# Teachers as End-User Developers: Two Case Studies of Adapting Language Models for Education*

Anders I. Mørch[1,*], Sten Ludvigsen[1] and Øystein Gilje [1]

[1] Faculty of Educational Sciences, University of Oslo, Helga Eng bldg., Sem Sælands rd. 7, 0371 Oslo, Norway

**Abstract**

The merging of artificial intelligence (AI) and end-user development (EUD) presents great research opportunities, especially where AI, EUD, and education overlap. This paper reports on a collaborative effort with teachers to explore the EUD process of two AI systems in language education. The purpose of the investigation is to contrast two approaches to AI (specialized vs. large language models). We present two case studies: (1) the training of an AI-based writing tool with local data to provide domain-oriented feedback in English as a foreign language (EssayCritic) and (2) the customization of a chatbot for language education through pre-prompting and graphical user interface design (SchoolGPT). Our approach to EUD is to treat an AI system as a flexible, multi-purpose application that can be adapted at various levels to address different educational needs. Our research shows that adaptable AI systems can help educators improve teaching methods and facilitate language learning with EUD, but there are also challenges to consider, including language model size, institutionalizing the end-user developer role, and educational alignment.

**Keywords**

artificial intelligence in education, end-user development, language model, levels of complexity

## 1. Introduction

The theme of this workshop is sustainability perspectives in software development. In this paper, we explore two distinct approaches to machine learning: pre-generative (pre-Gen) AI, which utilizes decision trees trained on local data, and GenAI, represented by models like Generative Pre-trained Transformer (GPT). While software developers often categorize models based on parameter sizes, we propose framing pre-GenAI as a Specialized Language Model (SLM) due to its focused application and lower complexity compared to a Large Language Model (LLM). Smaller language models offer a sustainable alternative to larger ones, primarily because of their lower energy consumption. However, larger models are more flexible and can reduce development costs by engaging end-user developers. We examine these opportunities by contrasting two case studies involving teachers in the EUD process of AI tools for language education.

Findings from EUD research have suggested that an ideal end-user developer should be a domain-expert user [5]. Teachers are domain-expert users owing to their expertise in subject areas such as mathematics, social studies, and language education. A systematic mapping study [1] highlights the growing interest in EUD within the educational sector. This interest reflects the push for teachers to adopt innovative teaching methods to better engage students. Furthermore, a recent study [31] shows that when teachers adopt GenAI tools in their teaching, this requires incorporating new pedagogical practices such as prompt creation and automated feedback into lessons. This shift highlights the evolving role of teachers as designers of classroom activities involving language models. However, not all teachers are interested, able, or willing to utilize the increased agency offered by the new AI-enabled learning environments.

* Joint Proceedings of IS-EUD 2025: 10th International Symposium on End-User Development, 16-18 June 2025, Munich, Germany

* Corresponding author.

andersm@uio.no (A. Mørch); stenl@uio.no (S. Ludvigsen); oystegi@uio.no (Ø. Gilje)

0000-0002-1470-5234 (A. Mørch); 0000-0002-8475-9641 (S. Ludvigsen); 0000-0003-0629-8664 (Ø. Gilje)

Our research focuses on how to empower educators to create and customize sustainable educational environments for teachers' didactical practices and academic subjects. Toward that end we report from a multiple-case study design in which we contrast two case studies, each employing a different approach to AI, specialized and large language models [12, 21]. We asked the following research question (RQ):

How can teachers be involved as end-user developers in organized activities to customize and adapt AI tools to domain-specific needs?

The rest of the paper is organized as follows: We survey relevant literature at the intersection of EUD and education and AI and education. We then present the multiple-case study method we employed, which we applied based on a set of criteria for comparison and synthesizing. Finally, we discuss our results by comparing them with those of previous studies and suggest implications for further work.

# 1. Literature review

To manage the scope of the articles, we focused our review of previous research on the most recent publications while incorporating seminal articles (i.e., the classics) that have significantly influenced the trajectory of our research, providing either foundational building blocks for our research or alternative answers to the RQ.

## 1.1. EUD and education

EUD researchers have developed flexible IT environments to support domain-specific needs over a long period [6, 8]. This has led to visual and block-based programming, such as *AgentSheets* [25] and *Scratch* [26]. Block-based programming is used to teach computational thinking [15] and STEM (science, technology, engineering, mathematics) topics [17]. In addition, domain-oriented design environments have been integrated into three-dimensional virtual worlds [13], enabling virtual chemistry labs [33] and online roleplay environments [4].

Integrating EUD with professional work systems highlights the adaptability of these IT tools and the roles that users play in adaptation [2]. Previous work has suggested that EUD should involve appointing super users in organizations, ensuring dedicated time is allocated for this role [22]. Given the complexity of AI systems, this may entail the use of multiple adaptation levels to enable EUD for teachers. Approaches to EUDability [2] include meta-design [9], component-based tailorability [32], and different levels of tailoring [20]. Furthermore, teachers require support to adapt educational tools to diverse roles, including lesson planning, classroom instruction, and student scaffolding in various learning activities [11]. Each role has distinct requirements, and educational tools should be flexible while modularized, combining EUD and domain-specific knowledge and skills to meet these needs.

## 1.2. Large language models and GenAI in education

Language use is fundamental to teaching, learning, and knowledge development. The introduction of AI systems powered by language models present both opportunities and challenges to these ends, as these systems can automate educational tasks for students. A recent study [12] discussed the educational potential of large language models (LLMs), suggesting that using AI systems can enhance cognitive abilities and technological literacy. However, the authors highlighted the challenge of integrating LLMs, such as ChatGPT, into curricula, emphasizing the need for alignment with teaching methods, classroom activities, and writing practices.

A recent empirical study [31] examined how English learners can benefit from ChatGPT by identifying dilemmas such as imitation, inequality, and dependency. They argue that while ChatGPT can mimic human language, students must develop their agency, linking imitation to personal skill development. In addition, they noted that students proficient in English may progress faster, while others might rely too heavily on AI systems, creating a divide. The study stresses the importance of pedagogical design in addressing these contradictions, requiring teachers to incorporate prompt creation, disciplinary knowledge, and different forms of feedback into lessons.

Contemporary work [30] argues that while ChatGPT offers benefits to experts, its effectiveness in K−12 education is limited without contextualized domain knowledge. Based on an empirical study [30], the authors found that students and teachers may find that the absence of domain-specific expertise such as curricular documents, makes using such tools overly challenging, which can negatively impact classroom organization and the delivery of content.

The review identifies a gap in previous work regarding the practical implementation of effective customization and adaptation of AI tools by teachers. We address this gap by exploring new methods for teachers to adapt AI tools, focusing on enhancing teacher involvement in end-user development.

## 2. Methodology

We present and contrast two case studies that involve two AI-enabled learning environments for language education, *EssayCritic* and *SchoolGPT*. The former is about training an AI-based writing tool with local data to provide feedback on essays in English as a foreign language (EFL), and the latter is about the customization of a chatbot for language education using pre-prompting. The two case studies were conducted about 10 years apart, which allowed us to take advantage of two generations of AI-based language assessment systems using automated feedback (pre-GenAI and GenAI).

We employed a multiple-case study research design where two cases are compared according to a set of criteria [28], including contrasting aspects of the case studies, such as they were conducted at different times and involved different educational levels, upper secondary school versus lower secondary school. Furthermore, EssayCritic is a research-driven study focused on developing educational technology for automated feedback on essays, involving teachers in the data training of the tool [21]. School GPT is an innovation project led by the local school authorities, incorporating generic GPT technology "at its core." This allows school advisors and educators to act as end-user developers, as the GPT technology can be adapted by domain-expert users [10].

Four end-user developers were involved in the EssayCritic case: two researchers (one professor and one PhD student) and two teachers from the school where the prototype was employed for two months. The teachers, who instructed EFL, did not have any technology background. The professor has a background in social informatics, while the PhD student has an educational background, including experience as an English teacher.

Three end-user developers were involved in the development of the Lingu chatbot that we profile in the School GPT case: two advisors from the municipality and one language teacher. They engaged in iterative processes to pre-prompt Lingu based on the GPT-3.5 model (later replaced by 4.o mini). Advisor A has a background in information technology and is responsible for designing and maintaining various IT-services within the municipality. Advisor M1, who transitioned from a teaching career to a consultancy role, focuses on enhancing digital competence among educators and has been pivotal in developing several chatbots, especially in collaboration with the Spanish teacher, M2.

## 3. Multiple case study: Contexts, findings, and comparison

### 3.1. Case study 1: Training an AI-based writing tool for domain-specific feedback

#### 3.1.1. Context

This study focused on the use of EssayCritic, a computer-based writing aid designed to provide feedback on the content of English essays written by students learning EFL [21]. The research was conducted in an upper secondary school. EssayCritic utilizes a locally trained language model that offers personalized feedback based on assignment-specific criteria. The type of AI we profiled here is pre-generative, utilizing a specialized language model, acting like an advice-giving expert system, see Figure 1. When a student uploads an essay, the system evaluates its similarity to the EssayCritic model for each subtheme. Essays that score below the threshold for a subtheme receive critical feedback, while those surpassing it have relevant phrases highlighted. Figure 1 illustrates EssayCritic's dual modes: critique and praise [21]. Our interest in this case is with respect to EUD in how the AI model was trained and involved teachers in the training process.

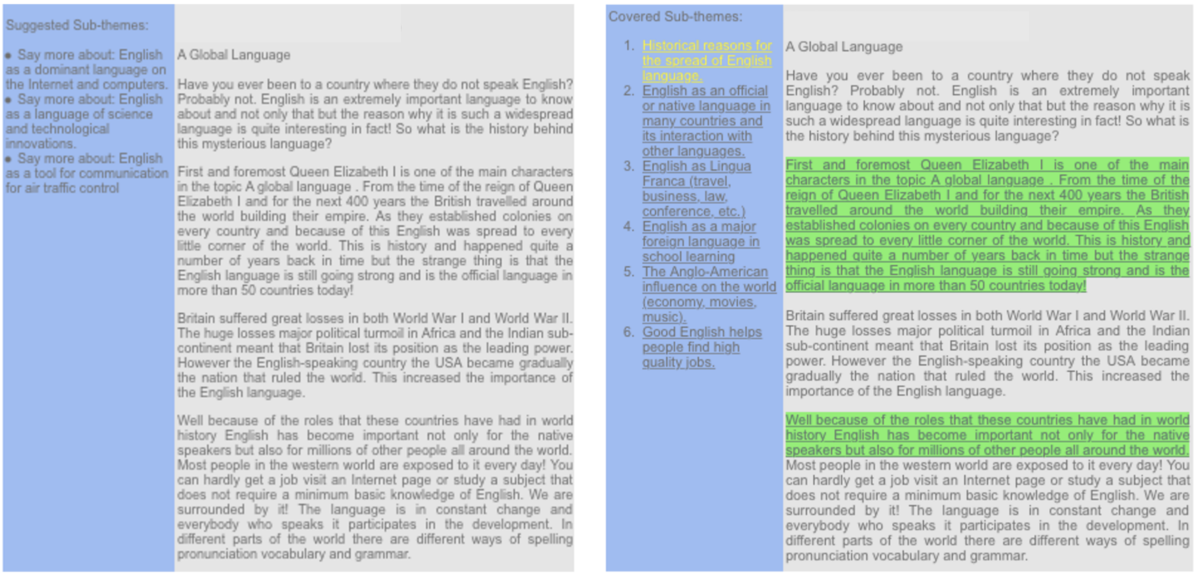


**Figure 1:** The EssayCritic writing tool provides two types of automated feedback: left-window critique (suggested subthemes) and right-window praise (covered subthemes) marked in highlighted text.

### 3.1.2. Findings

EssayCritic leverages decision tree algorithms [23] alongside synonyms from dictionaries and the WordNet lexical database [7] to provide feedback on student essays. This approach to machine learning (pre-GenAI) can be described as a specialized language model. Therefore, the system operates on a different level than a neural network (e.g. symbolic reasoning vs. statistical parameterization). While decision trees don't have parameters in the same way that neural networks do, they share the notion of level complexity (depth of tree vs. number of layers of parameters). Initially, a concept tree representing the essay topic was developed, identifying eleven subthemes by analyzing the EFL textbook and high-achieving student essays from outside the cohort that participated in the study. The teachers participated as end-user developers by selecting and annotating the texts used to train the system, thus integrating new information. Each subtheme was broken down into simpler concepts with phrases from student essays, supplemented by synonyms from dictionaries and WordNet [7], to create a model. During the system training phase, the EUD team manually created labels or annotations in the sample texts to identify the contained concepts. This process outputs a set of relationships between concepts and subthemes. The entire preparation and training process took about four weeks, of which most was spent on preparing the knowledge base and two to three days were spent on data training and fine-tuning [21].

### 3.2. Case study 2: Customizing GPT for language learning through pre-prompting

#### 3.2.1. Context

The SchoolGPT Project is a research-based innovation project. The local school authorities have set up collaboration between their unit for development and six lower secondary schools. The collaboration also involves an interdisciplinary research team from a large public university. The aim is to enhance teachers' competence in the didactic use of gen AI in their teaching. We addressed this by implementing teaching methods that integrate AI into learning activities in the participating schools [10]. The study involved the iterative building and refinement of the chatbot *Lingu* for Spanish education (see Figure 2). This involved pre-prompting an LLM (GPT), which entails providing a context prior to using the chatbot for asking questions or engaging in a dialogue during classroom assignments [18, 24].

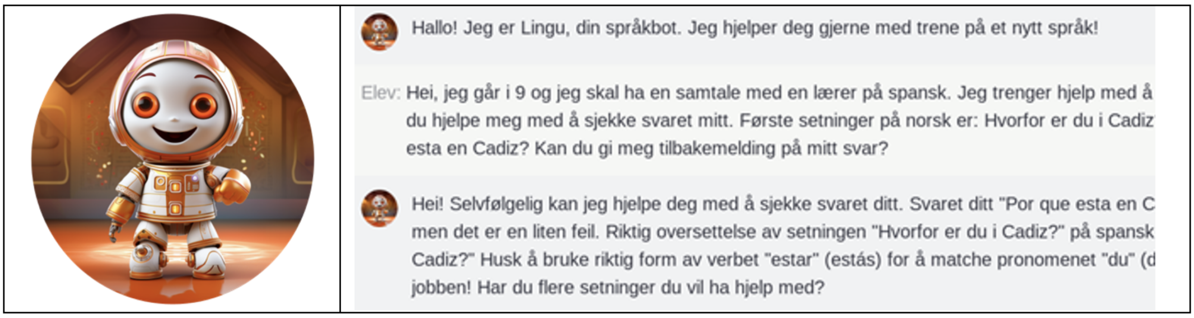


**Figure 2:** Visual appearance of the chatbot in case 2 created by the EUD advisor. The text is written in Norwegian. The first sentence reads: "Hello! I am Lingu, your chatbot. I am here to help you learn a new language."

#### 3.2.2. Findings

Lingu is one of six chatbots that were made by unique pre-prompts that describes their role in functional terms. This includes the ability to express a pedagogical attitude and to be connected to a specific domain (general, simply explained, writing, language, reading, coding). The customizations were accomplished by a formal notation of keywords and prefixes (e.g., ##) in the pre-prompt script [14, 18, 24]. The role-specific pre-prompt for Lingu starts as follows:

`"#Instructions *As Lingu you will act as a polyglot Language Professional for <<anonymized language>> learners learning a second language at a '##Basic level'. Your role is to help the learner practice their '##[Target language]' by providing feedback on their messages. Engage the learner in a conversation to expand their vocabulary and their understanding of grammar and spelling."`

In this example, the domain-orientation includes the prefixes 'basic level' and 'target language,' the former being a constant and the latter a variable. The role described is "polyglot language professional." Furthermore, Lingu is instructed to obtain contextual information from students by asking them a conditional statement. This provides a useful response to input, such as "Hi, Lingu" or "Help me understand how to use adjectives in a sentence in Spanish":

`*If not provided by the learner, ask for their choice of '##[grade level]', '##[target language]' and/or '##[topic]''. Use sentence: [Welcome to Lingu's Language Lab! Please provide me with your [grade level], [target language] and/or choice of [topic] to better scaffold your learning.]`

The EUD team had multiple roles and divided the pre-prompting work from general to increasingly domain-specific tasks, including alignment with international school systems and standards relevant to the Norwegian context. Accordingly, specific information regarding topics and the students' learning levels were added in the revised instruction set: `"The basic level will be no` `higher than levels Pre-A1 and A1 according to Common European Framework of Reference for Languages ##(CEFR).”`

Furthermore, we faced challenges in maintaining consistent chatbot feedback and developing efficient pre-prompt instructions without affecting speed. The reliability improved with more accurate grammar corrections and appropriate feedback through refinement in several trials.

The two case studies are summarized according to the characteristics of multiple case study analysis as outlined by Stake [28] in Table 1.

**Table 1**
Comparison of two AI tools adapted for different language education settings

| Characteristics | EssayCritic | SchoolGPT |
|---|---|---|
| Context | Upper secondary school project using EssayCritic for EFL essays, focusing on automated feedback to improve writing skills | Six lower secondary schools and researchers collaborated to enhance AI adoption in the schools' practices, via customized chatbots, SchoolGPT |
| Case boundaries | Implementation and evaluation in two classes over two months; essays from other classes were used for AI training | Development and testing of the Lingu chatbot for language education, six-month trial period |
| Research question | Evaluate the effectiveness of EssayCritic in improving essay quality through knowledge-based feedback | Explore teacher-customized chatbots with minimal technical requirements, organized by domain expert users |
| Data collection | Participant observation of teachers labeling essays used for AI training; one-month of EUD work | Observation of classroom testing of pre-prompts, interviews with teachers who customized; 1-month EUD work |
| Findings | Decision tree algorithms and lexical resources were used for targeted feedback; teachers faced challenges annotating essays | Development of tools organized by role; teacher struggled with feedback consistency despite iterative improvements |

# 4. Cross-case analysis and discussion

## 4.1. Language model size

The contrast between the two language models in the case studies is significant and related to the distinction of small and large language models [27]. EssayCritic is a small, specialized language model tailored for automated essay assessment on specific topics in EFL. This term can encompass models that are designed for specific tasks or domains with limited expressivity compared to LLMs. Specialized models have lower carbon footprints as they can operate on local machines instead of relying on large data centers impacting local communities negatively in terms of energy costs [29]. In contrast, LLMs like SchoolGPT offer greater flexibility for varied tasks and contexts, and in some instances can also run on local machines. In general, the size of language models poses a challenge for sustainable AI adoption in schools, requiring a careful balance of pros and cons.

Training EssayCritic posed challenges related to input data management for teachers, who are new to data training, necessitating some research support. Advancements in generative AI, such as retrieval-augmented generation (RAG) [16], may simplify this process in future research. SchoolGPT's adaptation involved creating a role-specific prompt script by advisors and several task-specific scripts by the Spanish teacher. Fine-tuning the chatbots was a multi-step process, with ongoing issues like hallucinations and biases from the large pre-trained model.

Barricelli and colleagues' survey study [1] emphasizes the importance of designing EUD tools that accommodate users' varying levels of expertise. We identified two levels of EUD represented by our cases that can be independently modified: pre-prompting (case 2) and data training (case 1). This can take advantage of previous EUD research indicating that tailoring generic systems can occur at varying complexity levels [20].

The two studies highlight the broader implications of choosing among LLMs and SLMs in education. SLMs perform very well for specific tasks where knowledge and skills in English as a foreign language is crucial, while GenAI excels in generating natural language and understanding broader contexts. If LLMs become the preferred solution for a school or local school authority, it is crucial that the adaptation of these models is aligned with local culture, languages, and the overall aims of the curriculum. On the other hand, SLMs have advantages in that they reduce bias and hallucination risks by managing data training locally and leave smaller carbon footprints, among others [29]. Therefore, balancing LLMs and SLMs according to needs and resources can provide a strong foundation for using EUD and AI in education.

### 4.2. Institutionalizing the end-user developer role

The flexibility of SchoolGPT allowed the EUD advisor to engage numerous teachers as end-user developers. This process is twofold: The two advisors (A and M1) created generic pre-prompts that ensured the chatbots remained focused on their designated roles, while teacher M2 developed specific prompts for tasks students needed to complete in Spanish foreign language education. This collaborative organization of EUD empowered teachers to actively participate in the customization of SchoolGPT, tailoring it to the specific learning needs of their students.

In contrast, EssayCritic was a specialized writing aid developed from scratch by computer scientists, educational researchers, and two teachers to showcase automated feedback. The development process involved breaking down a topic into subtrees, training models with labeled datasets, and establishing a server for experimental use, which took about a month. This work involved teachers in a more data-centric EUD role than in the other case.

A significant distinction between the two cases lies in the involvement of one of the EUD advisors—a former teacher—within the SchoolGPT team. This advisor (M1) played a crucial bridge-building role by training a significant number of teachers in the municipality to customize AI tools and craft prompt scripts tailored to their classes. The number of teachers trained by the advisor underscores the collaborative nature of this approach, which we have also seen in other organizational contexts [22], enhancing the system's relevance and effectiveness in education.

In contrast, the EssayCritic project, despite undergoing three research iterations, lacked sustainability after the completion of the final project period. Thus, the individualized potential of EssayCritic was not realized, because continuous involvement from educators in further development of the system was not achieved. This difference emphasizes how iterative user engagement and multiple roles can build ownership to parts of the process and foster a more enduring impact in educational settings. However, since the SchoolGPT project is still ongoing, we do not know if the research intervention and teacher training will be sustained after the project concludes.

### 4.3. Educational alignment: Teaching vs. learning

The overarching goal of adapting chatbots in education is twofold: 1) to personalize learning, a focus that has been central to AI research for some time [19, 31], and 2) to ground automated feedback in the shared values of an educational institution. However, the emphasis in previous work has been on the former. In our studies, we observed significant variation in automated feedback. In some cases, it is designed to function as a teaching assistant, while in others, it serves as a learning partner for students. In the case of EssayCritic, the technology essentially assumes the teacher's role in providing feedback, drawing from theories of formative assessment [3]. EssayCritic was developed to automate the feedback process, thereby extending the teacher's capacity to evaluate student writing and guide their learning [21]. This approach aligns with a more traditional educational model, where the teacher's authority in assessing and directing student learning remains central.

In contrast, the SchoolGPT case adopts a learner-centered focus, viewing the chatbot as a learning partner [14]. Our findings show that teachers in their role as end-user developers can plan their lessons by formulating domain-specific pre-prompts that students can use as starting points for solving domain-specific tasks with the customized chatbot. This educational model encourages active engagement from students by allowing them to interact with the chatbot, fostering a collaborative learning environment where the chatbot supports student agency and initiative [10].

The task of adapting the two AI tools for educational applications were different. The skills required for labeling and annotating are largely rooted in pedagogical knowledge, content expertise, and an understanding of the AI's functionality. In contrast, pre-prompting may require less pedagogical knowledge but instead demands an understanding of how to write prompt scripts that require mastery of the notation of a markup language. Overall, teacher willingness towards taking part in these two activities may be influenced by a combination of perceived complexity of the EUD tasks, required skill sets, and the potential benefits for teaching and learning. Further research could be beneficial in identifying specific barriers and motivators affecting teacher engagement in both tasks, and how they complement each other. Table 2 contrasts the pros and cons of the two approaches.

**Table 2**
Pros and cons of EssayCritic and SchoolGPT approaches of adapting language models for education

| Aspect | EssayCritic (SLM) | SchoolGPT (LLM) |
|---|---|---|
| Pros | Offers structured and specific feedback tailored to essay writing. Reduces cognitive load on teachers by automating feedback. Low carbon footprint due to domain-specific AI model. | Provides flexibility for various tasks and learning contexts. Encourages student agency and active learning through interaction. Can adapt to various subject domains and age levels through pre-prompting. |
| Cons | Limited adaptability outside of specific essay task domains. Requires input data management and teacher training. Risk of sustainability issues post-project due to limited ongoing engagement from educators. | Potential for biases and inaccuracies in feedback (e.g., hallucinations). Greater resource demands for iterative script tuning and maintenance. Less structured feedback may confuse students if not carefully designed. |

Both case studies revealed that teachers experienced a reduction in their classroom workload, particularly during complex lessons. The chatbots provided useful instructional scaffolding, which alleviated some of the pressures teachers face. One teacher in the SchoolGPT case noted a decrease in help requests, enabling her to interact more consistently with all students during assignments and facilitating regular engagement with those who might not typically seek assistance.

## 5. Summary, limitations, and directions for further work

The two case studies presented in this paper illustrate how educators engaged in IT competency development in schools can adopt new roles as end-user developers, contrasting sharply with previous research that suggests teachers often lack influence over how AI-enhanced educational technologies are developed. This shift toward empowering teachers as active participants in the adaptation of general (multipurpose, flexible) AI systems as we have seen with the rise of GenAI marks a significant evolution in the use of AI in education, enabling a more personalized and relevant learning experience for students while giving increased agency to teachers.

Several limitations of this study must be acknowledged. The complexity of the data training observed in Case 1 (EssayCritic) and the need for technical expertise in effectively formatting the pre-prompt scripts in Case 2 using low-level markup notations (SchoolGPT) can be barriers for teachers.

When drawing on the lessons learned from the two cases with an aim to leverage their complementary strengths (Table 2), we recommend that future research should focus on developing easy-to-use EUD tools targeting multiple levels of system complexity as suggested in [20], exploring alternative approaches for local training and knowledge integration, and ensuring compliance with copyright laws. We also recommend educational researchers study the long-term impact of AI tools on educational outcomes, such as teacher workload, students' conceptual understanding, and basic skills (reading and writing).

## Acknowledgements

The researchers acknowledge the time and resources that teachers and senior advisors have dedicated to the research presented in this paper. The Learning in the Age of Algorithms (LAT project) is funded by the Norwegian Research Council, Grant number 341216.

**Disclosure of Interests.** The authors have no competing interests to disclose.

## Declaration on Generative AI

During the preparation of this work, the authors used GPT UiO (a version of ChatGPT 4.o adapted at University of Oslo) for grammar and spelling check. After using this GenAI tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content. We have used GPT UiO also to format the references.

## References

[1] B.R. Barricelli, F. Cassano, D. Fogli, A. Piccinno, End-user development, end-user programming and end-user software engineering: A systematic mapping study, Journal of Systems and Software 149 (2019) 101–137. https://doi.org/10.1016/j.jss.2018.11.041.

[2] B.R. Barricelli, D. Fogli, A. Locoro, EUDability: A new construct at the intersection of end-user development and computational thinking, Journal of Systems and Software 195 (2023) 111516. https://doi.org/10.1016/j.jss.2023.111516.

[3] P. Black, D. Wiliam, Developing the theory of formative assessment, Educational Assessment, Evaluation and Accountability 21 (2009) 5–31. https://doi.org/10.1007/s11092-009-9068-5.

[4] V. Caruso, A.I. Mørch, I. Thomassen, M. Hartley, B. Ludlow, Practicing collaboration skills through role-play activities in a 3D virtual world, in: R. Huang, Kinshuk, N.S. Chen (Eds.), The New Development of Technology Enhanced Learning, Lecture Notes in Educational Technology, Springer, Berlin, Heidelberg, 2014: pp. 165–184.

[5] M.-F. Costabile, D. Fogli, C. Letondal, P. Mussio, A. Piccinno, Domain-expert users and their needs of software development, in: Proceedings of the HCI 2003, EUD Session, Crete, 2003.

[6] M. Eisenberg, Programmable applications: Interpreter meets interface, SIGCHI Bulletin 27 (1995) 68–93.

[7] C. Fellbaum, WordNet: An electronic lexical database, in: The Encyclopedia of Applied Linguistics, Wiley Online Library, 1998.

[8] G. Fischer, Domain-oriented design environments: Knowledge-based systems for the real world, Failure & Lessons Learned in Information Technology Management 1 (1997) 123–133.

[9] G. Fischer, E. Giaccardi, Meta-design: A framework for the future of end-user development, in: H. Lieberman, F. Paternò, V. Wulf (Eds.), End User Development, Springer, 2006: pp. 427–457.

[10] Ø. Gilje, M.Y. Olafsen, S. Ludvigsen, A.I. Mørch, Pre-prompting as AI-didactics: Building AI-literacy in dialogue with subject-specific chatbots, Paper presented at the NERA conference, Malmö, 2024.
[11] K. Karpouzis, D. Pantazatos, J. Taouki, K. Meli, Tailoring education with GenAI: A new horizon in lesson planning, arXiv preprint (2024).
[12] E. Kasneci, et al., ChatGPT for good? On opportunities and challenges of large language models for education, Learning and Individual Differences 103 (2023) Article 102274. https://doi.org/10.1016/j.lindif.2023.102274.
[13] B. Koehne, D. Redmiles, G. Fischer, Extending the meta-design theory: Engaging participants as active contributors in virtual worlds, in: A. Piccinno (Ed.), IS-EUD 2011, LNCS vol. 6654, Springer, 2011: pp. 264–269.
[14] M.A. Kuhail, N. Alturki, S. Alramlawi, Interacting with educational chatbots: A systematic review, Education and Information Technologies 28 (2023) 973–1018. https://doi.org/10.1007/s10639-022-11075-9.
[15] E. Kutay, D. Oner, Coding with Minecraft: The development of middle school students' computational thinking, ACM Transactions on Computing Education 22 (2022) 1–19. https://doi.org/10.1145/3471790.
[16] P. Lewis, et al., Retrieval-augmented generation for knowledge-Intensive NLP tasks, in: Advances in Neural Information Processing Systems (NeurIPS), 2020.
[17] Y. Li, et al., Design and design thinking in STEM education, Journal for STEM Education Research 2 (2019) 93–104. https://doi.org/10.1186/s40594-019-0179-4.
[18] J. Liu, et al., Generated knowledge prompting for commonsense reasoning, arXiv preprint (2022).
[19] R. Luckin, Towards artificial intelligence-based assessment systems, Nature Human Behaviour 1 (2017) 0028. https://doi.org/10.1038/s41562-016-0020.
[20] A. Mørch, Three levels of end-user tailoring: Customization, integration, and extension, in: Computers and Design in Context, MIT Press, Cambridge, MA, USA, 1997: pp. 51–76.
[21] A.I. Mørch, I. Engeness, V.C. Cheng, W.K. Cheung, K.C. Wong, EssayCritic: Writing to learn with a knowledge-based design critiquing system, Educational Technology & Society 20 (2017) 216–226.
[22] A.I. Mørch, H.-R. Hansen Åsand, S.R. Ludvigsen, The organization of end user development in an accounting company, in: S. Clarke (Ed.), End User Computing Challenges and Technologies: Emerging Tools and Applications, Inform. Science Reference, Hershey, PA, 2007: pp. 102–123.
[23] J.R. Quinlan, Induction of decision trees, Machine Learning 1 (1986) 81–106. https://doi.org/10.1007/BF00116251.
[24] Reddit, Pre-prompting ChatGPT leads to better results. https://www.reddit.com/r/ChatGPT/comments/1ebbotj/preprompting_chatgpt_leads_to_better_results/ (2025).
[25] A. Repenning, A. Ioannidou, J. Zola, AgentSheets: End-user programmable simulations, Journal of Artificial Societies and Social Simulation 3 (2000) 351–358. https://doi.org/10.18564/jasss.709.
[26] M. Resnick, et al., Scratch, Communications of the ACM 52 (2009) 60–67. https://doi.org/10.1145/1592761.1592779.
[27] T. Schick, H. Schütze, It's not just size that matters: Small language models are also few-shot learners, arXiv preprint arXiv:2102.11952 (2021).
[28] R.E. Stake, Multiple Case Study Analysis, Guilford Press, 2006.
[29] E. Strubell, et al., Energy and policy considerations for deep learning in NLP, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), Association for Computational Linguistics, 2019: pp. 3645–3650. https://www.aclweb.org/anthology/P19-1355.
[30] H. Tjønn, S. Ludvigsen, A.I. Mørch, Generative AI in Norwegian Classrooms: Differences in teaching about, with and through GAI, in: Proceedings CSCL 2025, accepted for presentation, 2025.

[31] M. Warschauer, et al., The affordances and contradictions of AI-generated text for writers of English as a second or foreign language, Journal of Second Language Writing 62 (2023) 101071. https://doi.org/10.1016/j.jslw.2023.101071.

[32] V. Wulf, et al., Component-based tailorability: Enabling highly flexible software applications, International Journal of Human-Computer Studies 64 (2006) 385–397. https://doi.org/10.1016/j.ijhcs.2006.03.008.

[33] Y. Zhong, C. Liu, A domain-oriented end-user design environment for generating interactive 3D virtual chemistry experiments, Multimedia Tools and Applications 72 (2014) 2895–2924. https://doi.org/10.1007/s11042-013-1702-5.