A learning module for generative AI literacy in a biomedical engineering classroom

Xianglong Wang*, Tiffany Marie Chan and Angelika Aldea Tamura

Department of Biomedical Engineering, University of California, Davis, Davis, CA, United States

Purpose: Generative artificial intelligence (GenAI), especially Large Language Model (LLM)-based chatbots such as ChatGPT, has reshaped students' learning and engagement in higher education. Yet, technical details of GenAI are largely inapproachable to most students. This article develops a learning module for GenAI and seeks to examine whether this module can potentially affect students' perceptions toward GenAI.

Methods: We implemented a one-lecture (60-min) module on GenAI models, with primary focus on structures of LLM-based chatbots, during the last week of a Biomedical Engineering (BME) Machine Learning course. A mixed-methods survey on perceptions of GenAI was distributed to the students before and after the module. Paired t-tests and regression analyses were used to analyze the Likert-scale quantitative questions and thematic coding was performed for the free-response questions.

Results: Students ($N = 13$) reported significantly stronger approval on favorability to use GenAI in medicine ($p = 0.015$), understanding of LLM-based chatbots ($p < 0.001$), confidence on using LLM-based chatbots ($p = 0.027$), optimism on future development of LLMs ($p = 0.020$), and perception of instructor's attitude toward GenAI ($p = 0.033$). Students maintained a neutral view on accuracy of LLM-generated answers and a negative view on the ability of generating bias-free answers in LLMs. The primary contributors identified in students' intentions to use LLMs are self-efficacy in using the LLM outputs and lower precepted bias of LLMs. The impression of GenAI for students shifted from primarily LLM-based chatbots and generative work to components and training process of GenAI. After the module, students reported a clear understanding of tokenizers and word embeddings while expressing confusion on transformers.

Conclusion: A module on the details of GenAI models shifted the students' attitudes to GenAI models positively while still being acutely aware of its limitations. We believe that inclusion of such modules in a modern engineering curriculum will help students achieve AI literacy.

1 Introduction

Generative artificial intelligence (GenAI) is often referred to as machine learning models that produce new information based on the training data (García-Peñalvo and Vázquez-Ingelmo, 2023). Despite the widespread attention received by GenAI products, such as large language model (LLM)-based chatbots including ChatGPT (GPT: Generative Pre-trained Transformer), the field with the most GenAI papers published in the past 3 years is medicine (García-Peñalvo and Vázquez-Ingelmo, 2023). In the aspect of education, GenAI products, especially LLM-based chatbots, have impressed students with their technical prowess and high accessibility (Yilmaz et al., 2023). These chatbots have been rapidly adopted by both students and educators alike. Estimates of actual adoption rates of these chatbots within the students vary, ranging from 24.6% (Abdaljaleel et al., 2024) in a survey conducted among undergraduates in multiple Asian countries, to a reported 58.2% among graduate students within a U.S. medical school (Hosseini et al., 2023), with varying numbers in between (Singh et al., 2023; Vest et al., 2024). Faculties have also reported using LLM-based chatbots in translating materials across languages (Kiryakova and Angelova, 2023), preparing lecture materials (Kiryakova and Angelova, 2023), generating assessments (Farrokhnia et al., 2023), and summarizing communication (White et al., 2024). Therefore, modern educators must pay special attention to the capabilities and limitations of GenAI products, while being acutely aware of their adoptions in classroom settings.

With more students think that using LLM-based chatbots is acceptable for coursework (White et al., 2024), especially for a specific subset of tasks (Vest et al., 2024), these chatbots will likely become an integral part of modern college programs, especially engineering programs. However, GenAI products, especially these LLM-based chatbots, differentiate themselves from other common engineering tools or office software, in ways that the performance characteristics of these chatbots are difficult to interpret and evaluate for non-machine learning (ML) experts (Singh et al., 2024). Most of the users of such LLM-based chatbots in current students, unfortunately, would classify as non-ML experts. OpenAI's website of their GPT models shows the performance of GPT on a series of text, video, and audio benchmarks. However, rarely do the users of the chatbots know what MMLU (Hendrycks et al., 2020), a prominent text-based benchmark that most modern LLMs get evaluated on, contains to make an accurate sense of the score on the MMLU benchmark. Neither do most users know the training data, the theoretical framework, or the structure of LLMs, making the nature of GenAI-based products inapproachable and incomprehensible.

The current education system is significantly challenged by these unknowns. The propagation of an unknown level of bias from the training data into GenAI products can pose ethical risks and perpetuate bias in education (Tlili et al., 2023). The high barriers to understanding the evaluation and components LLM-based chatbots contribute to the difficulty in evaluating the quality of responses generated by these chatbots (Ferrara, 2024), which can result in a sense of blind trust among the users (Jung et al., 2024) and potentially lead to degradation in students' high-order cognitive skills (Farrokhnia et al., 2023). These unknowns also discourage the users from taking responsibility for their actions in using these chatbots (Venkatesh, 2022) and thus, may encourage irresponsible behavior in learning, such as plagiarism and cheating (Farrokhnia et al., 2023), which in turn threatens academic integrity.

The goal of this paper is to investigate whether dispelling these unknowns by arming our BME students with knowledge of GenAI will affect the students' perceptions toward GenAI, especially toward the LLM-based chatbots. To achieve this goal, we designed a 60-min learning module on GenAI with a focus on construction of LLMs. We designed a 14-item survey from relevant theoretical frameworks for technology adoption to systematically investigate students' perceptions toward GenAI and LLMs. Through the survey, we characterized the effectiveness of this learning module and evaluated the most significant contributors to students' intention of using these chatbots. We intend to develop refined and tailored versions of our current learning module to fit various educators' needs.

2 Pedagogical framework and learning environment

The intervention, a 60-min lecture on GenAI models, was implemented as the last module of the "Machine Learning for Biomedical Engineering" technical elective course in the Spring 2024 term. Therefore, participating students tended to possess a high interest in machine learning and were knowledgeable in traditional machine learning methods. However, since full understanding of LLM requires knowledge in natural language processing (NLP) and deep learning, which were not covered in the course, we used the Cognitive Load Theory to guide the development of the learning module to achieve the best learning outcome when our students were not fully ready to tackle the material head-on.

The Cognitive Load Theory (Sweller, 2011) specifies that the extents of learning is affected by the intrinsic load of the material, which is the complexity of the knowledge presented. Even with the background and preparation level of our attending students, the intrinsic load of understanding LLM is extremely high. To reduce the intrinsic load, we designed the learning module which isolated the building blocks of a LLM model into its main building blocks, including tokenizers, word embeddings, and transformers. We also aimed to introduce more variability and promote interactivity by integrating discussion-based exercises after dense introductions of the concepts. A worked example was shown during the introduction of tokenizers to ease the transition to understanding difficult subjects. The 60-min lecture was structured as follows:

1 Introduction to flow of natural language processing (NLP) models and general structure of LLMs, assuming textual prompts and textual generation: *tokenizer* to *word embeddings* to *transformers* to *inverse word embeddings* and *inverse tokenization*.
2 Explaining the role of tokenizer, which converts sentences into a series of lexicographic tokens (in this case, an array of numbers). Students were reminded about the necessity of this step because computers can only understand numbers and not text. This section included a case study in Byte-Pair Encoding (Gage, 1994), the tokenizer adopted by GPT-series models, including GPT-4 (Berglund and van der Merwe, 2023; Hayase et al., 2024). A live demonstration of GPT-4o's tokenizer was

shown on the screen using the *tiktoken* Python package by OpenAI.

3 Introducing word embeddings as the way to project the word tokens into a lower-dimensional vector space with dimensions focusing on the meaning of the words instead of the words themselves. Students were first shown the size of the dictionaries used in GPT-4o's tokenizer, which includes 524,288 different words. Then, students were taught that the word embeddings used by GPT-4 can compress 524,288 dimensions into just 3,072 dimensions, demonstrating great savings in both time and space. The case study was an introduction to Word2Vec (Mikolov et al., 2013), a two-layer neural network-based approach to word embeddings. Students were informed that the GPT-4 uses a proprietary word embeddings model that is more complicated than Word2Vec.

4 Introducing transformers at a very high level. Transformers are neural networks that transform the input word embeddings (the processed prompt) into output word embeddings (the answer in numerical format). The transformation is made possible by the transformers learning about the statistical distributions of the training data. The case study was the network structure of the original transformer network (Vaswani, 2017), which closely resembles to the structure of the transformer in GPT-1 (Radford et al., 2018). The network structure was introduced at a block-diagram level without going into the details.

5 General training procedure of transformer networks, including estimates in size of network and data source, time, and monetary cost in training the transformer of GPT-4.

We assessed and identified relevant dimensions within the Technology Acceptance Model (TAM) (Davis et al., 1989), the Task-Technology Fit (TTF) Model (Goodhue and Thompson, 1995), and the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003, 2012) Models that were applicable toward using GenAI, especially those areas that have the potential to be impacted. These models were developed to explain adoption of emerging technology. The dimensions we identified as relevant include behavioral intention, attitude, performance expectancy/perceived usefulness, individual characteristics, and social influence. Behavioral intention and attitude were first identified as relevant due to the goal of the study, which is whether the perception of GenAI within the students will be changed due to this intervention.

We expect self-efficacy levels of using GenAI tools to increase after the intervention and be a potential positive contributor to students' intention of using GenAI. Although, we hypothesize that learning more about the components of LLMs may affect students' view of the performance expectancy of GenAI/LLM in conflicting ways. Knowing the components of how GenAI products are made can potentially enhance the interpretability of the contents generated by GenAI; however, the lesson plans also contained discussions on potential biases that GenAI could exhibit, which could cause students to trust GenAI less. Due to the transparency of the construction and evaluation of GenAI systems that this instruction module potentially brings to the students, we added self-efficacy and personality within the dimension of individual characteristics to our framework of examining GenAI. Additionally, we considered social influence to be a potential contributor to students' intention of using GenAI: the instructor of the module could potentially exhibit "advocacy bias" (Ellsworth, 2021) and thus affect the students' interest level or attitude toward GenAI.

Therefore, our research questions associated with this learning module are:

> *RQ1.* Is this module effective in developing understanding of GenAI systems within the participating students?
>
> *RQ2.* Is a better understanding of the inner workings of GenAI systems correlated with better self-efficacy of using GenAI products?
>
> *RQ3.* In which direction will students' perceptions toward GenAI products change when students have a better understanding of the construction of GenAI systems?
>
> *RQ4.* Will students recognize the instructor as a GenAI advocate if an instructor teaches a GenAI module in their course, irrespective of the instructor's stance of GenAI?

3 Evaluation methods

The class period was 80 min. At the beginning and the end of the class (10 min each), students were asked to complete a survey. The quantitative portion of the survey contained 11 statements based on levels of agreement. We chose a 6-point Likert scale (1: strongly disagree; 6: strongly agree) for better normality in the data (Leung, 2011) and having the participants take a position (Croasmun and Ostrom, 2011) so that we could better understand students' positionality on GenAI. Many statements were formulated to focus on LLMs due to the contents of the lecture. The 11 statements were based on areas of theoretical frameworks that we identified in the previous section: behavioral intention (Statement 5 or S5), performance expectancy (S7, 8, 10), attitude (S1, 2, 4), self-efficacy (S3, 6), optimism (S9) and social influence (S11). The wording of the survey questions can be seen in Table 1. To qualitatively assess the perception toward GenAI before and after the lecture, we included one additional open-ended question, "When you think of generative AI, what terms come to mind?" The post-survey also included two additional questions asking about the clearest points and the muddiest points from the lecture to evaluate and refine this lecture.

To ensure that the participants of the survey were actual participants of the intervention, the survey was distributed in person. Participants were informed that no grade bonus or penalty is associated with completing the survey, and they should not put their names on the survey. Instead, a number identifier was included in the surveys to link the pre- and post-surveys to a single participant. Since the machine learning course is a technical elective with a relatively small enrollment, adequate measures of ensuring anonymity were taken. The instructor left the room after the surveys were distributed and completed surveys were placed at the instructor podium facing down. For students who arrived late after the pre-survey had been submitted, they were instructed not to complete either survey but were allowed to attend the lecture. To further ensure anonymity,

TABLE 1 Average agreement levels with statements 1–11 on a 6-point Likert scale (1: strongly disagree; 6: strongly agree) in the pre-survey (Pre) and post-survey (Post).

Statements	N	Pre (/6.00)	Post (/6.00)	p-value	Sig.
1	12	4.167	4.583	0.1753	NS
2	10	4.400	4.900	0.0150	*
3	13	2.769	4.615	0.0001	***
4	13	3.923	4.231	0.2188	NS
5	13	4.538	4.923	0.0961	NS
6	13	4.154	4.615	0.0269	*
7	13	2.615	3.231	0.0712	NS
8	13	1.923	1.846	0.6727	NS
9	13	4.154	5.000	0.0205	*
10	13	3.692	3.769	0.8193	NS
11	8	4.375	4.875	0.0331	*

The number of participants who answered the question in both the pre- and the post-survey is denoted by N. For significance (sig.), $*p < 0.05$, $***p < 0.001$.

Statements list:

1. I am favorable toward the use of natural language processing models in medicine.
2. I am favorable toward the use of generative adversarial networks in medicine.
3. I am certain that I understand the building blocks of large language model (LLM)-based chatbots, such as ChatGPT.
4. I am favorable toward the use of LLM-based chatbots in medicine.
5. I am likely to use LLM-based chatbots in my study or line of work.
6. I am certain that I can use the outputs from LLM-based chatbots effectively.
7. I think that LLM-based chatbots generate highly accurate answers.
8. I think that LLM-based chatbots generate bias-free answers.
9. I am optimistic about the direction of development in future LLM models.
10. I think that LLM-based models have potential to replace humans in daily tasks.
11. The instructor is favorable toward the use of LLMs.

transcriptions of the survey results were performed by the non-instructor authors on this paper. This study was designated as Non-Human Subject Research by UC Davis IRB office (IRB #2209830-1).

4 Results

4.1 Quantitative results

Thirteen copies of pre- and post-survey were collected at the end of the lecture. Due to the nature of surveys being distributed via paper copies and the full anonymity, not all questions were completed by the students. Student demographics were not collected as part of this study; however, the overall makeup of the course is 52% female and predominantly BME senior undergraduates.

On the quantitative portion of the survey, students reported a perceived favorable attitude from the instructor toward LLMs, and the perception was significantly reinforced after the lecture (S11, $p = 0.033$). Students lean toward agreeing on statements 1, 2, 4, 5, 6, 9 before the lecture was given to them; after receiving the lecture, the levels of agreement on S2 (use of GANs in medicine, $p = 0.015$), S6 (self-efficacy in using chatbots, $p = 0.027$), and S9 (optimism on future of LLMs, $p = 0.020$) significantly increased. The positionalities of S3, 7, 10 were not clear in the pre-survey. Among these statements, students reported a major increase in understanding of the building blocks of LLM-based chatbots (S3, $p < 0.001$), demonstrating the

efficacy of the lecture. S8 (LLMs are bias-free) received a low level of agreement in the pre-survey, and this agreement level stayed low in the post-survey. Table 1 shows the full analysis of the Likert-scale questions and the full wording of these statements.

The results from the linear regression with RFE are presented in Table 2. The final model has only four predictors but achieved an excellent fit ($R^2 = 0.70$). Significant predictors on students' intention of using GenAI (S5) include their self-efficacy on using the LLM outputs (S6, slope = 1.68, $p = 0.001$) and their perception of LLM being bias-free (S8, slope = 0.78, $p = 0.016$). The favorability of using GenAI in medicine (S2, negative) and understanding of building blocks (S3, positive) are also contributors to this model, but these contributions are not statistically significant. The addition of any other statement into the model will cause the adjusted R^2 value to decrease, so we consider all other statements (S1, 4, 7, 9, 10, 11) to be non-contributors to students' intention of using GenAI.

4.2 Qualitative results

The final codebook for the free-response survey questions contains these major codes: components and training of LLMs, names of generative AI models, machine learning methods, generated data, medical AI, tool, ethics, and AI devices. The Cohen's Kappa for the coding was 0.840, demonstrating strong agreement between the two coders (McHugh, 2012).

We performed a Fisher's exact test on the coding frequencies in the common free-response question, "When you think of generative AI, what terms come to mind" (see Table 3). We found that the frequencies of the codes were significantly different ($p = 0.002$), further reaffirming our findings in the quantitative portion of the survey, that the students reported a significant increase in confidence in understanding of LLMs. Students have shifted from regarding GenAI as solely the names of GenAI products, such as ChatGPT and Dall-E, to components and training of LLMs. Another shift that we observed in the coding frequency is within the potential products of GenAI: more domain-specific codes in BME in "medical data processing" and "medical images" were identified instead of generic images, drawings, and letters.

We also applied this codebook to the other two questions in the post-survey, the clearest and muddiest points associated with the lecture. Students' answers to both questions, unsurprisingly, coded primarily into the category "Components and Training of LLMs." A more detailed analysis revealed that the students regarded the *tokenizers* as the clearest point, followed by *word embeddings* and *cost associated with training*; the structure and training of transformers remain the muddiest point for most students.

5 Discussion

In this study, we found that instructing the students about LLMs can shift students' perception of GenAI from naming the LLM-based chatbots to understanding the components of the products (S3, Table 3, codes 1, 2) as well as from general to domain-specific applications (S2, Table 3, codes 4, 5). These findings suggest that instructing students about GenAI, especially in a domain-specific context, may be beneficial for students to develop context for GenAI methods and products in their domain and

develop students' critical thinking levels. The instruction module also builds self-efficacy (S6) toward the usage and development of LLM-based chatbots in students. Overall, the learning module successfully fulfilled the role of bringing more clarity and interpretability for understanding and evaluating GenAI, especially LLM-based chatbots.

RQ2 was not fully supported from our initial cohort. The level of students' perception of adopting GenAI in their study or work (S5) received a near-significant increase. According to our linear regression model, the main contributor from the intervention toward students' tendency to adopt may be from a higher level of self-efficacy of using the *outputs* from LLMs (S6, total effect $0.461 \times 1.68 = 0.774$). The contribution from increased understanding of LLM components (S3) was present but much less effective (total effect $1.846 \times 0.17 = 0.314$). This finding suggests that if an educator's goal is to increase adoption rates of LLM-based chatbots in their classrooms and/or increase students' levels of GenAI literacy, lectures focusing on using the LLMs, for example, prompt engineering and/or evaluating the outputs from LLM-based chatbots, may be more effective than teaching the students about constructing LLM-based chatbots.

We were only able to partially validate RQ3 in our initial offering of the module. Although we did find a significant increase in S9 (optimism in using and developing LLM-based chatbots) and a near-significant increase in S5 (adopting GenAI), the observed effects were mostly from an increased self-efficacy shown in RQ2. The authors have originally hypothesized that more knowledge about the

components of LLMs will cause a decrease in S8, whether LLM models are regarded as bias-free. However, a prior module of this course has covered bias and equity issues in machine learning. Within the module, the study of word embeddings was used as an example for machine learning systems that exhibit bias. Possibly due to prior knowledge resulting from this prior module, students reported very low levels of agreement on S8 in the pre-survey. Therefore, we could not examine the effect in the awareness of bias level in GenAI in this cohort due to the pre-existing consensus. However, the regression model depicted that students who had a more optimistic view on bias and equity of LLMs tend to have a higher tendency of using LLMs in their study or work, partially confirming our initial hypothesis that better knowledge in biases exhibited in LLMs could potentially lead to a lower tendency of use. The previously proposed future work of developing a learning module for general students could potentially help us achieve better understanding in this RQ.

Although S11 (students' perception of instructor's attitude toward LLMs) was deemed a non-contributor to students' adoption of LLM-based chatbots, a lecture on constructing LLMs nonetheless increased an already-high level of perception that the instructor is favorable toward LLMs. From a post-lecture discussion with the students, the students were very surprised to know that the instructor is a complete non-user of LLM-based chatbots; the perceived favorability may have resulted from the identities of the instructor, i.e., a biomedical engineer teaching the course *machine learning in BME*, who *included a module of introductions to GenAI/LLM* in the syllabus and have multiple publications about *AI work in medicine*, including *generative AI work*. Therefore, the authors suggest that potential adopters of such modules in their own classrooms, whether teaching about components of LLMs or about using LLM-based chatbots, to be aware of students' perceptions about potential identities of the instructor, which can possibly affect the outcome of classroom instruction.

6 Limitations and future work

Our implementation of the one-lecture module promoted GenAI literacy in our machine learning students. However, we would like to

TABLE 2 Coefficients of the final linear regression model after RFE to predict S5 (behavioral intention).

Variable	Value	Standard error	Beta	*p*-value	Sig.
Intercept	−2.90	1.96	0.00	0.163	NS
S2	−0.35	0.25	−0.24	0.184	NS
S3	0.17	0.16	0.19	0.318	NS
S6	1.68	0.39	0.83	0.001	**
S8	0.78	0.28	0.49	0.016	*

The "Value" field is the value of the intercept and the slopes of all other statements. Value and Standard Error are unstandardized coefficients; Beta is the standardized slope. *$p < 0.05$; **$p < 0.01$.

TABLE 3 Codes, frequency (as measured by the number of references), and sample quotes in the question "when you think of generative AI, what terms come to mind" in both the pre- and the post-surveys.

Code	Pre-survey frequency	Post-survey frequency	Sample quotes
Components and training of LLMs	0	11	"Tokenizing, embedding, transforming" "…how complex and the amount of money put in to create these models"
Names of GenAI models	7	4	"ChatGPT," "DALL-E," "Google Gemini"
Machine learning methods	5	4	"Natural language processing," "neural networks," "machine learning," "deep learning"
Generative data	6	3	"AI-generated image," "create drawings or images," "form letters"
Medical AI	0	4	"…medical data processing," "medical images"
Tool	3.5	0	"…a tool that can assist us," "a useful tool"
Ethics	1	1	"…stolen work," "unapproved use of established works"
AI devices	1	0	"Robots"

Code frequency presented is the average of the two raters.

caution the readers on the results we obtained so far: participants of this study have almost completed a whole machine learning course, including modules on data exploration, visualization, linear regression, logistic regression, support vector machines, trees, fully connected neural networks, and clustering. These students are generally committed to learning more about AI and were receptive to knowledge related to GenAI. The cohort of students participating in the current study (maximum $N = 13$) is relatively small; more offerings of this course may be needed to increase the quality of statistics performed in this study.

Potential adopters of our strategies should mind students' level of background knowledge in machine learning and AI and should consider adjusting the complexity of the offering and/or increase the time allocated to this module for maximum benefits. With our students' preparation level, the topics that students have received adequate preparations for, such as the tokenizer and the word embeddings, were identified as the clearest points in the GenAI module. This course did not prepare the students to understand deep learning topics such as convolutional neural networks, and thematic analysis revealed that the details of deep learning-based transformers were too challenging for students to understand, even when introduced at the surface level.

One other future-facing challenge is the increasing opacity of GenAI products, especially LLM-based chatbots. The commercialization of LLM-based chatbots, now sometimes including audio, image, and/or video processing and generation capabilities, has shifted the scope of GenAI space. The training data, processes of word embeddings, the structure of the transformer network, and cost/time to train these chatbots, are no longer disclosed by commercial GenAI companies in their technical papers. The construction of this teaching module had to rely on best estimates in computer science literature and data from past models. We expect that our ability to update this module for adapting to future state-of-the-art GenAI products will be significantly challenged unless the companies become more transparent about the details of their GenAI products.

We intend to improve the module for the machine learning course: although understanding the neural network structures of transformers will be extremely challenging for students who are taking their first machine learning course, a more thorough introduction to deep learning methods will be beneficial in helping students understand important concepts such as layers, kernels/filters, and parameters. We also plan to develop two more instructional modules on GenAI. A technical module that assumes less background knowledge may be beneficial for easier adoption for interested instructors to develop their students' GenAI literacy and can be used as training for faculty to become more aware of GenAI. We also plan to develop a non-technical module, in collaboration with experienced LLM-based chatbot users, to increase participants' skills in prompting and evaluating the output of LLM-based chatbots. We intend to evaluate the outcomes of these modules with a more comprehensive survey among participants of these new modules.

Gamification has been reported to enhance students' engagement in class (Gari et al., 2018) and promote collaborative reasoning (Di Nardo et al., 2024) in a lecture-based context for assessment (Alhammad and Moreno, 2018). Our current machine learning course has integrated some major gamification components in instruction and assessment, for example, students are graded based on their placement on a leaderboard for the projects, which were private machine learning competitions. A way to address the absence of a formal assessment for the module in this course may be designing and implementing a bonus credit activity as an in-class GenAI trivia based on the materials of the module.

Data availability statement

The current version of the learning module can be accessed at https://cube3.engineering.ucdavis.edu.

Ethics statement

The studies involving humans were approved by University of California, Davis, Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because the study was designated as Non-Human Subject Research by the IRB Office due to the proposed activities not meeting the definition of human research (IRB 2219830-1).

Author contributions

XW: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. TC: Data curation, Formal analysis, Writing – original draft, Writing – review & editing. AT: Data curation, Formal analysis, Writing – original draft, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdaljaleel, M., Barakat, M., Alsanafi, M., Salim, N. A., Abazid, H., Malaeb, D., et al. (2024). A multinational study on the factors influencing university students' attitudes and usage of Chatgpt. *Sci. Rep.* 14:1983. doi: 10.1038/s41598-024-52549-8

Alhammad, M. M., and Moreno, A. M. (2018). Gamification in software engineering education: a systematic mapping. *J. Syst. Softw.* 141, 131–150. doi: 10.1016/j.jss.2018.03.065

Berglund, M., and Van Der Merwe, B. (2023). Formalizing Bpe Tokenization. *arXiv:2309.08715*. doi: 10.48550/arXiv.2309.08715

Croasmun, J. T., and Ostrom, L. (2011). Using likert-type scales in the social sciences. *J. Adult Educ.* 40, 19–22.

Davis, F. D., Bagozzi, R., and Warshaw, P. (1989). Technology acceptance model. *J. Manag. Sci.* 35, 982–1003.

Di Nardo, V., Fino, R., Fiore, M., Mignogna, G., Mongiello, M., and Simeone, G. (2024). Usage of gamification techniques in software engineering education and training: a systematic review. *Computers* 13:196. doi: 10.3390/computers13080196

Ellsworth, P. C. (2021). Truth and advocacy: reducing bias in policy-related research. *Perspect. Psychol. Sci.* 16, 1226–1241. doi: 10.1177/1745691620959832

Farrokhnia, M., Banihashem, S. K., Noroozi, O., and Wals, A. (2023). A Swot analysis of Chatgpt: implications for educational practice and research. *Innov. Educ. Teach. Int.* 61, 460–474. doi: 10.1080/14703297.2023.2195846

Ferrara, E. (2024). Genai against humanity: nefarious applications of generative artificial intelligence and large language models. *J. Comput. Soc. Sci.* 7, 549–569. doi: 10.1007/s42001-024-00250-1

Gage, P. (1994). A new algorithm for data compression. *C Users J.* 12, 23–38.

García-Peñalvo, F., and Vázquez-Ingelmo, A. (2023). What do we mean by Genai? A systematic mapping of the evolution, trends, and techniques involved in generative AI. *Int. J. Interact. Multimedia Artif. Intell.* 8, 7–16. doi: 10.9781/ijimai.2023.07.006

Gari, M. R. N., Walia, G. S., and Radermacher, A. D. Gamification in computer science education: A systematic literature review. 2018 Asee Annual Conference & Exposition (2018).

Goodhue, D. L., and Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Q.* 19, 213–236. doi: 10.2307/249689

Hayase, J., Liu, A., Choi, Y., Oh, S., and Smith, N. A. (2024). Data mixture inference: What do Bpe tokenizers reveal about their training data? *arXiv:2407.16607*. doi: 10.48550/arXiv.2407.16607

Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., et al. (2020). Measuring massive multitask language understanding. *arXiv:2009.03300*. doi: 10.48550/arXiv.2009.03300

Hosseini, M., Gao, C. A., Liebovitz, D. M., Carvalho, A. M., Ahmad, F. S., Luo, Y., et al. (2023). An exploratory survey about using Chatgpt in education, healthcare, and research. *PLoS One* 18:e0292216. doi: 10.1371/journal.pone.0292216

Jung, Y., Chen, C., Jang, E., and Sundar, S. S. Do we trust Chatgpt as much as Google search and Wikipedia? Extended Abstracts of the Chi Conference on Human Factors in Computing Systems (2024), 1–9.

Kiryakova, G., and Angelova, N. (2023). Chatgpt—a challenging tool for the university professors in their teaching practice. *Educ. Sci.* 13:1056. doi: 10.3390/educsci13101056

Leung, S.-O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *J. Soc. Serv. Res.* 37, 412–421. doi: 10.1080/01488376.2011.580697

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochem. Med. (Zagreb.)* 22, 276–282. doi: 10.11613/BM.2012.031

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. NIPS'13: Proceedings of the 27th International Conference on Neural Information Processing Systems, 26.

Radford, A., Narasimhan, K., Salimans, T., and Sutskever, I. (2018). Improving language understanding by generative pre-training. OpenAI. Available at: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

Singh, C., Inala, J. P., Galley, M., Caruana, R., and Gao, J. (2024). Rethinking interpretability in the era of large language models. *arXiv:2402.01761*. doi: 10.48550/arXiv.2402.01761

Singh, H., Tayarani-Najaran, M.-H., and Yaqoob, M. (2023). Exploring computer science students' perception of Chatgpt in higher education: a descriptive and correlation study. *Educ. Sci.* 13:924. doi: 10.3390/educsci13090924

Sweller, J. (2011). "Chapter two - Cognitive load theory" in Psychology of learning and motivation. Vol. 55. Eds. J. P. Mestre and B. H. Ross (Cambridge, MA, USA: Academic Press), 37–76. Available at: https://www.sciencedirect.com/science/article/pii/B9780123876911000028

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., et al. (2023). "What if the devil is my guardian angel: Chatgpt as a case study of using chatbots in education" in Smart Learning Environments, vol. 10.

Vaswani, A. (2017). Attention is all you need. *arXiv:1706.03762*. doi: 10.48550/arXiv.1706.03762

Venkatesh, V. (2022). Adoption and use of Ai tools: a research agenda grounded in Utaut. *Ann. Oper. Res.* 308, 641–652. doi: 10.1007/s10479-020-03918-9

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: toward a unified view. *MIS Q.* 27, 425–478. doi: 10.2307/30036540

Venkatesh, V., Thong, J. Y., and Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS Q.* 36, 157–178. doi: 10.2307/41410412

Vest, J. M. T., Steinwachs, M., Saichaie, K., and Heintz, K. (2024). Winter 2024 generative Ai student survey. Available online at: https://cee.ucdavis.edu/Genaisurvey.

White, L. L. A., Balart, T., Shryock, K. J., and Watson, K. L. (2024). Understanding faculty and student perceptions of Chatgpt. 2024 ASEE-GSW Conference. Canyon, TX: ASEE Conferences.

Yilmaz, H., Maxutov, S., Baitekov, A., and Balta, N. (2023). Student attitudes towards chat Gpt: a technology acceptance model survey. *Int. Educ. Rev.* 1, 57–83. doi: 10.58693/ier.114