



RESEARCH ARTICLE

Integrating Artificial Intelligence-Powered Large Language Models in English as a Foreign Language EFL Teacher Education Programs

Atieh Mohd Albadarneh¹, Amjad Mahmoud Daradkah^{1*}, Esraa Ahmed Telfah², Haggag Mohamed Haggag³, Fadi Fuad Ghawanmeh⁴, Yasmin Abdullah Al-Shunnaq⁵, Maria Salih Tawalbeh⁵, Ashraf Mahmoud Ahmed Mahmoud⁶, Karima Shahine⁶, Heba Sadiq Daradkah⁷

¹ Department of Educational Administration, Faculty of Educational Sciences, Ajloun National University, Ajloun, Jordan

² Department of School of Education, Faculty of Educational Sciences, Irbid National University, Irbid, Jordan.

³ Professor of Curriculum & EFL Methodology Hurghada Faculty of Education, South Valley University, Qena, Egypt,

⁴ Department of Educational Administration, Faculty of Educational Sciences, Zarqa University, Zarqa, Jordan

⁵ Talal Abu-Ghazaleh for education quality (TAG-EDUQA). Amman Jordan

⁶ Department of Comparative Education & Educational Administration, Hurghada Faculty of Education, South Valley University, Qena, Egypt,

⁷ PhD in Curriculum & EFL Methodology. Qena, Egypt,

⁸ Ministry of Education Jordan.

ARTICLE INFO

Received: Apr 24, 2024

Accepted: Jul 2, 2024

Keywords

Large Language Models
Artificial Intelligence - TEFL
Education

*Corresponding Author:

amjad.drardkah@anu.edu.jo

amdaradkah@yahoo.com

ABSTRACT

Large Language Models LLMs are revolutionary advancement in Artificial Intelligence AI studies that affected different disciplines including teacher education. This research aimed to identify the current LLM oriented practices in English as a Foreign Language EFL teacher education programs in the Arabian context. The research adopts the descriptive analytical design through the application and analysis of Large Language Models questionnaire designed by the researchers. The questionnaire included (12) key LLM types and it was run over (540) pre-service EFL teachers in different Arabian teacher education contexts. Results showed that two models were reported to be highly utilized by the participants; these models were (chat models, and questioning models). Further, six models reported to be (mid), these models were (copywriting models, creation models, tuning models, inference models, research models, and exams models). In addition, four models reported to be used in (low) levels by the participants, these models were (coding models, developers' models, operation models, and hardware models). Based on the obtained results, the research recommends integrating different types of LLMs in EFL teaching education programs based on their actual performances over these models. These models include copywriting, creation, tuning, questioning, and inference models as well; besides, considering its risks such as lack of adaptability, bias, cost and other use-oriented limitations.

INTRODUCTION

Teacher education programs bridge the gap between national and international visions and challenges in the different educational settings. Countries all over the world follow educational patterns and models for their prospective teachers within these frames. In the Arabian context, the

use of Artificial Intelligence and its models and applications is a challenging paradigm in the different teacher education programs since the level of practice, outcomes, and educational policies are varied. One of these challenges is the recent Artificial Intelligence based Models.

Language Models (LLMs) have emerged (Naveed, et al., 2024) as cutting-edge artificial intelligence systems that can process and generate text with consistent communication. Its historical background stems from the Natural Language Processing and Generative AI to Pre-Trained Language models, transformers, until the recent gigantic modeling such as Chatbots and other General Purposes models. Wu (2024) defined LLM to mean " Large language models (LLMs) are transformer-based neural network architectures pre-trained with language models as loss functions using extensive amounts of text data, often hundreds of billion or even trillions of words, and a large number of model parameters, often billions" (p.1). LLM depend on Artificial Intelligence and can perform various tasks in different fields including the educational context. LMS are models trained with huge amounts of data to accurately predict and structure the next word in a sentence.

Integrating Large Language Models in the educational context requires identifying the needs, perceptions, and possible integrations from the users' views. For instance, Gao, Wang & Wang (2024) explored EFL university teachers' beliefs in integrating ChatGPT as LLM and other large language models in language education China. The study aimed to investigate the beliefs of (95) EFL university teachers from Chinese universities regarding the integration of LLMs in language education, as well as the relationships between their beliefs and other factors. The study yielded several concerns among Chinese EFL teachers regarding LLMs integration, such as neglecting of traditional learning resources, academic integrity, and excessive reliance.

The integration of Artificial Intelligence in EFL context has largely affected the current and future practices of teachers of pre-service teachers (Zhang, e al., 2023). Further, they confirmed that "there has been a significant increase in the utilization of artificial intelligence (AI)-based educational applications in education" (p.1). Their perceptions and views may have impact on the learning outcomes of their future students.

Thus, the human interaction with LLM was explored in a study by Jeon & Lee (2023) which concluded (4) ChatGPT roles (interlocutor, content provider, teaching assistant, and evaluator) and three teacher roles (orchestrating different resources with quality pedagogical decisions, making students active investigators, and raising AI ethical awareness). This confirms the distinctive roles by the teachers themselves in the process of learning through Large Language Models.

In EFL context, Large Language Models LLM (Kosten, 2024) should be discussed within (TEFL) teacher education for two reasons: "Firstly, to respond to the easy availability of LLMs – and not only with “pedagogy of the gaps” – (future) teachers need what might be framed as an “AI-informed pedagogy”. The study of LLM in EFL context may provide insights about the actual practices and future implications for better teacher education practices. For instance, in their study Kasneci et al. (2023) explored ChatGPT as a large language model in teacher education programs. The study presented the potential benefits and challenges of educational applications of large language models, from student and teacher perspectives. They also argued that LLM in education require teachers and learners to enhance their competencies and literacies to understand the technology and its limitations.

In teacher education context, Haggag (2018) discussed implementing a European document (European Portfolio of Student Teachers of Language EPOSTL) as a self-assessment tool for pre-service teachers' competencies. The research design experiments the European portfolio for pre-service language teachers (EPOSTL), which has been translated by the researcher and published by the Council of Europe. Results of implementing the portfolio on a sample of (N. 30) students at Hurghada Faculty of Education during their practicum (3 groups in 3 official language schools) are

reported. The results indicate that using EPOSTL as a self-assessment tool has positive impacts on developing the five areas of the portfolio.

Kim & Kwon (2023) argued that although Artificial Intelligence (AI) significance in the educational contexts grows, still the lack of teacher readiness and beliefs hinder the integration of AI in schools. In this view, their study aimed to explore South Korean elementary school teachers' experiences teaching AI curricula and examine their competencies. A survey and interviews were conducted with (67) elementary school teachers who have been working in AI-leading schools in South Korea. Results showed that teachers were least confident in content knowledge, followed by technological knowledge and pedagogical knowledge relevant to AI. Additionally, interviews with 13 teachers revealed "five themes regarding AI education: (1) emphasizing the importance of instructional design in AI education; (2) redesigning the learning environment to promote learning experiences; (3) lowering teaching anxiety by acknowledging the limitations of content knowledge" (p.10).

Future teachers need to be confidently equipped to teach 21st century ICT skills. In their study, Kruskopf et al. (2024) investigated teaching self-efficacy (TSE) in ICT competencies among teacher students. They confirmed distinct ICT competencies among two cohorts from teacher training programs ($n = 347$; $n = 428$): practical (i.e., device and data management), and algorithmic (i.e., programming, and data security). Regression analyses indicated TSE-biases regarding younger age, male gender, and a background in natural sciences, with significant interactions between age, gender, and having learned such ICT-skills already in school. The findings point to a need for tailored strategies in teacher education to mitigate TSE disparities.

The following model by Gan & Wu (2023) depicts the various aspects of using edu-LLM in education.

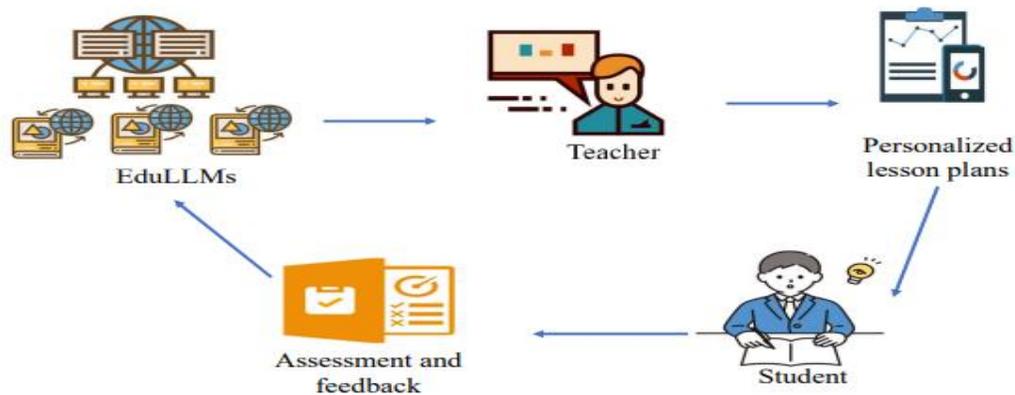


Fig. 1: Large Language Models in Education (Gan & Wu, 2023)

The figure shows the integration of using teacher-oriented practices that rely on LLM such as lesson planning and its impact on the student. Then, a comprehensive assessment and feedback is used in this circular process. The figure confirms the integrative view of these models in the process of education in general. Thus, this integrative view may be utilized in English as a Foreign Language teacher education context.

In sum, Large Language Models are advanced AI systems that are capable of understanding and generating human language. They are built using complex neural network architectures, such as transformer models, inspired by the human brain. LLMs are a type of artificial intelligence (AI) that uses deep learning techniques and big amounts of data to summarize, generate, and predict language (Ranta; Lammassaari & Lonka, 2024). These models are trained on vast amounts of data, enabling them to comprehend context and produce coherent text-based outputs, whether answering a question or crafting a narrative.

Therefore, this research aimed to explore integrating Large Language Models-based Artificial Intelligence in different Arabian EFL Teacher Education settings.

Problem and its significance

Despite the growing interest in the application of LLMs in language education, there is still a research gap in understanding teachers' beliefs in the integration of LLMs in language education (Wu, 2024; Gan & Wu, 2023; Gao et al., 2022; Kim & Kwon, 2023; Zhai, 2022). Based on the above literature and the researchers' teaching experiences to the participants, problem of the research can be stated in the need to identify pre-service EFL teachers' practices related to Large Language Models. It aims to verify the level of LLM use in their teacher education programs. Thus, the present research is thought to be significant due to the following considerations:

1. It highlights Large Language Models use in prospective EFL teachers' educational programs?
2. The interdisciplinary nature of the research may integrate the virtues of both AI and EFL paradigms.
3. EFL prospective teachers may use the insights in the research to develop their professional development in the light of their LLM oriented practices.
4. It compares between the levels of practices regarding Large Language Models' types.
5. This research is an attempt to call for reform in teacher education programs in the light of current technological advancements.

Questions of the research

1. What are the key Large Language Models used in EFL teachers' educational programs from the participants' views?
2. What are the mean differences between the different areas of Large Language Models in EFL teacher education programs?

Objectives of the research

The research aimed to the following:

1. Identifying the key Large Language Models used in prospective EFL teachers' educational programs.
2. Comparing between LLMs models' levels of practices based on the participants' views.

Design

This research follows the descriptive analytical design. It implements a questionnaire that includes (12) types of Large Language Models. It was applied over (540) participants in EFL teacher education programs in various Arabian contexts.

Participants

The students participated in the research are enrolled in different Arabian English as a Foreign Language teacher education programs. Using convenience sampling, we surveyed (590) pre-service EFL teachers during the academic year (2023-2024) in different Arabian contexts. After explaining the purpose of the research, the questionnaires were distributed, and the students were asked to complete the surveys independently. In the research, (50) completed questionnaires were considered to have insincere responses through the first coding process and excluded from the research, and (540) questionnaires were selected as the final valid sample for data processing.

Delimitations

The present research was delimited to only 12 models of AI-Large Language Models. These models were validated by the jury and reported to be valid models in English as Foreign Language context. Other models

can be used for different contexts. The research was also delimited to the Arabian teacher education context as these contexts follow similar sequence in the process of EFL teacher education.

Instrument of the research

The research utilized Large Language Models (LLMs) questionnaire; the questionnaire aimed to identify 12 areas of LLMs that can be used in EFL teacher education programs. These areas included (Copywriting Tools- Coding Tools- Developers Tools - Chat Tools - Operations Tools - Model - Creation Tools - Hardware Tools - Fine Tuning Tools - Inference Tools - Questioning & Answering - Research & Discovery - Quiz & Exam Generators). The questionnaire was validated statistically in terms of validity and reliability. The following table shows the reliability value of the questionnaire.

Table (1) Reliability Statistics

Number of Items	Cronbach's Alpha
12	.806

The obtained r value (.80) shows a high reliability value for the questionnaire and its items. Content validation was run through a jury of TEFL experts who recommended adapting and correcting some items to reach its final valid form with (12) main areas of Large Language Models that can be appropriate for the purpose of the research.

Validity of the questionnaire was calculated by jury validation who recommended adding some models as an image to the participants. They also modified the layout of the scale and its instructions. Besides, an internal consistency was statistically calculated as the following table (2) shows Pearson internal consistency levels.

Table (2) Descriptive Statistics

No	Model	Correlation
1	Copywriting models	0.513
2	Coding models	0.574
3	Developers models	0.340
4	Chat models	0.428
5	Operations models	0.601
6	Model Creation tools	0.362
7	Hardware tools	0.795
8	Fine Tuning models	0.578
9	Inference models	0.604
10	Questioning and answering models	0.440
11	Research and discovery models	0.549
12	Quiz and exam operators	0.528

The above table (2) shows the correlation between the 12 items of the questionnaire and their total score. The obtained values vary between (0.340) to (0.795); this shows strong positive correlation and internal consistency between the different items of the questionnaire, and thus its validity and usefulness to achieve its aims.

RESULTS

Using SPSS statistical analysis, the obtained quantitative results verified the questions of the research. The following table (1) shows the statistics of the models, means, standard deviations, and the degree of each of the 12 models in the questionnaire.

Table (3) Descriptive Statistics

LLM Types	N	Mean	Std. Deviation	Degree
Copyrighting models	540	1.71	.525	Mid
Coding models	540	1.29	.511	Low
Developers models	540	1.45	.760	Low
Chat models	540	2.73	.754	High
Operation models	540	1.32	.521	Low
Creation models	540	1.71	.680	Mid
Hardware models	540	1.40	.574	Low
Tuning models	540	1.67	.678	Mid
Inference models	540	1.73	.660	Mid
Questioning models	540	2.70	.726	High
Research models	540	1.86	.720	Mid
Exams models	540	1.88	.677	Mid
Valid N (listwise)	540			

According to table (3) results, two models were reported to be (high) these models are (Chat models 2.73, and questioning and answering models 2.70). Six models were reported to be used in (mid) levels, these models are (copywriting models 1.71, creation models 1.71, tuning models 1.67, inference models 1.73, research models 1.86, and exams models 1.88). Four models were reported to be (low) by the participants; these models are (coding models 1.29, developers' models 1.45, operation models 1.32, and hardware models 1.40).

The following figure (2) shows the percent in each model as viewed by the participants of the research.

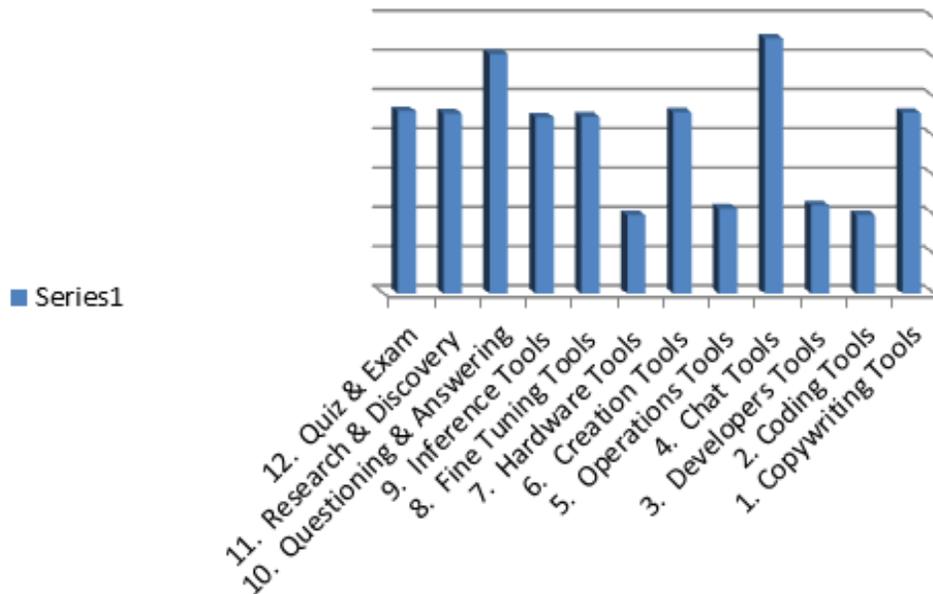


Figure 2: Percent of each model

The above table clearly shows that the top three large language models used by the participants are Chat models, and questioning and answering models. The following discussion contextualizes these obtained results in the light of pertinent studies.

DISCUSSION & IMPLICATIONS

The obtained results from the questionnaire highlighted 12 key LLMs that included (Copywriting Tools- Coding Tools- Developers Tools - Chat Tools - Operations Tools – Model - Creation Tools - Hardware Tools - Fine Tuning Tools - Inference Tools - Questioning & Answering - Research & Discovery - Quiz & Exam Generators).

The obtained results are consistent with many studies that highlighted the positive impacts of AI-powered LLMs in the area of language education (e.g. Kosten, 2024; Wang et al., 2024; Wu, 2024). For instance Gan & Wu (2023) stressed the points that LLM in education can provide personalized learning, adaptive feedback, access to diverse resources, natural language interaction, continuous learning support, content generation and creation, Multilingual capabilities, analyzing data, ethical consideration & collaboration). Similarly, the results are consistent with Kosten's (2024), in his study conducted within the EFL teacher education department of a German metropolitan university during the 2023 summer term, it analyzed 21 student essays written after an interactive 90-minute training session on ChatGPT, conducted in three intact TEFL teacher education seminars. The results also might be of interest for researchers and teacher educators interested in "(a) modeling digital text sovereignty as it applies to AI, and (b) developing teacher education programs that target digital text sovereignty in an AI context, or that aim to support future teachers in developing an AI-informed pedagogy" (p.1).

Wang et al. (2024) surveyed Large Language Models in education and provided important results about the effective use of these models in general as well as future recommendations. Although the present research is different in its scope as it reviewed the views of EFL pre-service teacher education programs in the Arabian context. They stressed the view that the advent of large language models (LLMs) has brought in a new era of possibilities in the realm of education. They summarized the various technologies of LLMs in educational settings from multifaceted perspectives, encompassing

student and teacher assistance, adaptive learning, and commercial tools. Results showed that the commonly used LLM by the participants is the questioning and answering models (2.70). This can be due to the exam based educational systems (Mousa, 2017) that foster assessment-driven exams. Students tend to use these models to respond to tasks, questions, and assignments.

The results are also consistent with the study by Gao, Wang & Wang (2023) that aimed to Explore EFL university teachers' beliefs in integrating ChatGPT and other large language models in language education: a study in China. Findings of the study showed that previous experiences with LLMs, frequency of LLMs use, and self-evaluation on stages of LLMs integration all played vital roles in shaping university teachers' beliefs in integrating LLMs in language education. It also showed that there were no significant correlation between university teachers' beliefs in integrating LLMs in language education and the availability of IT personnel. The present study also highlighted the use of chat models in EFL teacher education practices.

The limitations and risks that can be drawn in the light of these results are also consistent with literature. Kasneci (2023) summed the key limitations in the following: (a) copyright issues, (b) bias & fairness, (c) cost, (d) lack of expertise, and (e) lack of adaptability. Results showed the use of LLM in assessment and chatting purposes which highlight the issues of security, privacy and text originality. A key influence for LLM in teacher education is its impact on assessment fairness. Some educational contexts consider these issues in integrating these models in teacher education contexts.

Implications of the research stem from the theory and practice of using LLMs in language education in general and in Foreign Language Learning FLL in particular. TEFL courses can utilize these models in their exercises, texts, assignments and even in their assessments. Teachers can also refine and use these models in adaptive contexts to achieve the desired Intended Learning Outcomes ILOs of their courses. Contextualizing the use of LLMs can extend to professional development and teacher training contexts as well as pre-service teacher education programs.

CONCLUSION

Large Language Models are key advancements in the area of teacher education and consequently EFL teachers' education. The research explored 12 areas of LLM in the Arabian context. Participants were varied in their use and cognition about these models. Results revealed that the highly used models by the participants included testing and assessment models as well as chat models. Thus, there is a need to implement other AI-based models in teacher education programs to achieve quality and professionalism in teacher education programs. These models can extend to design, creation, copywriting, tuning and other AI based models. Limitations of using LMS in the Arabian context are consistent with global concerns related to cost, adaptability, fairness, and the other issues; these issues were obvious in both literature and in the obtained results.

REFERENCES

- Daradkah, A., Alotaibi, T., Badarneh, H., Momani, K., Hamadin, K., Alqudah, R., AL-Momani, M., Al-Ibrahim, A., Ashour, M., Almawadeh, N. & Mahmoud, A. (2023a). Proposed vision of the transformation of the Arab Universities into smart digital universities. *Information Sciences Letters*, 12(9), 2355- 2374.
- Daradkah, A., Alotaibi, T., Mahmoud, A., Awais, B., Al-Qudah, M., AL Shannaq, R., AL-Momani, M., AL_Ruheel, A., Albadarneh, A., Alqsaireen, E., Alkenani, R., Badarneh, H. & Mahmoud, S. (2023b). Proposed model for integration and cooperation between university, industry, and government in Arab countries: Innovation Triple Helix Model. *Applied Mathematics & Information Sciences*, 17(6), 1109- 1144.

- Daradkah, A., Mahmoud, A., AL-Momani, M., Al-nemrat, A., Badarneh, H., Hamadin, K., Almadawadeh, N., Alqudah, R., Ashour, M., AlKhatib, F., Mahmoud, S. (2023c). Degree of the requirements for improving human capital management in Arab universities as a gateway to achieving sustainable development. *Information Sciences Letters*, 12(10), 2617- 2640.
- Gan, W. & Wu, J. (2023). Large Language Models in Education: Vision and Opportunities. Conference Paper: *IEEE International Conference on Big Data (Big Data)*. DOI: 10.1109/BigData59044.2023.10386291
- Gao, Y., Wang, Q., & Wang, X. (2024). Exploring EFL university teachers' beliefs in integrating ChatGPT and other large language models in language education: a study in China. *Asia Pacific Journal of Education*, 44(1), 29–44. <https://doi.org/10.1080/02188791.2024.2305173>
- Haggag, M. (2018). Assessing Core Didactic Competencies of Pre-service English language Teachers during Practicum: a European Perspective, *European Scientific Journal*, 14(17), 12-32.
- Jeon, J. & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies Journal*, 28 (1), <https://doi.org/10.1007/s10639-023-11834->.
- Karimi, F. (2014). Didactic competences among teaching staff of universities in Kenya, *International Journal of Higher Education*, 3(2), Doi: 10.5430/ijhe.v3n2p28
- Kasneci, E., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences Journal*, 103 (1), <https://doi.org/10.1016/j.lindif.2023.102274>.
- Kim, K. & Kwon, K. (2023). Exploring the AI competencies of elementary school teachers in South Korea. *Computers and Education: Artificial Intelligence Journal* 4 (1) , <https://doi.org/10.1016/j.caeai.2023.100137>.
- Kosten, J. (2024). Just a Pocket Knife, Not a Machete: Large Language Models in TEFL Teacher Education and Digital Text Sovereignty. *Technology in Language Teaching and Learning*. 6(1), 1–16. DOI: <https://doi.org/10.29140/tl.v6n1.1192>
- Kruskopf, M.; Abdulhamed, R.; Ranta, M.; Lammasaari, H. & Lonka, K. (2024). Future teachers' self-efficacy in teaching practical and algorithmic ICT competencies – Does background matter?, *Teaching and Teacher Education*, 144, <https://doi.org/10.1016/j.tate.2024.104574>
- Mousa, M. (2017). Technology Based Assessment and Enhancement of Thinking Skills: A Case Study of the Educational System Development in Palestine. *International Humanities Studies*, 4 (2), 21-33.
- Paisi, M. (2014). Dimensions of teaching staff professional competences. *Social and behavioral science*, 140 (1), pp. 924-929
- Tigelaar, H.; Dolmans, M.; Wolfhagen, P.; & Van Der Vleuten, M. (2004). The development and validation of a framework for teaching competencies in higher education, *Higher Education*, 48, 253–268.
- Wang, S. et al. (2024). Large Language Models for Education: A Survey and Outlook. Conference: KDD 24: 30th, ACM SIGKDD Conference on Knowledge Discovery & Data Mining; Aug 25 - 29, 2024; Barcelona, Spain.
- Wu, Y. (2024). Large Language Model and Text Generation. In: Xu, H., Demner Fushman, D. (eds) *Natural Language Processing in Biomedicine. Cognitive Informatics in Biomedicine and Healthcare*. Springer, Cham. https://doi.org/10.1007/978-3-031-55865-8_10

- Zhai, X. (2022). ChatGPT user experience: Implications for education. *Available at SSRN 4312418*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4312418
- Zhang, C.; Schießl, J.; Plößl, L.; Hofmann, F.; and Gläser-Zikuda, M. (2023). Acceptance of artificial intelligence among pre-service teachers: a multi-group analysis. *International Journal of Educational Technology in Higher Education*, 20 (40). <https://doi.org/10.1186/s41239-023-00420-7>.