ADVANCED INTELLIGENT SYSTEMS
Open Access
www.advintellsyst.com

Large Language Model-Based Chatbots in Higher Education

Defne Yigci, Merve Eryilmaz, Ail K. Yetisen, Savas Tasoglu, and Aydogan Ozcan**

Large language models (LLMs) are artificial intelligence (AI) platforms capable of analyzing and mimicking natural language processing. Leveraging deep learning, LLM capabilities have been advanced significantly, giving rise to generative chatbots such as Generative Pre-trained Transformer (GPT). GPT-1 was initially released by OpenAI in 2018. ChatGPT's release in 2022 marked a global record of speed in technology uptake, attracting more than 100 million users in two months. Consequently, the utility of LLMs in fields including engineering, healthcare, and education has been explored. The potential of LLM-based chatbots in higher education has sparked significant interest and ignited debates. LLMs can offer personalized learning experiences and advance asynchronized learning, potentially revolutionizing higher education, but can also undermine academic integrity. Although concerns regarding AI-generated output accuracy, the spread of misinformation, propagation of biases, and other legal and ethical issues have not been fully addressed yet, several strategies have been implemented to mitigate these limitations. Here, the development of LLMs, properties of LLM-based chatbots, and potential applications of LLM-based chatbots in higher education are discussed. Current challenges and concerns associated with AI-based learning platforms are outlined. The potentials of LLM-based chatbot use in the context of learning experiences in higher education settings are explored.

1. Introduction

The study of artificial intelligence (AI) has grown exponentially over the past few decades, leading to scientific breakthroughs in a myriad of disciplines including genomics, materials science, robotics, climate science, astronomy, and linguistics (**Figure 1**).[1–6] An AI-based application that has gained widespread attention globally is the development of AI-based chatbots such as ChatGPT, a large language model (LLM) developed by OpenAI. Since its development, ChatGPT and other LLMs have been explored in scholarly publishing, healthcare, dentistry, oncological clinical practice, public health, global warming, solving bugs, finance, and education have been explored.[7–15] Despite promising applications of ChatGPT, concerns regarding data privacy and protection, regulatory processes, bias, and misinformation remain.[16–18] Particularly in education, the benefits versus the potential exploitation or misleading use of chatbots has initiated several controversial debates. Furthermore, some studies have focused on exploring the utilization of chatbots in learning, teaching, and administrative tasks to enable a higher degree of

D. Yigci
School of Medicine
Koç University
34450 Istanbul, Turkey

M. Eryilmaz, A. Ozcan
Department of Electrical & Computer Engineering
University of California Los Angeles (UCLA)
Los Angeles, CA 90095, USA
E-mail: ozcan@ucla.edu

A. K. Yetisen
Department of Chemical Engineering
Imperial College London
London SW7 2AZ, UK

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aisy.202400429.

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aisy.202400429

S. Tasoglu
Department of Mechanical Engineering
Koç University
34450 Sariyer Istanbul, Turkey
E-mail: stasoglu@is.mpg.de

S. Tasoglu
Physical Intelligence Department
Max Planck Institute for Intelligent Systems
70569 Stuttgart, Germany

A. Ozcan
California NanoSystems Institute (CNSI)
University of California Los Angeles (UCLA)
Los Angeles, CA, USA

A. Ozcan
Department of Bioengineering
University of California Los Angeles (UCLA)
Los Angeles, CA, USA

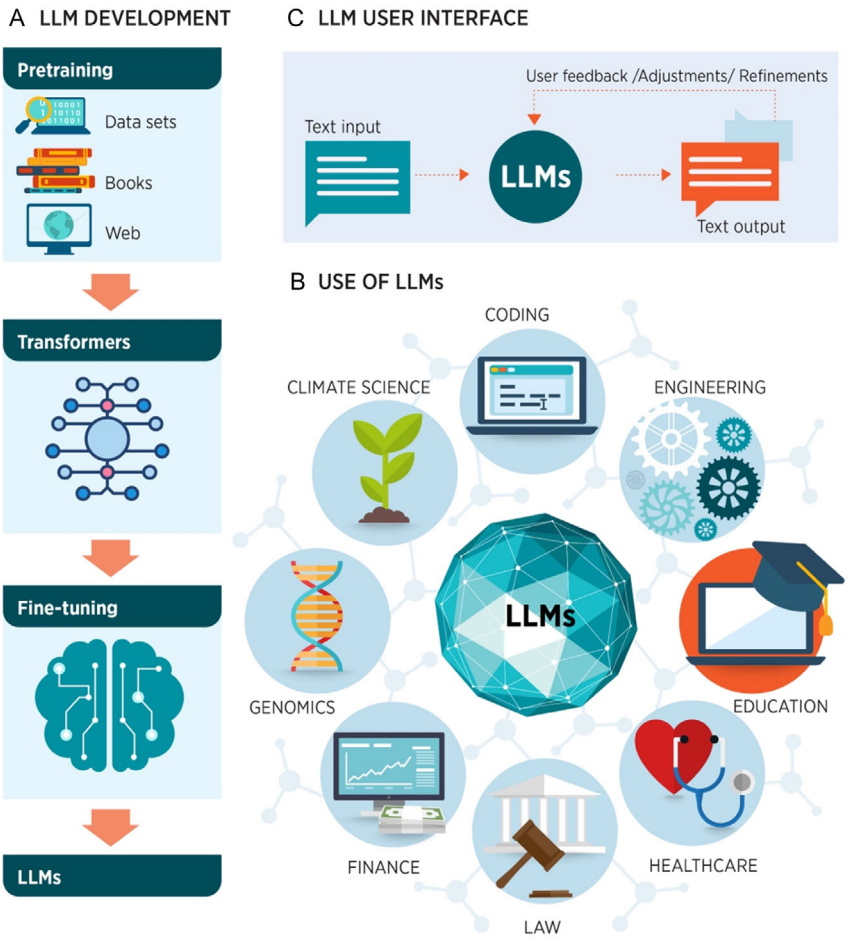


Figure 1. The applications of AI. A) LLM development includes the employment of pre-training using datasets, books, and web sources. Various transformers and fine-tuning methods are then used to facilitate LLM functionality. B) LLM user interface enables users to input text prompts and generates text outputs. User prompts can be adjusted or refined and LLMs respond to user feedback to fine tune outputs. C) The utility of LLMs in science, technology and social sciences. Some elements in this figure were designed using resources from freepik.com.

accessibility, fostering diversity and inclusion, and enhance personalized learning pathways.[19–21] Others have discussed the use of chatbots to aid instructors in course material production or as second language learning tools.[22,23] Concerns regarding academic integrity and plagiarism, data biases, misinformation or disinformation have also been raised and mitigation strategies have been discussed.[24–27] However, the widespread implementation of many of these strategies poses an ongoing challenge. As a result, promising applications regarding the use of ChatGPT in education have not been systematically incorporated into pedagogical practices.

Here, the development of LLMs and the characteristics of ChatGPT are briefly reviewed. Particularly, the potential applications of ChatGPT in higher education are explored. Finally, challenges and concerns regarding the utilization of ChatGPT in higher education are discussed and strategies to address such issues are pointed out.

2. Large Language Models (LLMs)

While the term had not been coined yet, the foundations of natural language processing (NLP) and language generation date back to the 1950s. The Georgetown-IBM demonstrations of the 1950s set early-stage public showcasing of machine-based language translation. Like the Russian-English language translation model, early language generation models of the 1960s also relied heavily on simple rule-based algorithms to detect keywords in input texts in order to generate natural language conversations between humans and machines.[28] However, such rule-based algorithms fell short in facilitating true language understanding. It wasn't until the 1980s and the transition to statistical language processing platforms, such as Hidden Markov Models, that an improved machine understanding of language was achieved.[29] Although contextual understanding and semantics interpretation remained lacking, computational grammar theory was studied

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

extensively during the late 1980s and 1990s, allowing for sophisticated linguistics processing.[30] The early 2000s marked the advent of deep learning tools, leading to a significant leap in the NLP models. Word embedding, neural language models, long short-term memory (LSTMs), and LLMs followed soon after.[31–34]

LLMs represent a class of AI-based models trained on massive amounts of data to analyze language and generate natural language text responses to free-text inputs without receiving input-specific training. LLM architectures typically consist of four major components: i) transformers enabling weighted input representation, ii) learning mechanisms (i.e., deep learning, self-supervised learning), iii) pre-training, and iv) fine-tuning methods.[35–37] The emergence of the initial Generative Pre-trained Transformer (GPT) in 2018 marked an important milestone in generative chatbot development. Sequence-to-sequence training tasks and 117 M parameters were used in the development of GPT.[38] Since then, up to 1750 B parameters have been implemented in generating LLMs, although the number of parameters used does not necessarily correlate with LLM performance at specific subject matters or tasks.[38] A variety of different strategies have also been implemented in LLM architecture, eventually leading to the development of novel generative chatbots (**Table 1**). Similar to GPT, sequence-to-sequence tasks were used to develop Text-to-Text Transfer Transformer (T5), GPT-2, GPT-3, GPT-4, Language Model for Dialogue Applications (LaMDA), and Open Pretrained Transformer (OPT).[39–45] Employing an alternative approach based on bidirectional context analysis, a Bidirectional Encoder Representations with Transformer (BERT) model was developed in 2019, rapidly

Table 1. LLM architectures enable general-purpose language generation.

Model Developer, [Year]	Transformer	Training Dataset	Pre-training Task	Fine-Tuning	Number of Parameters	Ref.
GPT-1 OpenAI, 2018	Sequence-to-sequence tasks	BooksCorpus dataset (1 × 109 words)	Token prediction, language remodeling	Discriminative fine-tuning, task-aware input transformations	117 mil	[38,39]
GPT-2 OpenAI, 2019	Sequence-to-sequence tasks	40 GB dataset: WebText	Token prediction, language remodeling	N/A	1.5 bil	[38,41]
GPT-3 OpenAI, 2020	Sequence-to-sequence tasks	45 TB dataset: Common Crawl, WebText2, Books1, Books2, Wikipedia	Token prediction, language remodeling	N/A	175 bil	[38,176]
GPT-4 OpenAI, 2023	Sequence-to-sequence tasks	N/A[a)]	N/A[a)]	Reinforcement learning from human feedback (RLHF)	N/A[a)]	[38,42]
BERT Google, 2018	Bidirectional context	Toronto Book Corpus and Wikipedia (3.3B tokens)	Transfer learning, MLM, NSP	N/A	0.34 bil	[38]
RoBERTa UW/ Google, 2019	Bidirectional context	BERT, CC News, OpenWebText, Stories (33B tokens)	Dynamic MLM	N/A	356 mil	[47]
ALBERT Google, 2019	Bidirectional context	Toronto Book Corpus and Wikipedia (3.3B tokens)	MLM, NSP	N/A	Up to 60 mil	[46]
ELECTRA Stanford/Google, 2020	Masked token replacement	Toronto Book Corpus and Wikipedia (3.3B tokens)	RTD	N/A	Up to 330 mil	[49]
XLNet CMU/Google, 2019	Permutation based training	BERT, Giga5, ClueWeb 2012-B, Common Crawl (110 GB)	Autoregressive language remodeling	N/A	Up to 360 mil	[48]
T5 Google, 2019	Sequence-to-sequence tasks	Collosal Clean Crawl Corpus, Common Crawl (750 GB)	DAE	N/A	Up to 11 bil	[43]
OPT Facebook, 2022	Sequence-to-sequence tasks	Concatenation of datasets used for RoBERTa, Pile, PushShift.io Reddit (180B tokens)	Token prediction, language remodeling	N/A	Up to 175 bil	[38,44]
LaMDA Google, 2022	Sequence-to-sequence tasks	Public dialogue and public web documents (1.56 T words)	Token prediction, language remodeling	Fine-tuning with annotated data and external knowledge source consulting	137 bil	[38,45]

[a)]remains confidential. MLM: masked language model, NLP: next sentence prediction, RTD: replaced token detection, DAE: Denoising Autoencoder.

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

followed by a Robustly Optimized BERT Pre-training Approach (RoBERTa) and a Lite BERT for Self-Supervised Learning of Language Representations (ALBERT).[46,47] Other training strategies have included the use of permutation-based training for XLNet and masked token replacement training for Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA).[48,49] Moreover, besides different transformers, various pre-training tasks, and training datasets, a number of new parameters and fine-tuning strategies have also been incorporated into LLM architectures, influencing LLM properties (Table 1).[50]

With a growing demand for generative AI, LLM development is likely to expand. The global chatbot market size is expected to increase from 44 B USD (2022) to 75 B USD by 2030.[51] Having reached 180 M users in December 2023, OpenAI's ChatGPT is one of the most widely used conversational chatbots globally, with OpenAI valued at 80 billion USD.[52] This rapid increase in valuation and growing user demand is closely related to ChatGPT's features. Pretrained to detect the next token and fine-tuned using reinforcement learning from human feedback (RLHF), GPT-4 exhibited superiority over other LLMs as demonstrated through standardized test performances.[42] Moreover, despite not being specifically trained for examinations intended for humans, GPT-4 performed in the 90th percentile and 88th percentile for the uniform bar exam and the law school admissions test (LSAT), respectively. Similarly, GPT-4 obtained 5/5 on a wide range of Advanced Placement (AP) examinations, including AP Art History, AP Biology, AP Macroeconomics, AP Microeconomics, and AP US Government placing in the 85th to 100th percentiles. However, despite outperforming other LLMs in language tests, GPT-4 received a 2/5 on AP English Literature and Composition placing in the 8th to 22nd percentile, revealing limitations in achieving sufficient literary understanding and generating coherent and logical textual analyses. Nevertheless, when ChatGPT was tested on a set of 80 multiple-choice United States Medical Licensing Examination (USMLE)-like questions assessing "soft skills" including empathy, communication and interpersonal skills, leadership, and complex ethical issues, it correctly answered 90% of questions without revising any answers, demonstrating a high aptitude for soft clinical skills.[53] The versatility of ChatGPT in higher education has set the ground for performing complex assessments and tasks in pedagogical approaches.

3. Capabilities of ChatGPT in Learning and Teaching

3.1. Next Generation Search Engines in Education and Research

Search engine optimization (SEO) enables users to access the most relevant pieces of information in response to their query. Traditionally, SEO has been achieved using keyword search, analytics monitoring, and user experience (UX) analysis.[54] Generative AI can transform SEO and eventually play an important role in developing next-generation search engines. Although unable to replace traditional search engines yet, ChatGPT has been demonstrated to excel in tasks requiring minimal specialization and provide a superior user experience when compared to search engines like Google.[55] Generative AI appears superior to conventional SEO strategies because it instantly responds to user-generated feedback and remembers previous inputs in the same conversation.[56] In fact, companies like Microsoft have attempted to improve the quality of search results produced using the search engine Bing by integrating ChatGPT.

In addition to efforts aiming to reinvent commonly used search engines, startups have also focused on developing AI-based or LLM-based search engines. Next- generation search engines can significantly expand the capabilities of conventional search engines for daily use as well as in the field of education and scholarly publishing. By allowing researchers to gain access to accurate content-driven sets of information, optimized search engines or next-generation search engines can revolutionize research and educational practices. For example, when ChatGPT and Google Scholar generated outputs on general medical knowledge on specific clinical cases were compared, ChatGPT outperformed Google Scholar.[57] However, ChatGPT outputs were far from perfect and fell short in delivering medical advice, suggesting that while ChatGPT can acquire prompt-specific information, outputs should be evaluated for accuracy. Despite current limitations, next-generation search engines can eventually replace conventional ones. Next- generation search engine development can also be seamlessly integrated into common daily practice as most legal issues and ethical concerns are indistinguishable from practices that are already in use.

Moreover, some of the most important advantages of developing next-generation AI-powered search engines include improvement of search algorithms and personalization. The contextual understanding of LLMs can power search engines to produce more accurate and personalized outputs to user queries. For example, a genAI-powered search tool PubTator 3.0 was shown to retrieve biomedical articles with higher precision compared to Google Scholar and PubMed, demonstrating promising SEO efficiency.[58] Similarly, LLM-augmented search tools such as Scite, Elicit, and Consensus can provide improved semantics search while others like LitMaps or ConnectedPapers can provide visual models to guide literature searches.[59] Despite promising results, several issues have to be addressed before next-generation search engines can be reliably used. Since search engines significantly impact users' access to information, it is imperative to systematically and quantitatively evaluate the ability of AI-powered search engines to generate accurate, reliable, and unbiased outputs. Although large-scale studies have not been conducted yet, it has been demonstrated that AI-powered search engines can exacerbate pre-existing biases of users due to the over-personalization of results.[60] Thus, it is essential to develop bias-detection tools to avoid amplifying biases. Diversification of training data and implementation of inclusivity practices is necessary. In addition, the spread of disinformation can be propagated, thereby making searches less accurate and reliable.[61] While it is challenging to ensure the accuracy of search results since an abundance of (both AI-generated and otherwise) misinformation exists, a combination of manual and systematic control measures is needed to improve the reliability of search engines.

While ChatGPT offers many chances to improve education methods, the ongoing discussions highlight the educational

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

community's diverse opinions. On one side, the rapid rise of this technology has raised concerns about academic integrity. The ability of ChatGPT to generate high-quality texts easily has sparked fears of increased plagiarism. Some academic institutions, like the New York City Department of Education, banned ChatGPT use on their networks and devices on January 2023, soon after the release of ChatGPT.[62] As ChatGPT gains widespread popularity, educational authorities worldwide are promptly taking positions on Generative AI tools such as ChatGPT disconcerting in cheating. Notably, the Departments of Education in the Australian states of Queensland and New South Wales have banned ChatGPT, implementing a firewall to restrict access.[63] The Department of Education in the Australian state of Victoria also opted to block access to ChatGPT based on its terms of use, which specify usage only for those above the age of 18), a condition that excludes most school-going students.[64] These bans aim to manage the use of the technology, whether on a permanent or temporary basis, while policies and regulations are being developed or revised.

While these bans could be practically ignored by students who used generative AI on their smartphones or non-school networks, they demonstrated a clear stance against the potential negative outcomes associated with the use of generative AI in education. Two major concerns were expressed in explaining the decision to ban access to ChatGPT: 1) ChatGPT provides rapid (although not always factually accurate) answers to prompts which can hinder the development of critical thinking and problem-solving skills, and 2) the safety and accuracy of ChatGPT-generated outputs is unreliable and can even be harmful to students. However, less than a year after this ban was imposed, the decision was reversed, and an AI Policy Lab was formed to ensure the safe and responsible incorporation of generative AI into educational practices.[65] Some of the major factors that influenced the decision to reverse the ban were 1) the potential of generative AI in aiding students and teachers in maximizing educational potential, 2) the need to carefully explore the potential and risks associated with the use of generative AI in education, and 3) enabling students to adapt to and prepare for the future which will inevitably involve generative AI and related technologies. Hence, the AI Policy Lab was initiated to ensure equitable access to AI technologies and foster the ethical use of generative AI in education.[66] Initially developed by the EDSAFE AI Alliance, the SAFE Benchmarks Framework guides AI Policy Lab activities. The framework is built on five major areas: engagement, procurement, interoperability and privacy, capacity building, diversity, equity and inclusion, and safety and security and ultimately aims to ensure safe, accountable, fair, and efficacious AI use in education.

Sciences Po Paris has imposed similar restrictions and the University of Hong Kong[67] reflects worries about ChatGPT undermining the educational process. Some early users see it as a groundbreaking technology boosting student motivation, while others worry about its potential to encourage superficial learning and weaken critical thinking.[68,69] The educational sector's interaction with ChatGPT mirrors society's broader engagement with AI. As later explained, perceptions of ChatGPT have evolved from initial enthusiasm to a more cautious recognition of its capabilities and limitations.[70] Similarly, it was suggested that while ChatGPT offers opportunities for creative and non-creative writing, teaching and learning, as long as it is used carefully, but also is a potential deception tool.[71] On the contrary, 38% of top-100 ranked universities have released statements or imposed regulations on ChatGPT use in higher education. Among the top 500-ranked higher education institutions that have adopted regulations regarding ChatGPT use, 23.3% of US colleges and 39.1% of UK universities have taken steps to ban ChatGPT.[72] Concerns regarding cheating and academic dishonesty were listed as the most common reasons for universities imposing bans on ChatGPT use. Despite these concerns, some institutions have embraced the use of ChatGPT in education while simultaneously updating their regulations of academic conduct. Following country-wide access restrictions from China, Iran, Russia, Cuba, and North Korea, Italy has also prohibited the use of ChatGPT after concerns over data privacy and unauthorized data collection by ChatGPT.[73]

Although both benefits and risks associated with using generative AI have been thoroughly explored, specific learning outcomes have not yet been evaluated using control groups. Such studies can guide best practices in generative AI integration into education.

3.2. Enrichment of Curriculum, Formative Feedback, and Assessments

ChatGPT can benefit to mentors, students, and researchers in higher education (**Figure 2**). Course instructors can leverage ChatGPT to develop curricula or weekly schedules, prepare course materials or assessments, and offer personalized study tools for students. Recognizing both the advantages and challenges of integrating AI into teaching is crucial. In contrast, course instructors can leverage the capabilities of ChatGPT as experts in their fields; students and non-expert individuals may encounter difficulties. A main advantage of utilizing ChatGPT is to generate a range of course materials from multiple choice questions to open-ended and case-based assignments, enabling instructors to offer more study materials for students in less time.[74] Dickey et al. introduced a GenAI content generation framework as an educational tool for educators. This framework promises quicker, high-quality, and more comprehensive educational content. They conducted an anonymous survey to support its potential, revealing that 50% of their students preferred GenAI applications, benefiting from the detailed content.[71,75] While the current limitations of ChatGPT such as the generation of misinformation should not be overlooked, such drawbacks can easily be identified by experienced instructors. Besides the generation of traditional course materials, instructors can also assign students to use ChatGPT to complete certain tasks. For example, ChatGPT-based problem-solving assignments in engineering classes can help students visualize concepts, receive step-by-step feedback to identify individual errors, and develop key problem-solving skills.[76]

In various studies, feedback emerged as a recurring theme, often involving providing students with text and/or images as part of formative evaluation. In 2020, Mousavi et al. took a step further by creating a system that delivered automated personalized feedback to first-year biology students.[77] This feedback was tailored to each student's demographics, attributes, and

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

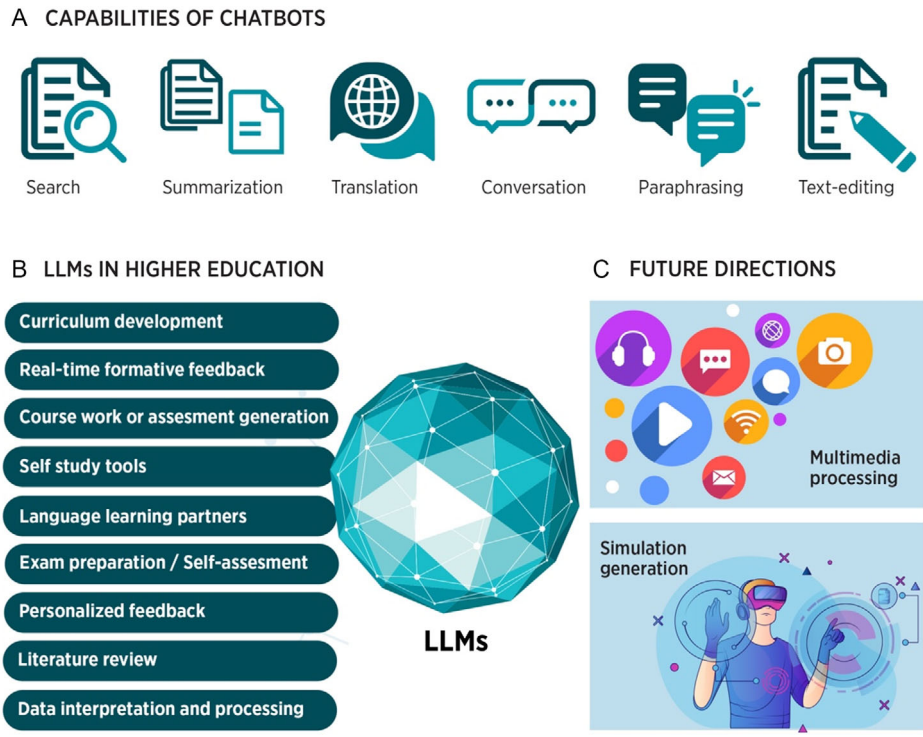


Figure 2. The advantages of Chatbots in learning and teaching applications. A) Commonly used capabilities of chatbots. B) LLMs can be employed by instructors in higher education to enrich curriculum, offer formative feedback, and generate coursework or assessment tools. Students can also utilize them to enhance learning opportunities as self-study tools or language learning partners. Students can leverage LLMs to provide personalized feedback or to develop exam preparation materials. LLMs can facilitate more efficient literature review processes and aid in data interpretation or processing for researchers in higher education. C) Future directions of LLM use in higher education include multimedia processing by LLMs and simulation generation. Multimedia processing can allow students to receive outputs in various formats including images or videos and thereby support learning processes. Simulators can be enhanced with LLMs, enabling user-friendly interfaces and facilitating interactive learning experiences. Some elements in this figure were designed using resources from freepik.com.

academic standing. LLMs present an exciting prospect for advancements in learning assessment within feedback and evaluation. A study questioning the efficacy of ChatGPT as an automated essay-scoring tool was examined.[78] The findings indicated that ChatGPT not only maintained consistency in scoring and expedited the grading process but also delivered immediate scores and valuable feedback on the students' writing proficiency. Similarly, tools such as the Intelligent Essay Assessor[79] are pivotal in evaluating students' written assignments and offering constructive feedback. Beyond its usage in reducing the grading workload for instructors, automatic grading has positively impacted students with varying needs. For instance, an automated assessment was employed to refine the academic writing skills of Uyghur ethnic minority students in China.[80] Given the cultural intricacies of writing, the research illustrated that students interacted cognitively with the automatic assessment system and facilitated self-regulated learning. All these developed tools underscore the transformative potential of GenAI in teaching and learning and ultimately have the potential to enhance the outcome of higher education.

Alternatively, ChatGPT can support clinical instructors by increasing the number of interactive problem-based learning (PBL) tools available to healthcare students.[81] Given that clinical cases provided to students are reviewed for realism and accuracy by instructors, the use of ChatGPT in creating PBL case studies can not only increase the number (and potentially diversity) of clinical scenarios students encounter without significantly adding to the workload of instructors but also promote engagement by presenting cases in an interactive or game-like manner. The interactive nature of such learning tools could be enriched by using AI-based chatbots to generate virtual patient dialogues, enabling students to practice their clinical skills by interacting with digital patients to improve their clinical decision-making.[82] While the use of generative AI in medical education is promising, several concerns regarding the creation of false information (a phenomenon known as hallucinations), outdated information presentation, and ethical practices remain.[83] It is essential to verify LLM outputs as outdated data or misinformation can significantly hinder medical education and practice. While LLMs that are designed specifically for biomedical or medical purposes have been developed, the problem of hallucinations has not been fully addressed yet.[84,85] In addition, depending on the data used to train LLM models, biases can be perpetuated, leading to discriminatory output generation.[86] Finally, over-reliance on LLMs for medical information retrieval or support in decision-making can ultimately deprive students of opportunities to practice

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

critical thinking and hinder their learning. Thus, the responsible and ethical implementation of LLMs in medical education is paramount.

Beyond medical education, similar ChatGPT-assisted PBL case studies can be employed by undergraduate and graduate-level course instructors including engineering, natural sciences, and social studies. In addition, ChatGPT-based assignments or assessments can be used in conjunction with AI-driven learning analytics to further evaluate student performance to enable more tailored support and enhance personalized learning experiences.[87] Similar strategies can also be used to assess the general understanding of course materials, allowing instructors to make appropriate curriculum changes to improve teaching outcomes.

3.3. Enhancing Learning Experience

In 1999, John Biggs introduced the 3 P model, "Presage - Process - Product" for learning and teaching, mentioning the significance of students' perceptions in education.[88] This well-established framework has been the fundamental educational guideline so far, emphasizing students' approaches to learning. Even though the evaluation criteria of enhanced learning may present challenges and require in-depth research by education experts, valuable insights can be derived from a survey about the efficiency of ChatGPT on the learning process.[89] The survey consisted of 399 higher education students and demonstrated that the majority of the students hold positive attitudes toward ChatGPT, an outcome that overlays with "presage" step of the 3 P model which defines students' acceptance and adaption of new techniques.[88] In addition, the survey demonstrated that ChatGPT provides a helpful starting point by answering simple questions and inspires students as virtual assistants who often face challenges initiating their assignments. Moreover, another review paper of 138 articles presented the potential strengths of the ChatGPT in education[90] which comply with the "process" stage of Biggs's 3 P model as follows: i) facilitating access to learning materials for inclusive learning,[91] ii) providing immediate answers to student questions, iii) delivering instant personalized feedback.[92] Considering all this positive feedback, ChatGPT seems in correlation with "Presage and Process" stages of the 3 P framework in terms of providing a stress-free educational setting and personalizing the learning experience.

Another benefit of ChatGPT on learning is its capability to summarize long texts and academic manuscripts and tutor the course content by enabling rapid access to key information and data.[93] For example, a chatbot-based learning system was developed to support coding students without a course instructor.[94] Another system, Coding Tutor, was able to respond to open-ended questions in natural language (similar to ChatGPT), assess source code, provide feedback, and offer step-by-step training exercises to help students improve their coding skills. Similarly, in subjects such as anatomy, physiology, or histology, obtaining a deeper understanding of topics by obtaining answers to questions on demand can enhance student understanding, develop more efficient problem-solving skills, provide critical thinking abilities, and improve information retention.[95] Moreover, ChatGPT effectively offers feedback on assessments and grades during the learning experience, improving the ability to work with different assignment tasks, which can be counted as enhancement in the learning outcome, related to the product stage of the 3 Ps. In this regard, ChatGPT assists in checking and initially grading completed homework, contributing to students' improvement in their assignments and writing skills. Due to its ability to generate natural language text outputs mimicking human interactions, ChatGPT can also enhance students' language skills when used as a virtual conversation partner in linguistics courses for foreign language learners.

An evolutional attempt in enhancing education with GenAI tools was presented by Arizona State University (ASU) in collaboration with OpenAI in the first quarter of 2024. ASU invited their staff and faculty to implement ChatGPT and submit applications to enhance student achievements, optimize the administrative plans and invent research projects.[96] Giving the school access to ChatGPT Enterprise to ASU, the COO of OpenAI mentioned that the reason for loving ChatGPT is learning. In addition, ASU aims to develop AI tutors to help their students in writing, studying, and learning specific courses.[97] Therefore, this collaboration can be evaluated as a revolution in educational techniques directly affecting students' learning performance and engagement, as well as clarifying the prominent role of ChatGPT in learning. To evaluate the role of ChatGPT and other GenAI tools to facilitate learning enhancements and to analyze the effect of ChatGPT in teaching and learning, Mai, D.T.T., et al. also posed similar questions referring to the three stages of Bigg's 3 P learning framework.[98] Overall, the improved learning experience by ChatGPT seems related mostly to the "process' and "product" stages of the 3 P model; and first-of-its-kind collaborations, like ASU & OpenAI, can lead to a new era in higher education and expand the AI impact in teaching and learning.

3.4. AI-Powered Research Support and Decision Making

Besides offering conceivable benefits for students and mentors in higher education, ChatGPT can also assist researchers in summarization and data collection. Initial attempts to assess the automatic summarization skills of ChatGPT focused on news content, demonstrating high factuality on ChatGPT-generated summaries and a correlation between instruction tuning and zero-shot summarizing capabilities.[99,100] Similarly, ChatGPT appeared superior to traditional fine-tuning methods on query-based summary generation tasks across diverse domains including stories, dialogue meetings, news, and Reddit posts.[101] While ChatGPT-based summarizing can enable researchers to save time by producing factually accurate summaries, for longer texts on topics that require a higher degree of specialization, such as medicine, ChatGPT falls short in achieving a similar level of success. ChatGPT-based clinical summaries have been found to contain factually inconsistent, overly convincing, and uncertain statements, which can potentially lead to the spread of misinformation.[102] To address the challenge of domain-specific ChatGPT performance, an iterative framework for ChatGPT was optimized to summarize radiology reports.[103] In addition to data summarization, ChatGPT can also play an important role in data collection or classification. For example, a zero-shot information extraction (IE) strategy has been generated using

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ChatGPT.[104] Moreover, LLM-based IE strategies can further facilitate data interpretation and processing without requiring advanced coding skills.

Other applications of ChatGPT include the generation of engineering simulations, user-friendly optimization strategies, and bug-fixing for coding applications, all of which can enable faster and more efficient design and manufacturing processes.[14,105] Similar LLM-based simulations can be applied to macroeconomics to mimic human-like decision-making processes.[106] Such simulations not only have significant implications for research (i.e., enabling researchers to make predictions or optimize designs) but can also allow students to have a deeper understanding of complex subjects. Although not yet transferrable to daily clinical practice, the potential use of ChatGPT for clinical applications has also been explored.[107] For example, ChatGPT treatment recommendations for breast cancer patients were evaluated to assess the use of ChatGPT as a support tool for tumor boards, demonstrating that 70% of ChatGPT-generated recommendations were comparable with those of the tumor board.[108] ChatGPT outputs for urology cases revealed poor quality responses, reiterating the need for further research before adopting such tools in clinical practice.[109] Alternative approaches for the use of ChatGPT in medicine include the employment of ChatGPT to identify research priorities for gastroenterology.[110] Despite obtaining high scores for relevance and clarity, originality scores remained low in ChatGPT generated information.

Despite promising case studies, an objective evaluation of ChatGPT and LLM capabilities is required. NLP performance metrics including the Bilingual Evaluation Understudy (BLEU), Recall Oriented Understudy for Gisting Evaluation (ROUGE), and Metric for Evaluation of Translation with Explicit Ordering (METEOR) can be employed to quantitatively evaluate LLM's success in generating summaries or supporting decision-making processes. When evaluated by these metrics on its ability to provide transparent and reliable outputs for educational questions, vanilla ChatGPT (the initial version of ChatGPT that does not use an information retrieval system (IRS)) scored 0.275, 0.444, and 0.379 on the BLEU-1, ROUGE-L, and METEOR tests, respectively.[111] ChatGPT with five-shot IRS demonstrated significantly improved outcomes with scores of 0.580, 0.745, and 0.543 on the BLEU-1, ROUGE-L, and METEOR tests, respectively. The use of such metrics can prove particularly helpful in advancing educational tools. Research guided by such metrics can also help identify and address shortcomings of ChatGPT and other LLMs. For instance, an educational question-answering system (EQA) was developed to overcome hallucinations and LLM's limited access to novel data (beyond its pre-training dataset). To this end, an IRS was integrated to the LLM and the EQA outperformed vanilla ChatGPT by 110.9%, 67.8%, 43.3%, and 9.2% on BLEU, ROUGE, METEOR, and BERTScores, respectively.

Similar quantitative LLM assessment strategies have also been employed to explore the efficacy of LLMs in summarizing content in specialized topics such as mental health or medical reporting with LLMs achieving satisfactory or even excellent scores (i.e., GPT 4.0 scoring 0.4011 on ROUGE-1 for clinical note summarization).[112–114] Notably, when BART, BERT-based summarization (BERTSUM), and GPT 3.5 or higher were evaluated on medical dialogue summarization, BART received higher ROUGE-1, ROUGE-2, ROUGE-L, and BERTScores than GPT.[115] However, when summaries were evaluated by medical professionals, GPT-generated summaries were preferred over BART, suggesting that quantitative scoring systems alone might not be sufficient to evaluate the capabilities of LLMs. Nevertheless, a combination of quantitative evaluations and expert reviews can help advance LLM capabilities significantly, allowing their implementation in education. Moreover, randomized controlled experiments can further evaluate the impact of LLMs in research and education.

3.5. AI as an Author of a Publication

The role of generative AI in scholarly publishing has also initiated several debates. While some have argued that AI-generated text outputs could be used in academic manuscripts, others have advocated against such practices. In December 2022, ChatGPT was listed as an author of an academic article focusing on using ChatGPT in medical education.[116] Other preprints and editorials where ChatGPT was cited as a first author followed soon after.[117,118] According to co-author Almira Osmanovic Thunström, the paper faced rejection from one journal after review. However, it found acceptance in a second journal after she revised the article in response to reviewer requests by acknowledging the GPT-3 as an author.[119] The team of well-known citation style, the APA format, posted instructions on their blog entitled "How to cite ChatGPT".[120] It is advised to describe the tool in the Methods section if the AI tools are used in a research paper, but authors might explain the use of AI tools in the introduction while preparing a literature review or a response paper. Despite this recommendation, the APA style guidelines related to ChatGPT and AI topics are not well-defined, and all publishers and academic institutions should focus on the policies for defining ChatGPT as a collaborator.

In contrast, several journals, editors, and scientists have disagreed with such practices, claiming that authors should have legal and ethical accountability and that LLMs fail to fulfill this.[119] Journals and publishers including Science, Nature, and arXiv have released official statements claiming that ChatGPT or other LLMs cannot be granted authorship as AI-based tools cannot meet authorship requirements such as having accountability or providing consent. Similarly, the International Committee of Medical Journal Editors (ICMJE) has updated their guidelines explaining that if applicable LLM use should be cited in the cover letter and manuscript but would not be accepted as an author. Elsevier has specified instances where AI-assisted technology would be permitted. These include using LLMs solely to improve readability. However, it has been explicitly described that uploading manuscripts to AI tools or using generative AI to evaluate manuscripts is prohibited. Moreover, among 300 journals, almost 60% have developed guidelines relating to the use of generative AI in scholarly publishing.[121] While most (96.6%) guidelines allow the use of generative AI for manuscript quality improvement, 98.9% of guidelines prohibit AI-based tools from being listed as an author. Although an international consensus has not been reached, these guidelines reflect concerns regarding accountability, reliability, and transparency when AI-assisted devices are employed in scholarly publishing.

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Furthermore, the U.S. Copyright Office, Library of Congress published a policy clarifying its practices in registering and examining the submitted material created by AI technology.[122] According to the guideline for copyright applicants, an author who arranges the human and non-human content in their work is recommended to claim the contributions from either human-authored or AI-generated. Like journals and publishers that do not grant authorship to AI-based tools, the U.S. Copyright Office also disclaimed the inclusion of AI tools as authors or co-authors.[122] Moreover, the U.S. National Science Foundation (NSF) declared a notice for research proposers by encouraging them to indicate why they prefer GenAI technology in their project description.[108] The NSF also mentioned that GenAI tools might lead to violations in the integrity and authentication of proposals and reports of NSF-funded projects. Furthermore, they have explained that their guidelines on falsification, plagiarism, or fabrication, would be updated accordingly to address potential research misconduct and guide researchers on the appropriate use of GenAI. In addition to copyright conflicts, the NSF is concerned about authors uploading any part of proposals and reviews on the open internet, which implies that GenAI threatens the merit review process by violating confidentiality principles.[123] However, interestingly, OpenAI has launched the "NSF Proposal Partner" for users of ChatGPT Plus and the virtual assistant claimed that it was trained on previous successful NSF proposals.[124]

Although an international consensus has not been reached, these guidelines reflect concerns regarding accountability, reliability, and transparency when AI-assisted devices are employed in scholarly publishing. Moreover, the stage of controversy of AI in science persists: King. R.D. et al. formulated the Stockholm Declaration on AI ethics for Science[125] to prevent the harms of AI on research and scholars and to help navigate the privacy and ethics issues. The team also invited researchers to sign this declaration and pledge to employ AI responsibly and ethically, which could gather scholars following the same frameworks.

4. Challenges and Concerns

Despite its conceivable benefits, the widespread use of ChatGPT or similar LLMs in higher education is complex due to several concerns. In addition to professional societies, publishers, and scholarly institutions, governments have also adopted different regulations and guidelines relating to the use of ChatGPT and LLMs. The global heterogeneity in generative AI regulation is clear.[126] While a comprehensive legislation has not been released yet, China has imposed regulations on AI-based tools on a case-by-case basis. In contrast, the European Union (EU) has released a detailed set of regulations on AI use and an AI Liability directive is underway. The legislation classified AI applications under high and low-risk categories with high-risk applications described as those potentially threatening fundamental rights, safety, and democracy. The United States has also planned to impose certain regulations to improve transparency and accountability for AI-based applications following an executive order released in late 2023. Alternatively, the UK has released a statement explaining that short-term plans do not include imposing regulations on AI use.

4.1. Copyright and Ethical Issues

Copyright and ownership issues relating to ChatGPT revolve around two fundamental components of LLMs: the data used to train the model and the output texts generated by the model.[127] Legally, ChatGPT cannot hold copyright stems as it is considered a non-human entity and copyright issues require human authorship in most countries. Thus, given that the output is sufficiently edited for originality by the user, the user could be considered the owner of the work. Alternatively, as ChatGPT cannot legally own work, ownership of ChatGPT- generated work can be granted to developers of ChatGPT. Moreover, the lack of legal consensus over the copyright of AI-generated content continues to fuel such debates. Legal issues regarding the originality or authenticity of AI-generated works stem from varying national policies and definitions of intellectual property ranging from the ability to make creative choices to a "minimal degree of creativity", thereby complicating the adoption of a national-level and international consensus. The fact that most AI-generated texts are difficult to detect using traditional plagiarism-detection tools further complicates the issue of copyright and intellectual property ownership. To mitigate this issue, digital watermarking has been suggested to detect and/or trace LLM-generated content.[128,129] Although watermarking alone cannot solve the issue of copyright, it can help resolve concerns regarding academic integrity.

4.2. Legal Issues

Another issue regarding ChatGPT-generated content is related to the use of creative works of others in training ChatGPT, which can lead to legal consequences if perceived as an exploitation of one's work.[127] In addition to copyright and intellectual property, the use of ChatGPT in fields such as medicine and healthcare can also give rise to legal and ethical issues regarding patient privacy. In many countries, patient privacy is protected under various laws such as the Health Insurance Portability and Accountability Act (HIPAA) in the US and the Data Protection Act in the UK. Furthermore, particularly for applications of ChatGPT in healthcare, regulations are required to prevent identifiable patient medical or personal information gathering and storing by ChatGPT. The lack of regulations can result in a breach of patient privacy, resulting in legal consequences and threatening patient safety.[130] For example, the Information Commissioner's Office in the UK has stated that the Royal Free Hospital failed to comply with the Data Protection Act as it passed the personal data of 1.6 M patients to DeepMind in 2015 and 2016. Ensuring transparency in healthcare data collection and usage as well as bias detection and prevention for AI-driven platforms is essential to prevent legal consequences.[131]

4.3. Data Security

The responsibility for safely handling private information falls on OpenAI as expressed by their privacy policy, making it difficult to conduct audit processes.[132] The lack of transparency in data collection and utilization by ChatGPT poses significant risks to data

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

privacy.[127] Moreover, with an increasing number of LLMs, public data can be exploited if strict regulations are not imposed. Moreover, it has been demonstrated that given an identification number, ChatGPT can retrieve personal information (i.e., date of birth) associated with the identification number, suggesting that sensitive information storage cannot be fully prevented.[132] In addition to the public data used during LLM training, despite security precautions, an increasing number of cybersecurity attacks can result in personal data leakage, presenting an imminent threat to data security.[133]

4.4. Accuracy, Misinformation, and Unclear Sources of Data

GPT-4 fails to assess the accuracy of information dating after September 2021 as most of the pre-training data was cut off during that time.[42] This not only indicates that some content generated by ChatGPT will inevitably be outdated, but also highlights the need for constant information feeding to the ChatGPT training database to obtain up-to-date content generation. Furthermore, even when the model's training dataset is current, the risk for misinformation remains significant as internet-based datasets are likely to contain misinformation, biases, and manipulative content and the model remains susceptible to potential counter-attacks to introduce false information.[134] ChatGPT can also generate fake citations, create false information (hallucinations), fail to provide sources for generated content, and propagate "infodemics" by allowing the spread of misinformation.[135–138] Other challenges regarding the accuracy of ChatGPT-generated content include the lack of realism for subject-specific content generation. For example, while ChatGPT can provide highly accurate medical information on clinical cases (except for those requiring experience in highly specialized areas), the inability to capture nuances of real-life scenarios remains an important challenge.[139]

4.5. Excessive Reliance of Students/Users on ChatGPT

Although the use of ChatGPT by students can enhance their learning experience, it can also lead to an excessive reliance on AI-generated content, eventually hindering diversified learning experiences. This can prevent the development of critical thinking skills, impede writing/coding abilities, and deter students from achieving deep content understanding. Plagiarism detection tools that differentiate between AI- and human-generated text have been suggested to prevent excessive student reliance on ChatGPT. Although AI-generated text detection software development is underway, diversified learning approaches can achieve better success in prompting students to gain a deeper understanding of concepts and not rely solely on ChatGPT to complete assignments.[140] Examples include assigning writing tasks that require students to include personal experience/perspectives (which is challenging to imitate using ChatGPT), requiring students to submit video reflections discussing assignments, adding in-person assessments and presentations, or developing assignments that incorporate the use of AI-based tools in creative ways.[76,140,141]

4.6. Plagiarism and Cheating

Whether ChatGPT-generated or amended (grammar, clarity, structure) texts can be considered original or should be treated as plagiarism has remained controversial. While some have pointed out that ChatGPT-generated texts if merely rephrased or edited do not constitute plagiarism, others have strongly refuted the use of ChatGPT to create written content. As any ChatGPT-generated text requires prompt input from the user, critics argue that there is a need for increased scrutiny in assessing plagiarism, cheating, and academic honesty in higher education.[142] Part of the concern regarding plagiarism is centered around the irresponsible use of ChatGPT by students to outsource written work and ultimately cheat on assignments.[140] Such academic dishonesty is especially concerning as traditional plagiarism-detection tools are limited in detecting AI-generated text. While plagiarism detection software providers such as Turnitin have proposed AI writing detecting capabilities, these solutions have not been tested in advanced LLMs. Such false positives result in plagiarism detection could lead to false academic misconduct cases and unfair treatment of students. Another point of concern regarding academic dishonesty and plagiarism relates to the use of ChatGPT in research. Moreover, several academic journals have already set regulations regarding using ChatGPT-generated texts for academic articles or listing ChatGPT as an author in submitted manuscripts.[143]

4.7. Bias

Some of the data used in training ChatGPT could be inherently biased. Thus, when ChatGPT or AI-based tools trained on biased data are used for research purposes, preexisting inequities can be perpetuated.[141] For example, it has been demonstrated that ChatGPT can assign gender and race to certain occupations or generate outputs that contain microaggressions.[144] Similarly, although less biased than humans in mathematical problems, ChatGPT presents biases in ambiguous and complex human matters.[145] In addition, the overconfidence of ChatGPT can exacerbate biases in such matters. In trying to avoid further propagation of biases through ChatGPT-generated outputs, several strategies have been employed. While these approaches have prevented ChatGPT from bluntly producing discriminatory content to some extent, the issue of bias in LLMs has not been fully resolved. Moreover, while inherently biased inputs can typically be detected by ChatGPT, several jail-breaking methods (exploiting the flaws) can allow users to bypass ChatGPT restrictions.[133]

4.8. Prompt Engineering

Another shortcoming of ChatGPT is that the accuracy of the output hinges on prompt engineering. The output content is highly dependent on the user-generated prompt. This presents an important challenge particularly in the field of higher education as output quality is of utmost importance for applications in specialized fields. For example, when inadequately engineered prompts are provided to ChatGPT, the possibility of obtaining factually unreliable outputs increases.[146] As a result, whether students use ChatGPT to study course content or instructors

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

use it to generate study materials, inadequate prompt engineering can significantly hinder the learning process. Other pitfalls of inaccurate prompt engineering include bias reinforcement through inherently discriminatory inputs (although mitigated to some extent) or the lack of satisfactory responses as a result of overfitting of prompts.[147] In addition, conflicting prompts can generate non-sensical outputs or result in inaccurate information dissemination.

4.9. Open-Source LLMs

Open-source LLMs, such as Large Language Model Meta AI (LLaMa) developed by Meta and Open Language Model (OLMo) developed by the Allen Institute for AI, have attracted widespread attention due to their accessibility and adaptability.[148–151] While open-source LLMs can have broad implications in education (i.e., to foster collaborative work spaces or empower researchers), the flexibility of open-source LLMs can also be prone to misuse. Open-source LLMs have less centralized oversight and less quality control or validation measures, potentially resulting in security vulnerabilities.[152] In addition, it can also prove challenging to determine responsibility or keep users accountable, thereby jeopardizing data security as well as output accuracy.

4.10. Adaptation to Real-World Educational Settings

Another important challenge is to ensure robust LLM performance when adapted to real-world educational settings. While most case studies have demonstrated promising results, some potential issues have not been thoroughly accounted for. For example, noisy data input can lower LLM accuracy and lead to contextual misunderstandings. Although fine-tuning can help mitigate high levels of static noise, it has been demonstrated that models can be prone to lower levels of dynamic noise.[153] Similarly, LLM robustness and capabilities can be affected by adversarial inputs. Moreover, models have adversarial vulnerabilities, which can be difficult to anticipate as these vulnerabilities are not consistent across different families of LLMs.[154] LLMs used in real-world educational settings are likely to encounter noisy data as well as adversarial inputs. This can result in variable output generation, inaccurate data dissemination, and superficial processing, which would hinder educational objectives. Thus, strategies to predict and address these limitations are required to maintain the robustness and accuracy of models, particularly for use in AI-powered decision-making.

LLM implementation in real-world educational settings also comes with practical challenges such as domain-specific LLM adaptation, deployment issues, and scalability and maintenance requirements. To ensure alignment with specific educational curricula and institutional regulations, domain-specific LLM fine-tuning might be necessary. However, while different fine-tuning strategies can allow LLMs to adopt desired functionalities, fine-tuning often significantly increases computational costs.[155,156] Infrastructure and LLM maintenance would also result in increased computational costs. Although these can be addressed partially with novel optimization strategies, scalability can prove challenging.[157,158] Thus, widespread implementation of LLM-powered educational models relies heavily on adjusting LLM structure to align with institutional needs and pre-existing systems.

5. Future Perspective

Recently, ChatGPT has expanded beyond text-based platforms to accept image, video, or sound-based inputs. The future versions of ChatGPT will be able to create multimedia outputs. Integration of ChatGPT with other generative AI-based tools can enable ChatGPT to obtain higher functionalities. For example, multimedia content synthesizing generative adversarial networks (GANs) can expand the use of ChatGPT in higher education.[159] To this end, a prompt manager that can bridge the gap between ChatGPT and visual foundation models has been developed.[160] This prompt manager has enabled the conversion between visual input/outputs and text-based content, allowing ChatGPT to process visual tasks. The development of advanced visual processing capabilities can play an important role in higher education by (i) aiding students in converting textual information into visuals (such as mind maps or diagrams), (ii) generating 3D or video-based engineering simulations, and (iii) facilitating radiological or pathological image translation.[161]

While chatbots have not been used for this purpose yet, AI and/or machine learning (ML)-based interactive tools have gained popularity owing to their diverse and adaptable nature. Recently, Google Gemini has emerged as a generative multimodal processing platform.[162] Using the ability to transition between audio, video, image, text, and code, Gemini achieves massive multitask language understanding (MMLU). As a result, it enables multimodal dialogues. However, similar to many generative AI models, Gemini has also faced significant criticism regarding the inaccurate depiction of historical images, amplification of stereotypes, and propagating racial biases.[163,164] As a result of growing controversy and concern over the spread of AI-generated disinformation, Google restricted Gemini from answering election-related questions in several countries including the United States and India.[165] Recent changes have also sparked debates regarding the safety and accuracy of generative AI use for other applications including health, education, and finance. Similar to Gemini, OpenAI's DALL-E has also emerged as a multimodal LLM-based tool. While studies have demonstrated that DALL-E could potentially be used as an educational tool depicting medical, societal, and economical perspectives, images lacked symbolic representation.[166] Likewise, when DALL-E was asked to illustrate congenital heart defects, images were labeled anatomically incorrect, hindering its use in education.[167]

Other strategies have included the integration of ML-based tools to develop simulator models. For example, ML-augmented fluid dynamics simulations were conducted using an interactive educational micromixer model.[168] Undergraduate, master's, and graduate student responses to the educational model were positive, demonstrating that interactive models can significantly contribute to student learning. Potential benefits of LLM integration in education have been suggested; however, the impact of LLMs on learning outcomes has not been quantified using some of the traditional methods that were developed for assessing the

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

value of new approaches in teaching and training. ChatGPT integration or complementation can allow interactive simulations to expand their applications by responding to user inputs, offering a more personalized interactive educational experience. The integration of interactive learning models and combinations with outstanding frameworks (e.g., John Biggs's 3 P) can assist all institutions and professionals to set rigorous learning policies with ChatGPT. These well-established "human-controlled" frameworks can be the best role models for GenAI tools to assist students with a real-world mindset, considering every critical aspect of learning and critical thinking.

Furthermore, the effect of ChatGPT on the learning experience needs a reliable assessment policy as it has been applied in traditional educational set-ups so far. The outstanding and most successful educational models, (SURE, CURE, and ROLE) were developed by Lapatto D. to gauge the effect of different interactive learning models on students.[169–171] For example, in the CURE model, the instructor provides expertise and a broad framework, refraining from predetermining the "correct answer." Instead, students undertake research projects, allowing them to generate new knowledge like an independent research endeavor. In addition, LLM-powered education can support non-traditional students in their learning processes. For instance, ChatGPT can offer continuous and on-demand support for non-traditional learners or help create virtual inclusive spaces to promote engagement.[172] Furthermore, the development of targeted generative AI models to assist in the learning processes of dyslexic individuals or individuals on the autism spectrum has also been suggested.[173,174]

Considering the benefits of ChatGPT in learning, such educational models and interactive evaluation approaches can be developed to monitor the advancements in student performance. They can potentially end up with guideline/policy updates and curriculum modifications in AI-based education. For instance, a team at Harvard Graduate School of Education had an attempt to monitor the changes in periodic teacher feedback with GenAI[175] and tested the impacts of student belonging and their productivity in that graduate-level course. Similar strategies and conventional performance evaluation approaches can be a guide to develop assessments to evaluate learning with GenAI tools. The increasing number of ChatGPT users seems to be a big cohort for researchers to develop models to test the efficiency of learning with ChatGPT and discuss the pedagogical impacts on students. Thus, a novel field exploring the role of AI-based regulations and GenAI tools to control and monitor learning advancements can emerge.

6. Conclusion

LLMs have attracted significant attention due to their ability to mimic natural human language. Chatbots such as ChatGPT have been used for various applications in a multitude of fields including engineering, medicine, and education. Particularly in higher education, ChatGPT can assist instructors in developing curricula and course materials, provide students with personalized learning tools, and optimize data summarization or coding processes for researchers. However, limitations and concerns regarding the use of ChatGPT in educational settings remain a debated topic. Although significant effort has been put in to mitigate misleading information on the ChatGPT platform, the risk of bias and inaccurate output generation persists. Additionally, legal issues regarding data privacy and protection, copyright matters, and concerns of originality and plagiarism continue to instigate debates. Finally, the potential excessive reliance of students on platforms such as ChatGPT may complicate the widespread adoption of diverse learning methods in higher education. Several strategies such as OpenAI restrictions preventing biased output generation have been implemented. The use of watermarking to identify ChatGPT-generated outputs, modifying course content to use ChatGPT in creative ways, or placing regulations to ensure data safety have been suggested to address some of the issues. In addition, the integration of ChatGPT with other AI-based tools can allow multimedia processing and enable the development of interactive simulations. Thus, the advances in LLMs will enable the delivery of personalized learning experiences to students in higher education settings. To overcome current challenges, objective strategies to evaluate the impact of GenAI tools on the learning experience are essential. Similarly, AI-powered educational tools should be evaluated on performance metrics to better analyze shortcomings. In addition, the issues of bias and misinformation should be addressed through repeated quality assessments. Finally, prior to the adaptation of LLM-based tools into real-world educational settings, regulations and guidelines addressing ethical and legal issues should be developed.

Acknowledgements

S.T. acknowledges Tubitak 2232 International Fellowship for Outstanding Researchers Award (118C391), Alexander von Humboldt Research Fellowship for Experienced Researchers, Marie Skłodowska-Curie Individual Fellowship (101003361), and Royal Academy Newton-Katip Celebi Transforming Systems Through Partnership award (120N019) for financial support of this research. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the TÜBITAK. This work was partially supported by Science Academy's Young Scientist Awards Program (BAGEP), Outstanding Young Scientists Awards (GEBIP), and Bilim Kahramanlari Dernegi the Young Scientist Award.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

ChatGPT, education, higher education, pedagogical approaches, teaching and learning

Received: June 3, 2024
Revised: June 30, 2024
Published online:

[1] C. Caudai, A. Galizia, F. Geraci, L. Le Pera, V. Morea, E. Salerno, A. Via, T. Colombo, *Comput. Struct. Biotechnol. J.* **2021**, *19*, 5762.
[2] J. Cowls, A. Tsamados, M. Taddeo, L. Floridi, *AI. Soc.* **2023**, *38*, 283.

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[3] K. Rajan, A. Saffiotti, *Artif. Intell.* **2017**, *247*, 1.
[4] D. Yigci, A. Ahmadpour, S. Tasoglu, *Adv. Sens. Res.* **2023** *3*, 2300109.
[5] C. J. Fluke, C. Jacobs, *WIREs Data Mining Knowl. Discov.* **2020**, *10*, e1349.
[6] B. E. Fund, in *AI for Climate and Nature Workshop*, **2023**.
[7] B. D. Lund, T. Wang, N. R. Mannuru, B. Nie, S. Shimray, Z. Wang, *J Assoc. Inform. Sci. Technol.* **2023**, *74*, 570.
[8] B. Meskó, E. J. Topol, *NPJ Digit Med.* **2023**, *6*, 120.
[9] H. Huang, O. Zheng, D. Wang, J. Yin, Z. Wang, S. Ding, H. Yin, C. Xu, R. Yang, Q. Zheng, B. Shi, *Int. J. Oral. Sci.* **2023**, *15*, 29.
[10] I. El Naqa, A. Karolak, Y. Luo, L. Folio, A. A. Tarhini, D. Rollison, K. Parodi, *Oncogene* **2023**, *42*, 3089.
[11] S. S. Biswas, *Ann. Biomed. Eng.* **2023**, *51*, 868.
[12] O. Baclic, M. Tunis, K. Young, C. Doan, H. Swerdfeger, *Can. Commun. Dis. Rep.* **2020**, *46*, 161.
[13] S. S. Biswas, *Ann. Biomed. Eng.* **2023**, *51*, 1126.
[14] N. M. S. Surameery, M. Y. Shakor, *Int. J. Inform. Technol. Comput. Eng.* **2023**, *17*.
[15] M. Dowling, B. Lucey, *Fin. Res. Lett.* **2023**, *53*, 103662.
[16] P. Hacker, A. Engel, M. Mauer, in *Proc. of the 2023 ACM Conf. on Fairness, Accountability, and Transparency (FAccT '23)*, Association for Computing Machinery, New York, NY **2023**, pp. 1112–1123.
[17] J. Kocoń, I. Cichecki, O. Kaszyca, M. Kochanek, D. Szydło, J. Baran, J. Bielaniewicz, M. Gruza, A. Janz, K. Kanclerz, A. Kocoń, B. Koptyra, W. Mieleszczenko-Kowszewicz, P. Miłkowski, M. Oleksy, M. Piasecki, Ł. Radliński, K. Wojtasik, S. Woźniak, P. Kazienko, *Inform. Fusion* **2023**, *99*, 101861.
[18] J. Paul, A. Ueno, C. Dennis, *Int. J. Consum. Stud.* **2023**, *47*, 1213.
[19] J. Q. Pérez, T. Daradoumis, J. M. M. Puig, *Comput. Appl. Eng. Educ.* **2020**, *28*, 1549.
[20] T.-C. Hsu, H.-L. Huang, G.-J. Hwang, M.-S. Chen, *Educ. Technol. Soc.* **2023**, *26*, 218.
[21] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, M. Stadler, J. Weller, J. Kuhn, G. Kasneci, *Learn. Individ. Differ.* **2023**, *103*, 102274.
[22] S. Shaikh, S. Y. Yayilgan, B. Klimova, M. Pikhart, *Eur. J. Investig. Health Psychol. Educ.* **2023**, *13*, 1937.
[23] S. Kim, J. Shim, J. Shim, *J. Multimed. Inf. Syst.* **2023**, *10*, 79.
[24] P. Adamski, M. Ostrowska, *J. Med. Res. Rev.* **2023**, *1*, 5.
[25] D. R. E. Cotton, P. A. Cotton, J. R. Shipway, *Innov. Educ. Teach. Int.* **2023**, *61*, 228.
[26] P. P. Ray, *Internet Things Cyber-Phys. Syst.* **2023**, *3*, 121.
[27] S. R. Mogali, *Anat. Sci. Educ.* **2024**, *17*, 444.
[28] J. Weizenbaum, *Commun. ACM* **1966**, *9*, 36.
[29] J. G. Wilpon, L. R. Rabiner, C.-H. Lee, E. R. Goldman, *IEEE Trans. Acoust. Speech Signal Process.* **1990**, *38*, 1870.
[30] D. Khurana, A. Koli, K. Khatter, S. Singh, *Multimed. Tools Appl.* **2023**, *82*, 3713.
[31] I. S. T. Mikolov, G. C. K. Chen, J. Dean, Distributed Representations of Words and Phrases and their Compositionality, *arXiv*, **2013**.
[32] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, T. Chen, *Pattern Recogn.* **2018**, *77*, 354.
[33] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, J. Schmidhuber, *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2222.
[34] L. Floridi, M. Chiriatti, *Minds Mach.* **2020**, *30*, 681.
[35] C. Subakan, M. Ravanelli, S. Cornell, M. Bronzi, J. Zhong, in *ICASSP 2021-2021 IEEE Inter. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, **2021**, pp. 21–25, http://dx.doi.org/10.1109/ICASSP39728.2021.9413901.
[36] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, S.-M. Hu, *Comput. Vis. Media*. **2022**, *8*, 331.
[37] C. Zhou, Q. Li, C. Li, J. Yu, Y. Liu, G. Wang, K. Zhang, A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT, *ArXiv*, abs/2302.09419, **2023**.
[38] A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, D. S. W. Ting, *Nat. Med.* **2023**, *29*, 1930.
[39] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, Improving Language Understanding by Generative Pre-Training, **2018**.
[40] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877.
[41] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, *Language Models are Unsupervised Multitask Learners*, OpenAI blog 1 **2019**, *1*, 9.
[42] O. J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, R. Avila, I. Babuschkin, S. Balaji, V. Balcom, P. Baltescu, H. Bao, M. Bavarian, J. Belgum, I. Bello, J. Berdine, G. Bernadett-Shapiro, C. Berner, L. Bogdonoff, O. Boiko, M. Boyd, A.-L. Brakman, G. Brockman, T. Brooks, M. Brundage, K. Button, *GPT-4 Technical Report*, **2023**. https://doi.org/10.48550/arXiv.2303.08774.
[43] M. Kale, A. Rastogi, *Text-to-Text Pre-Training for Data-to-Text Tasks,* arXiv:2005.10433, **2020**.
[44] S. Zhang, S. Roller, N. Goyal, M. Artetxe, M. Chen, S. Chen, Ch. Dewan, *Opt: Open Pre-Trained Transformer Language Models*, arXiv:2205.01068, **2022**.
[45] R. Thoppilan, D. De Freitas, J. Hall, N. Shazeer, A. Kulshreshtha, H.-T. Cheng, A. Jin, *Lamda: Language Models for Dialog Applications*, arXiv:2201.08239, **2022**.
[46] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, *Albert: A Lite Bert for Self-Supervised Learning of Language Representations*, arXiv:1909.11942, **2019**.
[47] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, *Roberta: A Robustly Optimized Bert Pretraining Approach*, arXiv:1907.11692, **2019**.
[48] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, Q. V. Le, *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
[49] K. Clark, M.-T. Luong, Q. V. Le, C. D. Manning, *Electra: Pre-Training Text Encoders as Discriminators Rather Than Generators*, arXiv:2003.10555, **2020**.
[50] X. Amatriain, *Transformer Models: An Introduction and Catalog,* arXiv:2302.07730, **2023**.
[51] *Chatbot Market Size, Share, Growth, Global Trends, Forecast 2030,* Facts and Factors, **2023**.
[52] M. Cheng, *OpenAI is Set to See Its Valuation at $80 Billion—Making it the Third Most Valuable Startup in the World*, **2023**, https://qz.com/openai-is-set-to-see-its-valuation-at-80-billion-makin-1850950928 (accessed: December 2023).
[53] D. Brin, V. Sorin, A. Vaid, A. Soroush, B. S. Glicksberg, A. W. Charney, G. Nadkarni, E. Klang, *Sci. Rep.* **2023**, *13*, 16492.
[54] V. P. S. Grover, L. Ahuja, in *2021 9th Inter. Conf. on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, **2021**, pp. 1–6, http://dx.doi.org/10.1109/ICRITO51393.2021.9596287.
[55] R. Xu, Y. Feng, H. Chen, *ChatGPT vs. Google: A Comparative Study of Search Performance and User Experience,* arXiv:2307.01135, **2023**.
[56] T. Wu, S. He, J. Liu, S. Sun, K. Liu, Q.-L. Han, Y. Tang, *IEEE/CAA J. Autom. Sin.* **2023**, *10*, 1122.
[57] N. F. Ayoub, Y.-J. Lee, D. Grimm, V. Divi, *Otolaryngol. Head Neck Surg.* **2023**.
[58] C.-H. Wei, A. A. Allot, P. T. Lai, R. Leaman, S. Tian, L. Luo, Z. Lu, *Nucl. Acids Res.* **2024**, gkae235.
[59] Q. Jin, R. Leaman, Z. Lu, *Ebiomed.* **2024**, *100*.

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[60] N. Sharma, Q. V. Liao, Z. Xiao. in *Proc. of the CHI Conf. on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY **2024**, pp. 1–17, Article 1033.
[61] S. A. Memon, J. D. West, Search Engines Post-ChatGPT: How Generative Artificial Intelligence Could Make Search Less Reliable, arXiv:2402.11707, **2024**.
[62] L. B. Eliot, *J. Bus. Ventur. Insights* **2023**, *19*, e00388.
[63] News, A. *Queensland to Join NSW in Banning Access to ChatGPT in State Schools,* **2023**, 2024, https://www.abc.net.au/news/2023-01-23/queensland-to-join-nsw-in-banning-access-to/101884288, (accessed: February 2024).
[64] C. Jaeger, AI Tool Banned in Victorian State Schools, **2023**.
[65] A. Klein, 180 Degree Turn: NYC District Goes from Banning ChatGPT to Exploring AI's Potential, in Education Week, **2023**, edweek.org, (accessed: January 2023).
[66] Overview of AI Policy Lab, https://www.nycdli.org/policylab#:~:text=This%20collaborative%20effort%20brings%20together,EDSAFE%20AI%20SAFE%20benchmarks%20framework (accessed: January 2024).
[67] M. W. T. Cheng, I. H. Y. Yim, *Discov. Educ.* **2024**, *3*, 1.
[68] R. H. Mogavi, C. Deng, J. J. Kim, P. Zhou, Y. D. Kwon, A. H. S. Metwally, A. Tlili, S. Bassanelli, A. Bucchiarone, S. Gujar, L. E. Nacke, P. Hui, *Comput. Human Behav. Artif. Humans* **2024**, *2*, 100027.
[69] R. H. Mogavi, C. Deng, J. J. Kim, P. Zhou, Y. D. Kwon, A. H. S. Metwally, A. Tlili, Exploring User Perspectives on ChatGPT: Applications, Perceptions, and Implications for AI-Integrated Education, arXiv:2305.13114, **2023**.
[70] C. Leiter, R. Zhang, Y. Chen, J. Belouadi, D. Larionov, V. Fresen, S. Eger, *Mach. Learn. Appl.* **2024**, *16*, 100541.
[71] A. J. G. Sison, M. T. Daza, R. Gozalo-Brizuela, E. C. Garrido-Merchán, *Int. J. Hum–Comput. Interact. 1*.
[72] P. Xiao, Y. Chen, W. Bao, Waiting, Banning, and Embracing: An Empirical Analysis of Adapting Policies for Generative AI in Higher Education, *ArXiv*, **2023**, abs/2305.18617.
[73] R. Browne, Italy Became the First Western Country to Ban ChatGPT. Here's What Other Countries are Doing, in CNBC, **2023**, CNBC.com.
[74] S. Onal, D. Kulavuz-Onal, *J. Educ. Technol. Syst.* **2024**, 00472395231196532301.
[75] E. Dickey, A. Bejarano, A Model for Integrating Generative AI into Course Content Development, arXiv:2308.12276, 2023.
[76] M.-L. Tsai, C. W. Ong, C.-L. Chen, *Educ. Chem. Eng.* **2023**, *44*, 71.
[77] A. Mousavi, M. Schmidt, V. Squires, K. Wilson, *Int. J. Artif. Intell. Educ.* **2021**, *31*: 603.
[78] A. Mizumoto, M. Eguchi, *Res. Meth. Appl. Linguist.* **2023**, *2*, 100050.
[79] T. K. Landauer, *Assess. Educ. Princ. Policy & Pract.* **2003**, *10*, 295.
[80] Z. Zhang, L. Xu, *J. Multiling. Multicult. Develop.*, **2022** *11*, 1.
[81] B. Stretton, J. Kovoor, M. Arnold, S. Bacchi, *Med. Sci. Educ.* **2024**, *34*, 215.
[82] F. Dolianiti, I. Tsoupouroglou, P. Antoniou, S. Konstantinidis, S. Anastasiades, P. Bamidis, in Brain Function Assessment in Learning: Second International Conference, BFAL 2020, Heraklion, Crete, Greece, October 9–11, 2020, Proceedings, Vol. *2*, Springer International Publishing, Cham **2020**, pp. 137–147.
[83] H. C. Lucas, J. S. Upperman, J. R. Robinson, in *Medical Education*, **2024**, pp. 1–10, http://dx.doi.org/10.1111/medu.15402.
[84] J. Qiu, L. Li, J. Sun, J. Peng, P. Shi, R. Zhang, Y. Dong, *IEEE J. Biomed. Health Inform.* **2023**, *27*, 6074.
[85] D. Q. Wang, L.-Y. Feng, J.-G. Ye, J.-G. Zou, Y.-F. Zheng, *MedComm–Future Med.* **2023**, *2*, e43.
[86] O. Temsah, S. A. Khan, Y. Chaiah, A. Senjab, K. Alhasan, A. Jamal, F. Aljamaan, K. H. Malki, R. Halwani, J. A. Al-Tawfiq, M. H. Temsah, A. Al-Eyadhy, *Cureus* **2023**, *15*, e37281.
[87] G.-J. Hwang, H. Xie, B. W. Wah, D. Gašević, *Comput. Educ. Artif. Intell.* **2020**, *1*, 100001.
[88] J. Biggs, D. Kember, D. Y. P. Leung, *Br. J. Educ. Psychol.* **2001**, *71*, 133.
[89] C. K. Y. Chan, W. Hu, *Int. J. Educ. Technol. High. Educ.* **2023**, *20*, 43.
[90] H. Crompton, D. Burke, *Int. J. Educ. Technol. High. Educ.* **2023**, *20*, 22.
[91] S. Gupta, Y. Chen, *J. Inform. Syst. Educ.* **2022**, *33*, 98.
[92] E. Mousavinasab, N. Zarifsanaiey, S. R. N. Kalhori, M. Rakhshan, L. Keikha, M. G. Saeedi, *Interact. Learn. Environ.* **2021**, *29*, 142.
[93] H. Lee, *Anat. Sci. Educ.* **2024**, *17*, 926.
[94] S. Hobert, in *Inter. Conf. on Interaction Sciences*, **2019**.
[95] S. A. Miller, W. Perrotti, D. U. Silverthorn, A. F. Dalley, K. E. Rarey, *Anat. Rec.* **2002**, *269*, 69.
[96] A. S. University, A New Collaboration with OpenAI Charts the Future of AI in Higher Education, **2024**, https://news.asu.edu/20240118-university-news-new-collaboration-openai-charts-future-ai-higher-education, (accessed: February 2024).
[97] T. Swaak, Arizona State and OpenAI Are Now Partners. What Does That Mean? **2024**, https://www.chronicle.com/article/arizona-state-and-openai-are-now-partners-what-does-that-mean, (accessed: February 2024).
[98] D. T. T. Mai, C. V. Da, N. V. Hanh, in *Frontiers In Education*, Frontiers Media SA **2024**.
[99] T. Goyal, J. J. Li, G. Durrett, News Summarization and Evaluation in the Era of GPT-3, *arXiv*, **2023**.
[100] T. Zhang, F. Ladhak, E. Durmus, P. Liang, K. McKeown, T. B. Hashimoto, Benchmarking Large Language Models for News Summarization, *arXiv*, **2023**.
[101] X. Yang, Y. Li, X. Zhang, H. Chen, W. Cheng, Exploring the Limits of ChatGPT for Query or Aspect-based Text Summarization, *arXiv*, **2023**.
[102] L. Tang, Z. Sun, B. Idnay, J. G. Nestor, A. Soroush, P. A. Elias, Z. Xu, Y. Ding, G. Durrett, J. F. Rousseau, C. Weng, Y. Peng, *NPJ Digit. Med.* **2023**, *6*, 158.
[103] C. Ma, Z. Wu, J. Wang, S. Xu, Y. Wei, Z. Liu, X. Jiang, L. Guo, X. Cai, S. Zhang, T. Zhang, D. Zhu, D. Shen, T. Liu, X. Li, ImpressionGPT: An Iterative Optimizing Framework for Radiology Report Summarization with ChatGPT, *arXiv*, **2023**.
[104] X. Wei, X. Cui, N. Cheng, X. Wang, X. Zhang, S. Huang, P. Xie, J. Xu, Y. Chen, M. Zhang, Y. Jiang, W. Han, Zero-Shot Information Extraction via Chatting with ChatGPT, **2023** *arXiv*: 2302.10205.
[105] S. Badini, S. Regondi, E. Frontoni, R. Pugliese, *Adv. Ind. Eng. Polym. Res.* **2023**, *6*, 278.
[106] N. Li, C. Gao, Y. Li, Q. Liao, Large Language Model-Empowered Agents for Simulating Macroeconomic Activities, *arXiv*, **2023**.
[107] V. Sorin, Y. Barash, E. Konen, E. Klang, *J. Cancer Res. Clin. Oncol.* **2023**, *149*, 9505.
[108] V. Sorin, E. Klang, M. Sklair-Levy, I. Cohen, D. B. Zippel, N. Balint Lahat, E. Konen, Y. Barash, *NPJ Breast Cancer* **2023**, *9*, 44.
[109] A. Cocci, M. Pezzoli, M. Lo Re, G. I. Russo, M. G. Asmundo, M. Fode, E. Durukan, *Prostate Cancer Prostat. Dis.* **2024**, *27*, 103.
[110] A. Lahat, E. Shachar, B. Avidan, Z. Shatz, B. S. Glicksberg, E. Klang, *Sci. Rep.* **2023**, *13*, 4164.
[111] Y. Wu, A. Henriksson, M. Duneld, J. Nouri, in *Responsive and Sustainable Educational Futures*, Springer Nature, Cham, Switzerland **2023**.
[112] P. K. Adhikary, S. Aseem, S. Kumar, S. M. Singh, P. Manuja, J. K. Gopinath, V. Krishnan, S. Kedia, K. S. Deb, T. Chakraborty, Exploring the Efficacy of Large Language Models in Summarizing Mental Health Counseling Sessions: A Benchmark Study, *ArXiv*, **2024**, abs/2402.19052.

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[113] D. V. Zandvoort, L. Wiersema, T. Huibers, S. v. Dulmen, S. Brinkkemper, in *Inter. Joint Conf. on Biomedical Engineering Systems and Technologies*, arXiv: 2311.13274, **2023**.
[114] M. Yadav, N. K. Sahu, M. Chaturvedi, S. Gupta, H. R. Lone, Fine-Tuning Large Language Models for Automated Diagnostic Screening Summaries, *ArXiv*, **2024**, abs/2403.20145.
[115] Y. Liu, S. Ju, J. Wang, *BMC Med. Inform. Decis. Making* **2024**, *24*, 75.
[116] H. K. Tiffany, M. Cheatham, A. Medenilla, C. Sillos, L. D. Leon, C. Elepaño, M. Madriaga, Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models, *medRxiv*, **2022**, p. 2022.12.19.22283643.
[117] G. G. P. Transformer, A. O. Thunström, S. Steingrimsson, Can GPT-3 write an Academic Paper on Itself, with Minimal Human Input? **2022**.
[118] ChatGPT Generative Pre-Trained Transformer 2, A. Zhavoronkov, *Oncoscience* **2022**, *9*, 82.
[119] C. Stokel-Walker, *Nature* **2023**, *613*, 620.
[120] T. McAdoo, How to Cite ChatGPT, **2024**.
[121] B. D. Lund, K. T. Naheem, *Can ChatGPT be an Author? A Study of Artificial Intelligence Authorship Policies in Top Academic Journals*, Learned Publishing **2024**, p. 13.
[122] H. Perlmutter, Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, **2023**, R.O.C.A.D.O.T.U.S.C. Office.
[123] Notice to Research Community: Use of Generative Artificial Intelligence Technology in the NSF Merit Review Process, **2023**, U.S.N.S Foundation.
[124] N. P. Partner, https://chat.openai.com/g/g-00Nu0FoNA-nsf-proposal-partner (accessed: June 2024).
[125] R. D. King, T. Scassa, S. Kramer, H. Kitano, *Nature* **2024**, *626*, 716.
[126] M. H. T. Ryan-Mosley, Z. Yang, *MIT Technol. Rev.* **2024**.
[127] N. Lucchi, *Eur. J. Risk Regul.* **2023**, *1*.
[128] T. Lancaster, *Int. J. Educ. Integr.* **2023**, *19*, 10.
[129] R. Tang, Y.-N. Chuang, X. Hu, The Science of Detecting LLM-Generated Texts, *ArXiv*, **2023**, abs/2303.07205.
[130] C. Wang, S. Liu, H. Yang, J. Guo, Y. Wu, J. Liu, *J. Med. Internet Res.* **2023**, *25*, e48009.
[131] K. Adhikari, N. Naik, B. Z. Hameed, S. K. Raghunath, B. K. Somani, et al., *Curr. Urol. Rep.* **2023**, *25*, 1.
[132] X. Wu, R. Duan, J. Ni, *J. Inf. Intell.* **2024**.
[133] M. Gupta, C. Akiri, K. Aryal, E. Parker, L. Praharaj, *IEEE Access* **2023**, *11*, 80218.
[134] S. Monteith, T. Glenn, J. R. Geddes, P. C. Whybrow, E. Achtyes, M. Bauer, *Br. J. Psychiatry* **2024**, *224*, 33.
[135] L. De Angelis, F. Baglivo, G. Arzilli, G. P. Privitera, P. Ferragina, A. E. Tozzi, C. Rizzo, *Front. Public Health* **2023**, *11*, 1166120.
[136] G. Deiana, G. Deiana, M. Dettori, A. Arghittu, A. Azara, G. Gabutti, P. Castiglia, *Vaccines* **2023**, *11*, 1217.
[137] S. C. Briand, M. Cinelli, T. Nguyen, R. Lewis, D. Prybylski, C. M. Valensise, V. Colizza, A. E. Tozzi, N. Perra, A. Baronchelli, M. Tizzoni, F. Zollo, A. Scala, T. Purnat, C. Czerniak, A. J. Kucharski, A. Tshangela, L. Zhou, W. Quattrociocchi, *Cell* **2021**, *184*, 6010.
[138] Y. Bang, S. Cahyawijaya, N. Lee, W. Dai, D. Su, B. Wilie, H. Lovenia, Z. Ji, T. Yu, W. Chung, Q. V. Do, Y. Xu, P. Fung, A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity, *arXiv*, **2023**.
[139] K. Lower, I. Seth, B. Lim, N. thSe, *Indian J. Orthop.* **2023**, *57*, 1527.
[140] S. T. Jürgen Rudolph, S. Tan, *J. Appl. Learn. Teach.* **2023**, *6*.
[141] J. G. Meyer, R. J. Urbanowicz, P. C. N. Martin, K. O'Connor, R. Li, P.-C. Peng, T. J. Bright, N. Tatonetti, K. J. Won, G. Gonzalez-Hernandez, J. H. Moore, *BioData Mining* **2023**, *16*, 20.
[142] A. M. Jarrah, Y. Wardat, P. Fidalgo, *Online J. Commun. Media Technol.* **2023**, *13*, e202346.
[143] M. R. King, *Cell. Mol. Bioeng.* **2023**, *16*, 95.
[144] N. Gross, *Soc. Sci.* **2023**, *12*, 435.
[145] Y. Chen, M. Andiappan, T. Jenkin, A. Ovchinnikov, in *A Manager and an AI Walk into a Bar: Does ChatGPT Make Biased Decisions Like We Do?* SSRN **2023**, 4380365.
[146] T. Heston, C. Khun, C. Khun, *Int. Med. Educ.* **2023**, *2*, 198.
[147] L. Giray, *Ann. Biomed. Eng.* **2023**, *51*, 2629.
[148] META, *LLaMa*, https://llama.meta.com/ (accessed: June 2024).
[149] AI, A.I.f. *OLMo*, https://allenai.org/olmo (June 2024).
[150] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière et al., *LLaMA: Open and Efficient Foundation Language Models*. *ArXiv*, **2023**, abs/2302.13971.
[151] D. Groeneveld, I. Beltagy, P. Walsh, A. Bhagia, R. Kinney, O. Tafjord, A. H. Jha, *Olmo: Accelerating the Science of Language Models*, arXiv:2402.00838, **2024**.
[152] Y. Huang, S. Gupta, M. Xia, K. Li, D. Chen, *Catastrophic Jailbreak of Open-Source LLMS Via Exploiting Generation*, arXiv:2310.06987, **2023**.
[153] A. Havrilla, M. Iyer, *Understanding the Effect of Noise in LLM Training Data with Algorithmic Chains of Thought*, arXiv:2402.04004, **2024**.
[154] V. Raina, A. Liusie, M. Gales, *Is LLM-as-a-Judge Robust? Investigating Universal Adversarial Attacks on Zero-shot LLM Assessment,* arXiv:2402.14016, **2024**.
[155] B. Zhang, Z. Liu, C. Cherry, O. Firat, *When Scaling Meets LLM Finetuning: The Effect of Data*, *Model and Finetuning Method*, arXiv:2402.17193, **2024**.
[156] H. Huang, Y. Qu, J. Liu, M. Yang, T. Zhao, *An Empirical Study of LLM-as-a-Judge for LLM Evaluation: Fine-tuned Judge Models are Task-specific Classifiers*, arXiv:2403.02839, **2024**.
[157] I. Bouzenia, P. Devanbu, M. Pradel, *RepairAgent: An Autonomous, LLM-Based Agent for Program Repair,* arXiv:2403.17134, **2024**.
[158] S. Shekhar, T. Dubey, K. Mukherjee, A. Saxena, A. Tyagi, N. Kotla, *Towards Optimizing the Costs of LLM Usage*, arXiv:2402.01742, **2024**.
[159] L. Kumar, D. K. Singh, *Multimed. Tools Appl.* **2023**, *82*, 40585.
[160] C. Wu, S. Yin, W. Qi, X. Wang, Z. Tang, N. Duan, Visual ChatGPT: Talking, Drawing and Editing with Visual Foundation Models, *arXiv*, **2023**.
[161] K. Armanious, C. Jiang, M. Fischer, T. Küstner, T. Hepp, K. Nikolaou, S. Gatidis, B. Yang, *Comput. Med. Imaging Graph.* **2020**, *79*, 101684.
[162] *Google Gemini*, **2023**, https://deepmind.google/technologies/gemini/#introduction (accessed: December 2023).
[163] IV, A.P. *Google's Gemini Controversy Explained: AI Model Criticized By Musk and Others Over Alleged Bias*, **2024**, https://www.forbes.com/sites/antoniopequenoiv/2024/02/26/googles-gemini-controversy-explained-ai-model-criticized-by-musk-and-others-over-alleged-bias/?sh=6882b93a4b99 (accessed: February 2024).
[164] S. Shamim, Why Google's AI Tool was Slammed for Showing Images of People of Colour, https://www.aljazeera.com/news/2024/3/9/why-google-gemini-wont-show-you-white-people#:~:text=The%20AI%2Dgenerated%20images%20of,democracy%20protests%20in%20Hong%20Kong (accessed: March 2024).
[165] N. Robins-Early, Google Restricts AI Chatbot Gemini from Answering Questions on 2024 Elections, https://www.theguardian.com/us-news/2024/mar/12/google-ai-gemini-2024-election (accessed: March 2024).
[166] K. Puteikis, R. Mameniškienė, *Epilepsy Behav.* **2024**, *156*, 109822.
[167] M.-H. Temsah, A. N. Alhuzaimi, M. Almansour, F. Aljamaan, K. Alhasan, M. A. Batarfi, I. Altamimi, A. Alharbi, A. A. Alsuhaibani, L. Alwakeel, A. A. Alzahrani, K. B. Alsulaim, A. Jamal, A. Khayat, M. H. Alghamdi, R. Halwani, M. K. Khan, A. Al-Eyadhy, R. Nazer, *J. Med. Syst.* **2024**, *48*, 54.

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[168] M. T. Birtek, M. M. Alseed, M. R. Sarabi, A. Ahmadpour, A. K. Yetisen, S. Tasoglu, *Biomicrofluidics* **2023**, *17*, 044101.
[169] D. Lopatto, A. G. Rosenwald, R. C. Burgess, C. Silver Key, M. Van Stry, M. Wawersik, J. R. Diangelo, A. T. Hark, M. Skerritt, A. K. Allen, C. Alvarez, S. Anderson, C. Arrigo, A. Arsham, D. Barnard, J. E. J. Bedard, I. Bose, J. M. Braverman, M. G. Burg, P. Croonquist, C. Du, S. Dubowsky, H. Eisler, M. A. Escobar, M. Foulk, T. Giarla, R. L. Glaser, A. L. Goodman, Y. Gosser, A. Haberman, et al., *J. Microbiol. Biol. Educ.* **2022**, *23*, e00208-21.
[170] D. Lopatto, The Essential Features of Undergraduate Research, Vol. *24*, Council on Undergraduate Research Quarterly, **2003**, pp. 139–142.
[171] D. Lopatto, *Cell Biol. Educ.* **2004**, *3*, 270.
[172] M. Al-Emran, *Interact. Learn. Environ.* 1.
[173] B. Botchu, K. P. Iyengar, R. Botchu, *Disabil. Rehabil. Assist. Technol.* **2023**, *19*, 2131.
[174] D. Choi, S. Lee, S.-I. Kim, K. Lee, H. J. Yoo, S. Lee, H. Hong, *Unlock Life with a Chat(GPT): Integrating Conversational AI with Large Language Models into Everyday Lives of Autistic Individuals. In Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, Association for Computing Machinery, New York, NY Article 72, pp. 1–17, https://doi.org/10.1145/3613904.3641989.
[175] G. Sung, L. Guillain, B. Schneider, in *Proc. of the 17th International Conf. of the Learning Sciences-ICLS 2023*, International Society of the Learning Sciences, Buffalo, NY **2023**, pp. 2093–2094.
[176] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language Models are Few-Shot Learners, *arXiv*, **2020**.



Defne Yigci is a medical student at Koç University, School of Medicine. She obtained her B.Sc. in molecular environmental biology (2021) from UC Berkeley. Her research interests include CRISPR-based diagnostics and therapeutics, utilization of AI in healthcare and medicine, and continuous health monitoring tools for early disease diagnosis.



Savas Tasoglu is currently an associate professor of Koc and (adjunct) Bogazici University. Dr. Tasoglu received his Ph.D. from UC Berkeley and held a postdoctoral appointment at Harvard Medical School. Dr. Tasoglu holds six patents and has published 100+ articles in journals such as Nature Materials, Nature Communications, Nature Reviews Urology, Advanced Materials, PNAS, Small, ACS Nano. His work was highlighted in Nature, Nature Physics, Nature Medicine, Boston Globe, Reuters Health, and Boston Magazine.



Aydogan Ozcan is the chancellor's professor and the Volgenau Chair for Engineering Innovation at UCLA and an HHMI professor with the Howard Hughes Medical Institute. He is also the associate director of the California NanoSystems Institute. Dr. Ozcan is elected Fellow of the National Academy of Inventors and holds >75 issued/granted patents, and is also the author of one book and the co-author of >1000 peer-reviewed publications in leading scientific journals/conferences. Dr. Ozcan is elected Fellow of Optica, AAAS, SPIE, IEEE, AIMBE, RSC, APS and the Guggenheim Foundation, and is a Lifetime Fellow Member of Optica, NAI, AAAS, SPIE, and APS.

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH

26404567, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aisy.202400429 by Koc University, Wiley Online Library on [30/12/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License