatgpt needs spade (sustainability, privacy, digital divide, and ethics) evaluation: A review. *arXiv*, <https://arxiv.org/abs/2305.03123> (2023).
81. Pahl, S. *An emerging divide: Who is benefiting from AI* (United Nations Industrial Development Organization, 2023).
82. Bostrom, N. & Yudkowsky, E. The ethics of artificial intelligence. In *Artificial Intelligence Safety and Security*, 57–69 (Chapman and Hall/CRC, 2018).
83. Kooli, C. Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability* **15**, 5614 (2023).
84. Talanquer, V. Interview with the chatbot: How does it reason? *J. Chem. Educ.* **100**, 2821–2824 (2023).
85. Kuhail, M. A., Alturki, N., Alramlawi, S. & Alhejori, K. Interacting with educational chatbots: A systematic review. *Educ. Inf. Technol.* **28**, 973–1018 (2023).
86. Kamalov, F. & Gurrib, I. A new era of artificial intelligence in education: A multifaceted revolution. *arXiv*, <https://arxiv.org/abs/2305.18303> (2023).
87. Li, Y. et al. Can large language models write reflectively. *Comput. Educ. Artif. Intell.* **4**, 100140 (2023).
88. Zeng, Z. et al. Towards automatic boundary detection for human-ai hybrid essay in education. *arXiv*, <https://arxiv.org/abs/2307.12267> (2023).
89. Qadir, J. Engineering education in the era of chatgpt: Promise and pitfalls of generative ai for education. In *2023 IEEE Global Engineering Education Conference (EDUCON)*, 1–9 (IEEE, 2023).
90. Casal-Otero, L. et al. AI literacy in k-12: a systematic literature review. *Int. J. STEM Educ.* **10**, 29 (2023).
91. Trust, T., Whalen, J. & Mouza, C. Chatgpt: Challenges, opportunities, and implications for teacher education. *Contemp. Issues Technol. Teach. Educ.* **23**, 1–23 (2023).
92. Ng, D. T. K., Leung, J. K. L., Chu, S. K. W. & Qiao, M. S. Conceptualizing ai literacy: An exploratory review. *Comput. Educ. Artif. Intell.* **2**, 100041 (2021).
93. Lin, P.-Y. et al. Modeling the structural relationship among primary students' motivation to learn artificial intelligence. *Comput. Educ. Artif. Intell.* **2**, 100006 (2021).
94. Bai, Y. et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv*, <https://arxiv.org/abs/2204.05862> (2022).
95. Lee, H. et al. Rlaif: Scaling reinforcement learning from human feedback with ai feedback. *arXiv*, <https://arxiv.org/abs/2309.00267> (2023).
96. Sun, Z. et al. Principle-driven self-alignment of language models from scratch with minimal human supervision. *arXiv*, <https://arxiv.org/abs/2305.03047> (2023).
97. Lee, H. et al. Kosbi: A dataset for mitigating social bias risks towards safer large language model application. *arXiv*, <https://arxiv.org/abs/2305.17701> (2023).
98. Wang, Y. et al. Aligning large language models with human: A survey. *arXiv*, <https://arxiv.org/abs/2307.12966> (2023).
99. Schwartz, R. et al. Towards a standard for identifying and managing bias in artificial intelligence. *NIST Spec. Publ.* **1270**, 10.6028 (2022).
100. Casper, S. et al. Open problems and fundamental limitations of reinforcement learning from human feedback. *arXiv*, <https://arxiv.org/abs/2307.15217> (2023).

101. Rong, Y. et al. Towards human-centered explainable ai: A survey of user studies for model explanations. *IEEE Trans. Pattern Anal. Mach. Intell.* **46**, 2104–2122 (2023).
102. Leemann, T., Rong, Y., Nguyen, T.-T., Kasneci, E. & Kasneci, G. Caution to the exemplars: On the intriguing effects of example choice on human trust in XAI. In *XAI in Action: Past, Present, and Future Applications* (NIPS, 2023).
103. Chen, Y. et al. Longlora: Efficient fine-tuning of long-context large language models. *arXiv*, <https://arxiv.org/abs/2309.12307> (2023).
104. Peng, B., Quesnelle, J., Fan, H. & Shippole, E. Yarn: Efficient context window extension of large language models. *arXiv*, <https://arxiv.org/abs/2309.00071> (2023).
105. Hacker, P., Engel, A. & Mauer, M. Regulating chatgpt and other large generative ai models. In *Proceedings of 2023 ACM Conf. Fairness, Accountability, and Transparency*, 1112–1123 (ACM, 2023).
106. Al-Khiami, M. I. & Jaeger, M. Leveraging chatgpt in android app development: A case study on supporting novice developers, in creating successful apps, *Comput. Sci. Math.* (2023).
107. Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695 (IEEE, 2022).
108. Borisov, V., Sessler, K., Leemann, T., Pawelczyk, M. & Kasneci, G., Language models are realistic tabular data generators. In: *The Eleventh International Conference on Learning Representations* (ICLR, 2022).
109. Cascante-Bonilla, P. et al. Going beyond nouns with vision & language models using synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 20155–20165, (IEEE, 2023).

### Author contributions

S.K., K.E.A., Y.D., C.H., N.R., V.R., NiS., StS., F.F., M.F., E.K., GiK., T.K., GJK., S.M., MiS., A.S., MaS., J.W., and J.K. contributed to the conception of the

research, the writing of the first version and the revision of the manuscript. All authors have read and approved the manuscript.

### Funding

Open Access funding enabled and organized by Projekt DEAL.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to Stefan Küchemann.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025