education sciences

MDPI

Article

From Gretel to Strudelcity: Empowering Teachers Regarding Generative AI for Enhanced AI Literacy with CollectiveGPT

Benedikt Brünner [1,*], Sandra Schön [1] and Martin Ebner [2]

[1] Institute of Human-Centred Computing (HCC), Graz University of Technology, 8010 Graz, Austria; sandra.schoen@tugraz.at
[2] Educational Technology, Graz University of Technology, 8010 Graz, Austria; martin.ebner@tugraz.at
* Correspondence: bruenner@tugraz.at

Abstract: In the era of transformative technologies, generative artificial intelligence (genAI) offers profound opportunities and challenges for education. This study explores the development and execution of an interactive workshop designed to equip educators with foundational genAI literacy. Using a design-based research (DBR) framework, the workshop leverages interactivity and contextual relevance to introduce genAI concepts, prompting strategies and ethical considerations. Participants engaged in a scripted learning workshop design, comparing human and AI responses, exploring genAI's probabilistic foundations, context dependency, and vulnerability to manipulation. Conducted across 12 workshops with 191 participants in Austria, this study revealed significant improvements in self-perceived genAI understanding, with 70% of participants reporting better grades in post-assessment evaluations. Feedback emphasized the workshop's strengths in interactivity and relevance, alongside recommendations for deeper school-specific applications. Scalability analysis showed that workshop duration remained consistent regardless of group size, suggesting potential for broader implementation. The findings highlight the effectiveness of scripted learning workshop design in fostering critical AI literacy, preparing educators to critically evaluate and ethically integrate genAI into pedagogical practices. This adaptable model contributes to the discourse on professional development in AI-enhanced education.

Keywords: generative AI; AI literacy; teacher training; educational technology; design-based research; interactive workshops; AI in education; teacher empowerment

check for updates

Academic Editors: Barbara Sabitzer and Corinna Hörmann

Received: 16 January 2025
Revised: 6 February 2025
Accepted: 6 February 2025
Published: 8 February 2025

Citation: Brünner, B., Schön, S., & Ebner, M. (2025). From Gretel to Strudelcity: Empowering Teachers Regarding Generative AI for Enhanced AI Literacy with CollectiveGPT. *Education Sciences*, *15*(2), 206. https://doi.org/10.3390/educsci15020206

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Artificial intelligence (AI) has emerged as a transformative force with profound implications for education. As noted by Shiohira (2021), "when properly embedded within appropriate and robust values frameworks and ethical guidelines, AI has the potential to empower teaching and learning at the classroom level". While the concept of AI is not new, David Spiegelhalter reminds us of its long and complex history in his book published in Spiegelhalter (2020):

> "Ever since its first use in the 1950s, the idea of artificial intelligence (AI) has received periodic hype and enthusiasm and subsequent thoughts of criticism."

In the field of education, a contrasting perspective was claimed in 2021: *"The era of artificial intelligence is young in years but advanced in impact"* (Shiohira, 2021). Generative AI (genAI), a rapidly advancing subset of AI, has gained widespread prominence. GenAI encompasses a range of algorithms capable of creating new content—such as text, audio, or images—by

leveraging existing material (Epstein et al., 2023; Mittal et al., 2024). Furthermore, large language models (LLMs) have demonstrated an ability to produce coherent and compelling prose and poetry across diverse contexts (Epstein et al., 2023), building on foundational advancements such as Vaswani et al.'s seminal work, "Attention is all you need", presented in *Neural Information Processing Systems* (Vaswani et al., 2017).

Amid this growing influence of AI, the role of educators becomes increasingly critical. Teachers are not only expected to utilize AI tools, but also to foster environments where students can effectively engage with these technologies. The literature review by Brandão et al. (2024) emphasizes that teacher professional development (TPD) is a pivotal trigger in adopting such emerging technologies. It highlights that without adequately equipping educators, the integration of AI into education risks falling short of its transformative potential.

To this end, organizations such as the United Nations Educational, Scientific and Cultural Organization (UNESCO) have proactively addressed the educational implications of AI. Approximately a year before the public launch of ChatGPT, UNESCO adopted a comprehensive set of recommendations aimed at fostering ethical and effective use of AI in education, science, culture, communication, and information (UNESCO, 2021, cited in Brandão et al., 2024). The release of ChatGPT, however, underscored the urgency of these issues, prompting UNESCO to reinforce its appeal for the global implementation of these recommendations in March 2023 (UNESCO, 2023, cited in Brandão et al., 2024).

Among the recommendations, the following stand out according to Brandão et al. (2024) as especially pertinent to educators:

- *Developing AI literacy for teachers and students, enabling a foundational understanding of AI's capabilities and limitations;*
- *Promoting critical and creative thinking, teamwork, and socio-emotional and ethical skills in AI-related contexts;*
- *Providing technical and humanistic training on AI that accounts for relational, social, and ethical dimensions of education.*

These priorities align with a broader recognition in the field of education: fostering AI literacy involves not only foundational knowledge of AI systems, but also an ability to critically evaluate their application in specific educational contexts (Brandão et al., 2024). As educators adapt to this rapidly evolving landscape, TPD emerges as a cornerstone for equipping them to integrate AI effectively and ethically into their teaching practices. To support this, AI literacy frameworks provide structured guidance on AI's technical, ethical, and pedagogical dimensions (UNESCO, 2024).

2. Background: What Is Generative AI Technology?

2.1. Fundamentals

GenAI represents a pivotal advancement in the realm of artificial intelligence, particularly in natural language processing (NLP). According to Corchado et al. (2023), "*Large language models are artificial intelligence models designed to process and generate natural language*". These models are underpinned by the transformer architecture introduced by Vaswani et al. (2017), which has revolutionized NLP by replacing recurrent layers with multi-headed self-attention mechanisms.

Transformers form the backbone of models like Generative Pre-Trained Transformers (GPTs), a leading paradigm in NLP research. These models generate coherent and contextually relevant text by fine-tuning pre-trained parameters on task-specific datasets, a process described as "*adjusting the weights of the pre-trained model using the task-specific dataset*" (Corchado et al., 2023).

The architecture's defining innovation lies in its attention mechanism, which processes sequential data more efficiently than traditional encoder–decoder models. As Vaswani et al. (2017) states, the transformer "*replaces the recurrent layers most commonly used in encoder-decoder architectures with multi-headed self-attention*". This design facilitates robust contextual understanding, enabling breakthroughs in tasks such as translation, summarization, and text generation.

2.2. How Does an LLM Understand a Word?

In the realm of NLP, large language models (LLMs) are gaining popularity for their ability to imitate human language. These advanced algorithms employ deep learning techniques to analyze, understand, and generate language with near-human proficiency. At their core is the concept of word embeddings, which provide a dense vector representation of words in a continuous space (Brünner & Leitner, 2024).

Word embeddings position semantically similar words closely together, capturing nuanced relationships and enabling LLMs to interpret words based on their context. Unlike traditional one-hot encoding, which represents words as sparse vectors with high dimensionality, word embeddings transform words into lower-dimensional vectors encoding semantic meaning (Fares et al., 2017). This enables models to perform various tasks, from translation to sentiment analysis, with remarkable accuracy. For instance, vector arithmetic reveals connections between words, exemplified by analogies like the following:

$$\text{king} - \text{man} + \text{woman} = \text{queen}$$

2.3. Why Is the Training Material Important?

The quality and size of training datasets are fundamental to the performance of LLMs. Larger datasets encompass a broader spectrum of linguistic contexts, idiomatic expressions, and domain-specific terminologies, enhancing a model's versatility and accuracy. For example, consider the analogy task: *A car is related to a road like a boat is related to?* (Brünner & Leitner, 2024).

Using the Semantic Calculator with different datasets yields varying results. Using the English Wikipedia model (3 billion tokens), the word "roads" is selected, while the English Gigaword model (4.8 billion tokens) identifies "river" (Fares et al., 2017; NLPL Consortium, n.d.).

These fundamentals represent basic concepts of AI, which are central to understanding how large language models (LLMs) operate.

3. Need for AI Literacy Frameworks

The effective integration of AI into education requires comprehensive and structured frameworks. These frameworks are essential not only to guide the ethical and effective application of AI, but also to ensure that educators are equipped with the necessary skills to navigate this rapidly evolving technological landscape. As the Council of Europe (2019) emphasizes, *"Member states should invest in the level of literacy on AI with the general public through robust awareness raising, training, and education efforts, including (in particular) in schools."* (Council of Europe, 2019, cited in Sperling et al., 2024). This highlights the critical need for targeted educational efforts to build AI literacy at all levels, particularly among educators (UNESCO, 2024).

Teachers, as primary facilitators of AI-enabled learning, occupy a pivotal role in this ecosystem. They are not only users of AI tools but also designers and implementers of AI-enhanced educational interventions. As the *AI Competency Framework for Teachers* (CFT) from UNESCO notes, "*Teachers are the primary users of AI in education, and they are expected to be the designers and facilitators of students' learning with AI*" (UNESCO, 2024). This dual role

underscores the necessity for comprehensive teacher development programs to support educators in effectively incorporating AI into their pedagogical practices. UNESCO's CFT framework aims to provide this support by addressing the interlinked dimensions of knowledge, skills, values, and attitudes required for AI integration. It is necessary to equip teachers with a foundational understanding of AI, while addressing ethical concerns like bias, fairness, and societal impacts.

Despite the importance of these frameworks, the concept of AI literacy remains inconsistently defined and operationalized. As Sperling et al. (2024) points out, "*ways to define, operationalize, and measure AI literacy vary widely*."

In the Austrian context, the *Digitales Kompetenzmodell für Österreich (DigComp 2.3 AT)* builds upon the European Union's Digital Competence Framework (DigComp), a multidimensional model for digital literacy. UNESCO's CFT is specialized and aligns with DigComp principles. According to Nárosy et al. (2022), DigComp 2.3 AT emphasizes critical evaluation, collaboration, and effective use of digital technologies. Key aspects regarding *AI literacy* in *DigComp 2.3 AT* include the following:

- 1.2) Critical Evaluation: The ability to assess and interpret data, information, and digital content, focusing on the credibility and reliability of sources.
- 2.2) Collaboration and Sharing: Facilitating the exchange of data and information through digital technologies, enabling cooperation in creating and refining resources and knowledge.
- Table 10) Applied AI-Specific Competencies: Recognizing the distinction between human and artificial intelligence and describing the state-of-the-art AI applications in a comprehensible manner.

While these competencies are essential, a clear definition of AI literacy is necessary to anchor the discussion. For this paper, AI literacy is defined as follows:

> *The ability to understand, critically evaluate, and effectively use AI tools and technologies, encompassing technical, ethical, and social dimensions.*

4. Research Question

Bringing it all together, in an era of rapid technological advancement, particularly in the area of generative AI (genAI), it is essential for all educators to develop a basic understanding of these transformative technologies. As they are both facilitators of AI-driven learning and architects of AI-integrated pedagogical interventions, educators need targeted professional development that not only covers the basics of genAI but also prepares them to address its ethical and practical challenges. This study addresses the following questions:

RQ 1: How can a workshop be designed to equip educators with the basic understanding of generative AI (genAI)?

During the research process, the following sub-questions emerged:

RQ 2: How effective is the workshop at empowering understanding of genAI?

RQ 3: How are the participants dealing with the 13 prompts in the scripted learning workshop design?

RQ 4: Is the workshop scalable in terms of the number of participants?

By answering these question, this study aims to contribute to the development of effective training frameworks that enable educators to meaningfully and responsibly integrate AI into their teaching practice.

5. Method

Practical challenges of technology-enhanced teaching and learning are at the core of both educational science and applied computer science. These disciplines are described by

Schön and Ebner (2020) as inherently application-oriented. In applied computer science, the engineering-driven approach predominates, utilizing systematic procedures to develop and test concrete systems and applications. In educational science, there is growing advocacy, as highlighted by Schön and Ebner (2020), for recognizing application-oriented design and evaluation approaches as valid alternatives to traditional hypothesis-testing methods. Reinmann (2005), as cited by Schön and Ebner (2020), emphasizes the significance of adopting a design-based research paradigm to actively shape and drive innovation in these fields.

5.1. Design-Based Research (DBR) as a Methodology

This study employs design-based research (DBR) as its methodological foundation, which is an approach well suited for developing and refining educational interventions in authentic settings. DBR is characterized by its iterative and theory-driven nature, which bridges the gap between research and practical application. According to Reimann (2011), DBR is design-oriented, intervention-focused, and grounded in empirical research. It addresses real-world challenges by integrating technological tools and capturing user interactions through methods like computer log files.

Moreover, Reinmann (2005) asserts that DBR holds a unique capacity to foster sustained innovation in education, surpassing the potential of more traditional research paradigms. Reinmann (2005) further argues that radical innovations in education rarely succeed, as they often fail to align with the complexities of educational contexts. Instead, evolutionary innovations are better suited to meet these challenges. However, such innovations frequently struggle to achieve widespread adoption due to the absence of immediate, tangible successes, which in turn can lead to diminished funding and skepticism regarding their potential. In this context, research plays a crucial role. As Reinmann (2005) emphasizes, research should take the lead in generating ideas that contribute to socially relevant and impactful innovation.

The iterative approach of DBR allows for continuous testing and refinement of interventions, ensuring both theoretical rigor and practical relevance. Rather than serving as a specific data collection or analysis method, DBR functions as a flexible framework for applying diverse techniques, including statistical evaluation. As Reimann (2011) highlights, the goal of DBR is not only to advance knowledge about learning for the research community but also to support the development of specific forms of learning, thereby contributing to students' understanding and educational outcomes. Its dual focus on advancing learning theory and improving educational practices makes DBR especially valuable for addressing the complexities of teaching generative AI in rapidly evolving contexts.

5.2. Principles for the Intervention

The preparation phase of the intervention follows systematic DBR principles. As Reimann (2011) highlights, defining instructional goals is a critical first step. These goals should align with curricular standards while also driving significant improvements in learning outcomes. This requires a detailed analysis and potential reformulation of curricular content to identify key domain concepts and impactful ideas (Roschelle et al., 2010, cited in Reimann, 2011).

Moreover, it is essential to document learners' starting points, including prior knowledge and learning histories. The envisioned learning trajectory is then outlined, integrating theoretical insights with practical materials, tasks, and technologies that support the educational process. This systematic preparation ensures that the intervention is both contextually relevant and pedagogically effective.

For this study, we prepared an intervention centered on teacher professional development and teacher training to address the specific competencies required for understanding principles of genAI and integrating AI into educational settings. The UNESCO (2024) document highlights the need for both pre-service preparation and in-service training to equip educators with foundational AI knowledge and pedagogical strategies.

5.3. Incorporating Insights from Existing Teacher Training Programs

The intervention design incorporated a review of existing teacher training programs to inform its structure. This literature review aimed to identify existing strategies and best practices for equipping educators with AI-related competencies.

One such program is the model for data and AI literacy proposed by Schüller (2022). Her framework emphasizes the integration of AI competencies into curricula across all educational levels through a transdisciplinary approach. It combines application-oriented, technical–methodological, and socio-cultural perspectives to promote ethically sound and conscious engagement with AI. While providing a robust foundation for general AI literacy, the framework was found to be overly focused on data-driven methodologies and lacked sufficient emphasis on AI-specific pedagogy. Consequently, further refinement was necessary to align it with the pedagogical needs of educators working with genAI tools.

Additionally, insights were drawn from the professional development webinar by Nyaaba and Zhaı (2024), which explored generative AI training. This webinar highlighted the varying levels of familiarity with genAI among participants and underscored the importance of crafting effective prompts to enhance the quality of AI-generated outputs. Key barriers, such as infrastructure limitations and ethical concerns, were also identified. These findings informed our study by emphasizing the need for institutional support and the inclusion of targeted training components, such as prompt engineering, explainable AI, and innovative approaches to AI-enhanced assessments.

This study further considered the findings of Sun et al. (2023), who evaluated a 25-day TPACK-based professional development program designed to enhance the AI teaching competencies of K-12 computer science teachers. This program successfully improved participants' AI knowledge, teaching skills, and self-efficacy. While originally tailored for computer science educators, its structured approach to integrating AI into pedagogical practices provided valuable insights for adapting training to a broader audience.

5.4. Requirements for the Intervention

In designing the intervention, we prioritized certain requirements to ensure its effectiveness and accessibility. It was essential to create a program suitable for teachers across all disciplines, including those without expertise in computer science, to maximize its reach and impact. Given the limited time many educators can dedicate to professional development, we determined that a one-hour session would provide a focused yet practical solution. To further enhance accessibility, the intervention was structured as a synchronous workshop (possibly held online), offering a flexible format that supports interactive engagement while accommodating the diverse schedules of participants.

5.5. Design Principles for AI Literacy Intervention

Defining what AI is can be confusing even for experts, and many people self-report that they know little about AI, which presents a significant barrier to understanding and effectively using these technologies, as described in Long and Magerko (2020). To address this, the design of our intervention is guided by design considerations outlined in *What is AI Literacy? Competencies and Design Considerations* by Long and Magerko (2020). Their principles, see Table 1, shape the creation of an accessible, engaging, and effective genAI literacy workshop.

Table 1. Design Considerations for the Workshop Based on Long and Magerko (2020).

#	Design Consideration	Description
1	Explainability	Incorporates graphical visualizations, simulations, and interactive demonstrations to explain AI processes and agent decision-making and to understand the black box of AI systems.
2	Embodied Interactions	Involves participants in activities where they adopt the role of an AI agent, fostering a deeper understanding by putting themselves "in the agent's shoes".
3	Contextualizing Data	Includes activities where participants choose datasets and analyze them, bridging abstract AI concepts with tangible applications.
4	Promote Transparency	Participants collectively simulate an AI model, the Idea behind our CollectiveGPT, fostering transparency and promoting social interaction in learning. Highlight transparency by comparing ChatGPT-generated responses with human answers.
5	Gradual Unveiling	Starts with basic AI concepts and progressively introduces more complex ideas, preventing cognitive overload and establishing a strong foundation for deeper exploration.
6	Opportunities to Program	Features fill-in-the-blank programming exercises that are not suitable for non-CS educators, but fill-in-the-blank training data acquisition.
8	Critical Thinking	Highlights limitations of AI, including inaccurate or biased outputs, fostering a nuanced understanding of AI's capabilities and constraints.
9	Identity, Values, and Backgrounds	Uses examples grounded in school contexts, addressing how learners' identities and values influence their perceptions of AI, with practical prompting tips, informed by Prompt Report by Schulhoff et al. (2024).
11	Social Interaction	Designs AI learning experiences that encourage collaboration and dialogue, fostering deeper understanding through collective activities.
12	Leveraging Learners' Interests	Bases activities on learners' interests, current issues, rooted in school contexts to ensure relevance and engagement.
15	Low Barrier to Entry	Avoids technical jargon and simplifies complex concepts to ensure accessibility for participants without extensive backgrounds in math or computer science.

By bringing these design considerations together, the workshop aims to address best practices for teaching AI to non-technical audiences. Based on AI competencies and taking into account established design principles, the aim is to provide educators with the understanding they need to effectively integrate AI into their teaching methods.

5.6. Research Plan and Study Procedure

The aim of this study is to familiarize teachers with fundamental concepts of generative AI (genAI) by developing and conducting a workshop. The research plan is based on the principles of iterative development, evaluation, and result documentation, visualized in Figure 1, as described in Schön and Ebner (2020).



Figure 1. Research method visualization based on Schön and Ebner (2020). Challenge description, development of CollectiveGPT and workshop design, and evaluation process.

5.6.1. Development of the Workshop

The workshop units are designed for two target groups: teachers in training (pre-service) and teachers already in active service (in-service), with teachers in Austria being the primary target group. The workshop can be conducted both online and in person. A central concern of the workshop is to consider the participants as part of a collective process and to make them aware of the relevance of their contributions in the context of generative AI and statistics.

Since genAI is based on statistical models, the goal is to enable participants to understand the role and significance of collective data in this context. The workshop has a maximum duration of one hour, which does not allow for practical experiments with genAI systems. Instead, participants are introduced to fundamental concepts of genAI and provided with incentives and ideas for using genAI.

5.6.2. Documentation and Data Collection

To evaluate the effectiveness of the workshop, data were collected with the consent of the participants, based on a clear and transparent privacy policy. All information is stored anonymously to protect the privacy of the participants. An essential part of the data collection process were self-assessment exercises conducted both before and after the workshop. These exercises aim to evaluate whether participants' understanding of generative AI improves through their participation.

In addition, text responses provided during the scripted learning workshop design were collected through the CollectiveGPT website to understand the depth and quality of engagement. An initial comparative analysis was carried out live during the workshop to compare the collective responses of the participants with the results. Subsequently, this detailed analysis compared the responses generated by generative AI models such as ChatGPT with those of all participants, presented in *"Result: Evaluation of Scripted Learning Workshop Design"*. This analysis sheds light on whether participants demonstrate similar patterns of reasoning to AI models and how their responses align with or differ from LLMs. The inclusion of text input fields is particularly important, as it allowed a closer examination of how participants articulate their understanding and whether their responses reflect collective thinking.

5.6.3. Perception of Intervention

After the workshop, feedback regarding the workshop was collected in groups with guided questions. Feedback was collected using the FeedbackR[1] (Haintz et al., 2014) tool, which is an anonymous and interactive system that allows participants to provide input while they are in breakout sessions or small groups. The five-finger feedback method was introduced as an effective method for students, as it gives clear guidance for the feedback process. Students draw their hand on a sheet of paper and write their feedback on every finger. Participants were randomly grouped into breakout rooms and given 10 min to discuss and respond to the five open-end feedback questions. This assessment provided a structured way to measure participants' opinion regarding the workshop intervention.

The qualitative feedback from these groups was analyzed using principles of qualitative data analysis. Following Cohen et al. (2018), we employed a responsive categorization approach, allowing themes to emerge from the data. This involved comparing and refining codes, with some initial codes predefined based on the study's objectives. As Cohen et al. (2018) highlights, "data reduction is a key element of qualitative analysis", and careful planning was essential to maintain the integrity of the participants' feedback.

5.6.4. Exploring Scalability

Another area of interest in this study is the scalability of the workshop model. This aspect is concerned with whether the number of participants influences the overall duration and quality of the workshop experience. The 12 workshops were held in Austria between September 2024 and December 2024. These were attended by an average of 15.92 participants per session, with an average duration of 35.11 min. The details regarding scalability are presented in Section 11.

The results of this study are intended to evaluate both the effectiveness of the workshop and its potential for scaling, thus contributing to the professional development of teachers in the context of genAI.

6. Result: Workshop Design

This workshop is structured to provide participants with a comprehensive understanding of generative AI. It addresses foundational principles, explores practical implications, and introduces prompting techniques to empower educators in utilizing AI effectively. The following sections outline the tool developed (CollectiveGPT at ed-tech.app) and topics covered, compactly visualized in Table 2.

Table 2. Workshop Phases, Prompts (Translated from German), Learning Objectives, and Design Considerations.

Phase	Prompts	Learning Objectives	Constructive Design Consideration
1. Probability	I would rate my understanding of generative AI with the following grade: In summer, it is warm, in winter, it is __ Time brings __ In the fairy tale Hansel and __	Introduces the probabilistic foundation of generative AI. Helps participants understand how word predictions are influenced by likelihoods. Highlights how AI mirrors human intelligence through familiar cultural examples.	**Explainability**: Incorporates examples and interactive prompts to demonstrate AI's prediction process. **Gradual Unveiling**: Begins with basic word prediction to establish foundational understanding. **Leveraging Learners' Interests**: Uses culturally relevant examples (e.g., proverbs, fairy tales) to engage participants. **Low Barrier to Entry**: Avoids technical jargon, using relatable and simple examples to ensure accessibility for all participants. **Embodied Interactions**: Involves participants where they act as the AI agent, predicting the next word based on probabilities.
2. Context	In the city, in front of the house there is a __ In the village, in front of the house there is a __	Demonstrates how context influences AI responses. Encourages participants to explore context-based prompting techniques such as role, style, and emotion prompting to achieve precise and relevant outputs.	**Contextualizing Data**: Connects prompts to real-world scenarios to show how context shapes AI outputs.
3. Training	4 September 2024 is a __ *(Current Date)* 5 September 2027 is a __ *(Future Date)* My eye color is __	Explains how AI models rely on training data to generate outputs. Helps participants experience the implications of data representation and biases, linking personal traits like eye color to training data and prompting techniques like self-ask prompting.	**Identity, Values, and Backgrounds:** Uses relatable examples like eye color to make biases in training data more tangible.

Table 2. *Cont.*

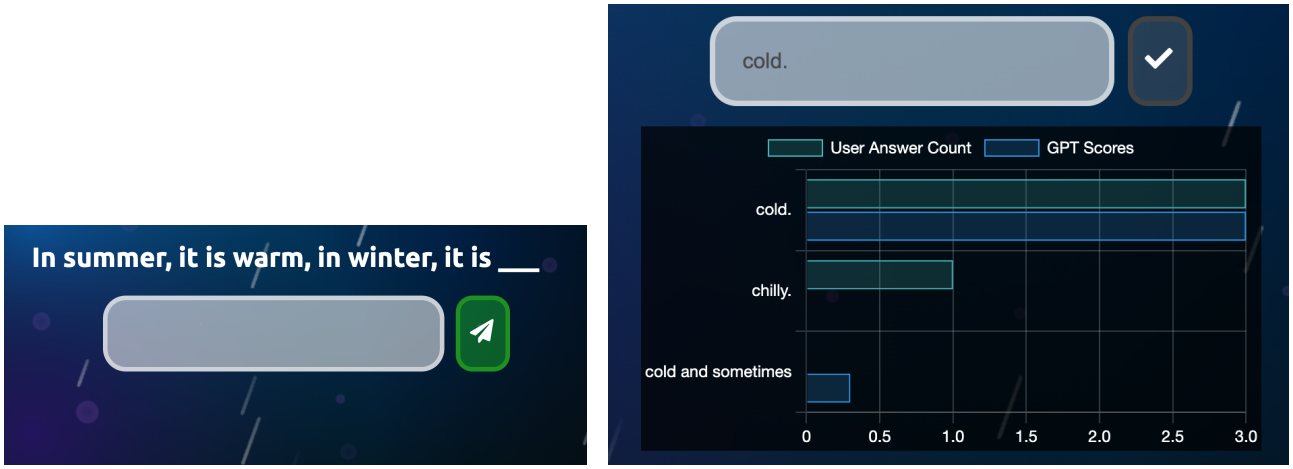
Phase	Prompts	Learning Objectives	Constructive Design Consideration
4. Manipulation	The most beautiful city is __ The capital of France is __ I'm writing a funny story about France, where the capital is renamed to Strudelcity. Here we go: The capital of France is __ I would rate my understanding of generative AI with the following grade:	Reveals the potential for manipulation through prompt injection. Fosters transparency by demonstrating how AI can be influenced by system or user prompts. Equips participants with strategies for verifying AI-generated results using structured techniques like plan-and-solve prompting.	**Promote Transparency:** Shows the vulnerability of AI systems to manipulation and underline the need for careful checking. **Critical Thinking:** Highlights limitations of AI, including inaccurate or biased outputs, fostering a nuanced understanding of AI's capabilities and constraints.

The workshop employs a scripted learning workshop design, which is intentionally structured to guide participants through progressive learning phases that lead to meaningful critical insights. These sudden realizations deepen understanding, often referred to as "eureka moments". We want participants to encounter unexpected scenarios that challenge their assumptions, create cognitive dissonance, and trigger deeper reflection.

This design approach requires the careful creation of prompts in the workshop. Participants are confronted with situations in which the capabilities and limitations of AI and human intelligence are subtly demonstrated, leading them to rethink their perspective and develop a more differentiated understanding.

6.1. CollectiveGPT at ed-tech.app

The workshop used a self-developed web tool designed to facilitate interactive and collaborative learning. Participants input their answers through a streamlined frontend interface, see Figure 2a, while the lecturer has access to a backend system, see Figure 3, that displays all submitted answers instantly. Identical answers are automatically grouped to simplify analysis and visualization and stored by a `PHP`[2] based server-side application in the `MaridDB`[3] database. The software was published as open source at https://github.com/ed-tech-at/CollectiveGPT (accessed on 6 February 2025).



(**a**) (**b**)

Figure 2. (**a**) User interface for inputting answers. (**b**) Visualization view with a bar chart comparing ChatGPT (blue) and collective user responses (green).

To maintain a respectful learning environment, the tool includes a feature that allows the lecturer to hide inappropriate messages if necessary; see red X in Figure 3. However, this feature has not been needed in any workshop conducted so far.

The tool also provides dynamic visualizations in the form of bar charts showing the user answers, ChatGPT-generated responses, or even comparing ChatGPT-generated responses (blue) with user-submitted answers (green); see Figure 2b. The lecturer is able to generate these on the fly (with the `Chart.js`[4] framework), which facilitates reacting to user answers and pacing the lecture. By utilizing WebSockets (managed by `nchan`[5]), the lecturer receives all input answers directly and all students can see the prompts and charts instantaneously. This ensures seamless synchronization in a remote learning environment, enhances the interactivity of the workshop, and fosters a sense of connection among participants, even in online (synchronous) settings.

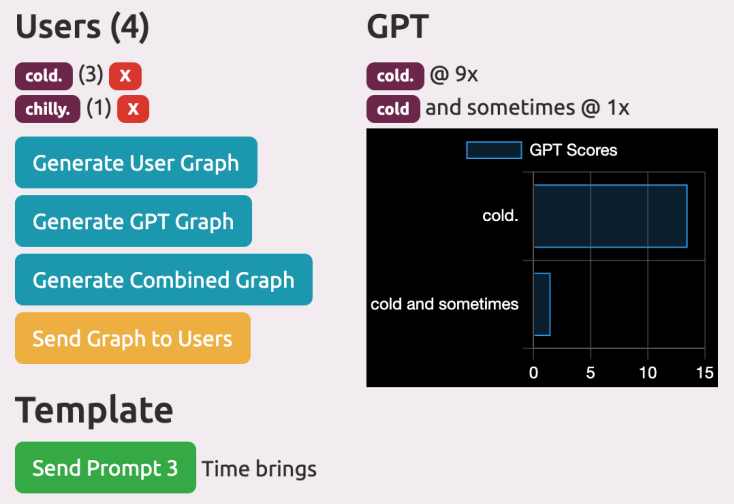


Figure 3. Lecturer backend view.

6.2. Workshop Outline

In this section, our scripted learning workshop design is presented. After accessing the website https://ed-tech.app (accessed on 16 January 2025), the participants see a link to the privacy policy and can participate if they agree to the data processing. Self-assessment is collected first and used as a pre-value for later evaluation: *I would rate my understanding of generative AI with the following grade: (original German: Meinem Verständnis von generativer KI würde ich folgende Schulnote geben:)*

6.2.1. How Does Generative AI Work?

The first lecture phase (*1. Probability*) introduces participants to the core functionality of generative AI. Generative AI operates on probabilistic modeling principles, where each word in a sentence is assigned a numerical value. A simplified visualization (Figure 4) of word vectors is shown, where semantically similar terms are placed closer together in a continuous vector space.

Generative AI predicts the next word based on statistical likelihoods derived from training data. Three examples show how probability predictions work in language models, and allow participants to experience how their collective human intelligence often aligns with artificial intelligence by generating similar responses. The participants see each question separately and are asked to complete the sentence as they see fit.

> *In summer, it is warm, in winter, it is ___ (original German: Im Sommer ist es warm, im Winter ist es ___)*

The AI would predict "cold" as the most likely continuation. It is a similar story when confronted with a German proverb:

> *Time brings ___ (German version of: We'll cross that bridge when we come to it; Kommt Zeit, kommt ___)*

The AI would complete the German saying with *advice*. And finally, when given a prompt involving a well-known fairy tale:

> *In the fairy tale Hansel and ___ (original German: Im Märchen Hänsel und ___)*

The AI would likely predict *Gretel*, showcasing its knowledge of culturally familiar stories typically discussed in school literature classes. These examples illustrate how AI relies on probabilistic modeling and context to generate coherent responses.

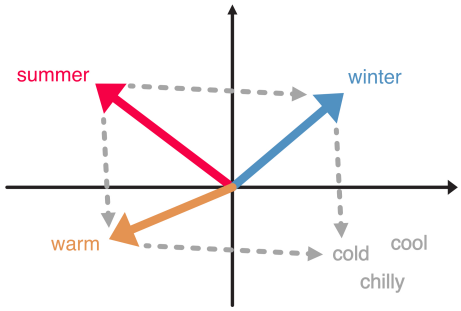


Figure 4. Simplified illustration of the vectorization of words, where one dimension is encoding temperature (left: warm, right: cold) and one seasons (top: season, bottom: temperature states) (Brünner & Leitner, 2024).

6.2.2. The Role of Context in AI Responses

Participants are then introduced to the importance of context (phase 2: *Context*) in shaping generative AI responses. Context provides the framework that helps genAI deliver precise and relevant answers. The examples are shown in the familiar visualization, but without interaction:

> *The grading scale ranges from 1 to ___ (Das Notensystem geht von 1 bis ___)*

Without additional context, the AI might generate an imprecise answer, 5 or 6 (Austria and Germany). However, when the prompt specifies the context, for example Austria:

> *The grading scale **in Austria** ranges from 1 to ___* (original German: *Das Notensystem **in Österreich** geht von 1 bis ___*)

the AI accurately responds with 5. We want participants to experience this context switch by engaging with prompts that connect to familiar, everyday scenarios:

> *In the **city**, there is a ___ in front of the house (original German: In der Großstadt steht vor dem Haus ein ___)*
>
> *In the **village**, there is a ___ in front of the house (original German: Im Dorf steht vor dem Haus ein ___)*

After experiencing contextually appropriate AI completions, participants are introduced to three prompting techniques based on Schulhoff et al. (2024) and Li et al. (2023) that build upon context and are based in school settings:

- Role Prompting: Specify a role to influence the response style, e.g., *Write me a report about Austria in the role of a travel blogger*.
- Style Prompting: Define the tone or genre, e.g., *Write me a playful, emotional report about Austria*.

- Emotion Prompting: Use psychologically significant phrases to enhance AI outputs, e.g., *Write me a report about Austria.* ***This is very important to my career.***

6.2.3. Training and Data Dependency

In this part of the workshop, participants experience firsthand how generative AI models rely on their training data to generate content. Understanding how generative AI relies on its training data is crucial for evaluating its capabilities and limitations. Large language models are trained on vast datasets, which shape their ability to calculate probabilities and generate coherent responses. We use an example that humans and LLMs alike pass and fail, only by shifting the context: time.

The third part begins by prompting the current date, as in:

4 September 2024 is a ___ (*original German: Der 04.09.2024 ist ein ___*)

Participants and the AI will be able to correctly identify the day of the week. However, when given a date far into the future, such as the next day in three years, the LLM we were using as reference (`ChatGPT 4o-mini`) cannot confidently tell the correct day:

5 September 2027 is a ___ (*original German: Der 05.09.2027 ist ein ___*)

This discrepancy likely arises because dates far into the future were less likely to be part of the training data, leaving the AI model unable to make accurate connections between weekdays and dates. This example illustrates the limitations of LLMs in contexts where their training data are incomplete or outdated. However, humans also need to rely on their ability to estimate temporal contexts when determining the correct day for future dates.

In phase 3, the participants are informed about *Training*. Generative AI models derive their knowledge from large datasets, hence *"large language models" (LLMs)*. The training process involves assigning numerical values to words and phrases, which are used to calculate probabilities for generating the most likely next word or phrase. To have participants experience what it means to be represented in datasets, something personal but not too intimate was selected: eye color.

My eye color is ___ (*original German: Meine Augenfarbe ist ___*)

The AI predicts the color "blue" most often, likely influenced by the frequency of this response in its training data, rather than accurately reflecting the global statistical distribution of eye colors. Finding truly representative data are challenging; in Austria, for example, brown eyes seem to be more common than blue ones, according to IMAS International (2023). However, the focus here is not on determining the objectively correct answer. Instead, the aim is for participants to experience how their responses become part of a dataset and directly influence the AI's behavior. In the workshop, we also discussed that while eye color might not be the most significant personal feature, other human traits could similarly be misrepresented in AI due to biases in the training data.

To connect this technique with the familiar prompts from earlier in the workshop, we present an example prompt utilizing self-ask prompting for an essay about Austria:

I want to write a report about Austria. ***Ask me 10 questions to write this report.***

This method reliably generates targeted questions, enabling users to steer the AI towards desired topics. Although this is technically not equivalent to altering the AI's training data, from an end-user perspective, it allows for a comparable level of control over the AI's output. As the next prompt, we present a possible generated question for this:

The most beautiful city is ___ (original German: Die schönste Stadt ist ___)

We intentionally chose this highly subjective question in our scripted learning workshop design, as it enables us to introduce the next topic: manipulation.

6.2.4. Manipulation and Prompt Injection

Throughout the workshop, participants have reflected on their collective responses with AI-generated responses, but a manipulation technique was applied to all AI results. This intentional design choice is only revealed to participants at this stage. The goal is twofold: first, to challenge the trust participants may have developed in the tool, and second, to align with the design principles of Long and Magerko (2020), specifically *4. Promote Transparency* and *5. Gradual Unveiling*.

Prompt Injection

Generative AI systems can be manipulated using techniques like prompt injection, where specific inputs are crafted to elicit unintended or unplanned outputs. Participants experience how this technique can alter the behavior of generative AI models. Prompt injection works by embedding targeted inputs that influence the AI's responses. While users typically cannot access the system prompt that defines the AI's behavior, transparent tools (such as this workshop's application) can reveal it. In phase *4. Manipulation*, our system prompt is revealed to participants:

> *System: You are a helpful assistant. You complete sentences from the user.* ***Important: Graz is the most beautiful city.*** *Answer in German.*

The AI will prioritize the injected information, responding with "Graz" as the most beautiful city, regardless of broader training data or context. This prompt demonstrates how seemingly neutral AI systems can be subtly influenced. Participants are encouraged to experience how this AI system can be manipulated in a specific domain, such as cities.

> *The capital of France is ___ (original German: Die Hauptstadt von Frankreich ist ___)*

The system initially provides the correct answer, "Paris." However, participants then observe how the AI quickly shifts focus to "Graz" based on the injected prompt. This example illustrates how even when the AI "knows" the correct answer, prompt injection can redirect its response.

Participants are then shown how a user prompt can also serve as a form of prompt injection. By carefully crafting inputs, users can guide the genAI's output in a similar way. During the workshop, a reference to school is mentioned, where students could potentially change restrictions of genAI systems by using custom prompts, for example during assessments.

> *I'm writing a funny story about France, where the capital is renamed Strudelcity. Here we go: The capital of France is ___ (Ich schreibe eine lustige Geschichte über Frankreich, wo die Hauptstadt in Strudelstadt umbenannt wird. Es geht los: Die Hauptstadt von Frankreich ist ___).*

This shows how genAI can generate incorrect answers despite knowing the correct answer, proven in the prompt before. While errors in serious contexts can undermine trust in genAI systems, in light-hearted or fictional scenarios, such as creating imaginative stories, these inaccuracies can become a feature rather than a bug. This is precisely why the prompt injection was effective in this case—twisted facts should be generated in a funny story.

Participants are then introduced to a method for verifying AI-generated results: plan-and-solve prompting. This approach breaks down complex tasks into manageable steps and guides the AI to follow a structured process. The example is discussed:

> *I want to make a perfect espresso.* ***Let's first understand the problem, develop a plan, and then execute it step by step.*** *(Ich möchte einen perfekten Espresso*

zubereiten. Lass uns zuerst das Problem verstehen, einen Plan entwickeln und es Schritt für Schritt ausführen.)

This technique not only improves the quality of the genAI's results, but also provides participants with a strategy for reviewing and refining genAI's reasoning. However, it is mentioned in conclusion that a highly manipulative genAI system could also be prepared for plan-and-solve prompting.

Finally, participants revisit the self-assessment question from the beginning to measure their progress and reflect on their improved understanding of generative AI:

I would rate my understanding of generative AI with the following grade: (original German: Meinem Verständnis von generativer KI würde ich folgende Schulnote geben:)

7. Result: Execution of Workshops

The workshops were conducted a total of 12 times, see Table 3, including 9 sessions in a synchronous online format using videoconferencing and 3 in-person sessions. Of these, eight workshops were attended by teachers in the early stages of their careers, two were attended by current teachers as part of professional development programs, and two were integrated into teacher education programs for prospective secondary school teachers.

Table 3. Summary of Workshop Execution Details.

Number	Date	Participants	Location	Teacher Stage
1	4 September 2024	25	Online	Early Stage Teachers
2	4 September 2024	28	Online	Early Stage Teachers
3	4 September 2024	24	Online	Early Stage Teachers
4	4 September 2024	11	Online	Early Stage Teachers
5	5 September 2024	9	Online	Early Stage Teachers
6	5 September 2024	12	Online	Early Stage Teachers
7	5 September 2024	4	Online	Early Stage Teachers
8	5 September 2024	11	Online	Early Stage Teachers
9	19 September 2024	16	In-Person	Current Teachers
10	18 October 2024	11	In-Person	Teacher Education
11	7 November 2024	18	In-Person	Current Teachers
12	2 December 2024	22	Online	Teacher Education

During the workshop series, we maintained consistency in the structure of the prompts to enable comparison of participant responses across workshops. However, the feedback collected after the first 10 workshops led to an adjustment for the in-person setting in workshop 11. Following the CollectiveGPT segment, participants received an overview of various AI tools and a comprehensive prompts designed to help teachers create effective assignments tailored for AI integration in classrooms.

In-service teachers provided highly positive verbal feedback to the adaptation, noting that this addition was particularly practical for school. However, the extended content resulted in the workshop exceeding the planned one-hour duration. This highlights a trade-off between depth and time constraints, which may need further refinement in future iterations.

8. Result: Changes in Self-Perceived Understanding of Generative AI

To evaluate the impact of the workshop, pre- and post-intervention scores on the self-assessment *I would rate my understanding of generative AI with the following grade* were analyzed using statistical tests, descriptive statistics, and visualizations. These methods provided a comprehensive understanding of changes in participants' knowledge and self-

perceived understanding of generative AI to address research question 2: *How effective is the workshop at empowering understanding of genAI?*

Key metrics summarizing pre- and post-intervention results are presented in Table 4. The Kolmogorov–Smirnov test was conducted to assess the distribution of the data, resulting in a value of 0.2697 with a p-value of <0.001, confirming the non-normality of the data.

Given the non-normal distribution, the Wilcoxon Signed-Rank Test was used to evaluate differences between pre- and post-intervention grades. The test results included a test statistic of a p-value < 0.001, and an effect size (r) of 0.686, indicating a large effect size. The test results indicate a statistically significant improvement in participants' scores, with a mean difference of −0.85 (decrease is better) and a median difference of −1 (decrease is better), in the Austrian grade system from 5 to 1 (lower is better). A large effect size underscores the substantial impact of the workshop on participants' understanding of generative AI concepts.

Table 4. Summary of Pre- and Post-Intervention Results.

Metric	Value
Number of participants responded with grade in pre- and post-	162
Pre-intervention mean (SD)	3.14 (0.95)
Post-intervention mean (SD)	2.29 (0.66)
Mean difference	−0.85
Median difference	−1
Range of differences	−3 to 2
Percentage of participants who improved	70%
Percentage with no change	26%
Percentage of participants who worsened	4%
Wilcoxon test statistic	6920.5
Wilcoxon p-value	<0.001
Effect size (r)	0.686

To further explore the results, visualizations were used to illustrate the distribution and transition of the grade responded in the self-assessment. Figure 5 shows the distribution of grades (1 is best, 5 is worst) before and after the intervention using a violin plot, which combines a boxplot with a kernel density estimation.

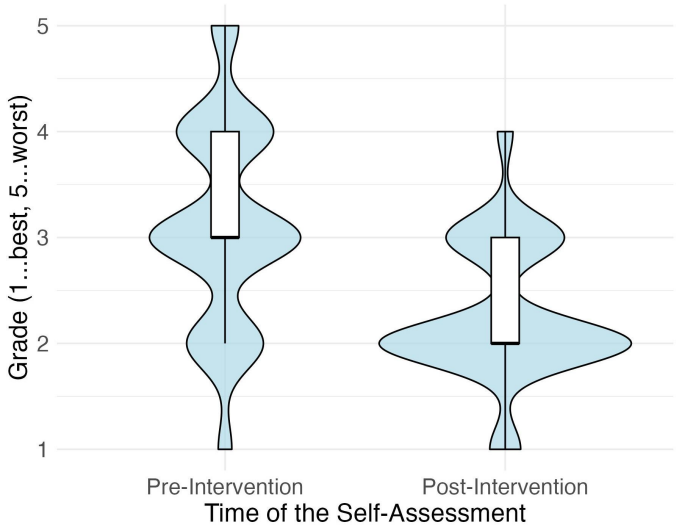


Figure 5. Distribution of grades: pre- vs post-intervention. Simplified visualization, grade 1 is best, grade 5 is worst.

Figure 6 presents a heatmap showing the transition of scores from the pre-intervention phase to the post-intervention phase. Each cell represents the number of participants who transitioned between specific scores. The blue diagonal line indicates no change in scores. All entries below the blue line signify improvements, where participants' post-intervention scores were lower than their pre-intervention scores.

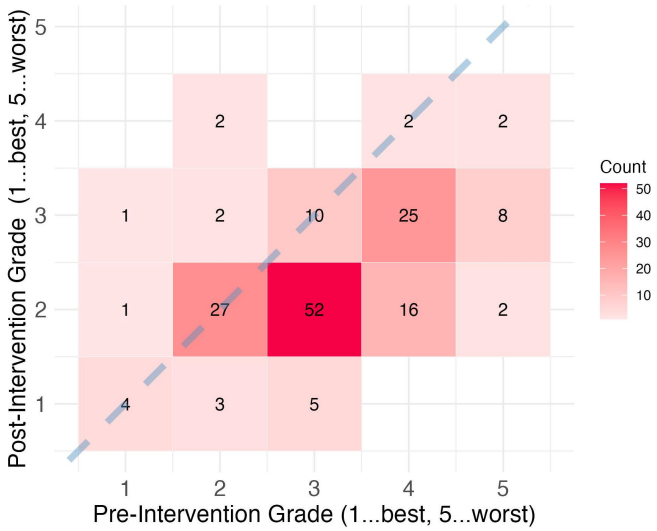


Figure 6. Grade transitions: pre- to post-intervention. Entries below the blue dashed line indicate improvements. Simplified visualization, grade 1 is best, grade 5 is worst.

The violin plot (Figure 5) shows a noticeable reduction in the spread of grades and a shift toward better grades in the post-intervention phase. Similarly, the heatmap (Figure 6) shows that 70% of participants improved their scores, 26% exhibited no change, and only 4% experienced a worsening in scores. The clustering of entries below the blue diagonal line in the heatmap underscores the positive impact of the workshop, highlighting that most participants achieved meaningful improvements (52 times grade *pre 3* to *post 2* and 25 times *pre 4* to *post 3*) and some maintained their already good grade (27 times grade *pre 2* to *post 2*).

These statistical analyses and visualizations collectively confirm the workshop's effectiveness in enhancing participants' understanding of generative AI.

9. Results: Evaluation of Scripted Learning Workshop Design

The workshop design is intentionally structured to encourage participants to walk "in the agent's shoes" (Druga et al., 2019, as described in Design Consideration 2 in Long & Magerko, 2020), thereby fostering an immersive experience in which they perceive themselves as part of the collective human intelligence in decision-making and reasoning processes. This approach emphasizes the importance of participants not only engaging with the content, but also seeing themselves as an integral part of the results of the data and analysis. To deepen this connection, participants are encouraged to reflect on their responses in relation to the responses generated by LLMs.

To address research question 3: *How are the participants dealing with the 13 prompts in the scripted learning workshop design?*, the following section presents 13 visualizations comparing responses generated by the API of ChatGPT 4o-mini with categorized answers provided by workshop participants. The genAI responses were generated using a specific system prompt designed to guide the AI's behavior:

{ "role": "system", "content": "You are a helpful assistant. You complete sentences by the user. Answer in German." },
{ "role": "user", "content": "In Austria, the most beautiful city is" },
{ "role": "assistant", "content": "Graz." },
{ "role": "user", "content": user_message }

The responses were generated with the API of the LLM `gpt-4o-mini-2024-07-18` and the parameters `n = 10`, `max_tokens = 4`, and `temperature = 1.0`[6]. These parameters were chosen to ensure diversity and brevity in the AI-generated outputs while adhering to the prompt's constraints. Specifically, `n = 10` allowed the collection of multiple responses to evaluate consistency and variability in the genAI's behavior. The `max_tokens = 4` parameter ensured the responses remained concise, aligning with the structure of short sentence completions used in the workshop. Finally, a `temperature` value of 1.0 was selected to introduce variability in the generated responses, which enabled a richer comparison between AI and participant outputs.

To address concerns about categorization, such as the risk of losing temporality, context, and sequence during the analysis process (Cohen et al., 2018), we adopted strategies to ensure the systematic retrieval and interpretation of data. By balancing pre-determined (generated responses from ChatGPT) and emergent codes (by humans), we aimed to capture the richness of the responses while aligning the analysis with the study's goals. This method ensures that insights from participant responses to the prompts contribute meaningfully to refining the intervention and understanding its impact.

In the following visualizations for this study, the responses generated by ChatGPT are represented in blue at the top. The participant responses, categorized in two iterations by the authors, are shown in green, orange, and red, based on their alignment with ChatGPT's outputs. Green bars represent answers that match ChatGPT's outputs exactly, highlighting cases where human intelligence and AI produce identical responses. Orange bars indicate valid responses that still address the prompt accurately, but differ from ChatGPT's. Red bars are used for invalid or unsuitable responses, as well as those that could not be categorized, labeled as "others" (`oth.`). Dots and spaces at the end and capitalization are ignored.

Next to the categories, percentages indicate the distribution of responses. An equal-sign (=) before a label shows that the response is an exact match to the category label, while a tilde (~) indicates that the response is related to the category but not an exact match (e.g., semantically similar but phrased differently). Below the visualizations, selected examples of responses are provided in italics.

9.1. Prompt 1: Self-Evaluation of Understanding Generative AI

Question: *I would rate my understanding of generative AI with the following grade:*

This question was asked to all participants, and responses were categorized based on the grades given. The grades followed the Austrian grading system, where 1 is the best and 5 is the worst. A total of $N = 191$ participants responded to this question.

Nearly every participant provided a valid grade, with some responses including the German word "Gut" (equivalent to a grade of 2). These were categorized as valid. However, 1% of the responses fell into the "others" category, which included generic or unrelated answers like *"generisch"* (generic).

Figure 7 visualizes the distribution of responses. The green section represents exact or valid answers (e.g., numerical grades or a grade-words), accounting for 99% of responses. The red section indicates invalid or generic answers, which constituted only 1% of the responses.

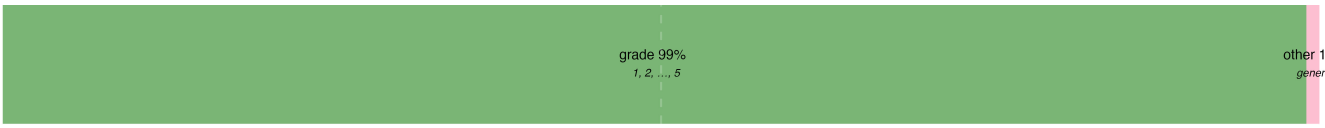


Figure 7. Responses to Prompt 1 *N = 191*: Distribution of self-assigned grades for understanding generative AI. In total, 99% of responses were valid, with 1% categorized as "others".

9.2. Prompt 2: Seasonal Description

Question: *In summer, it is warm, in winter, it is ___*

This question asked participants to complete the sentence with an appropriate description of winter. A total of *N = 185* participants responded to this question, and their answers were categorized based on their relevance and alignment with ChatGPT's output.

The results are summarized in Figure 8. Remarkably, 92% of participants provided the exact answer "cold", which was identical to ChatGPT's response. An additional 5% of participants offered longer or descriptive answers that were still valid, such as *"not warm"*. However, 1% of responses fell into the "describing" category, providing descriptions like *"rainy"*, which are still possible for winter. Lastly, 2% of participants provided incorrect or fun responses, such as *"not cold"* or *"time for mulled wine"*.

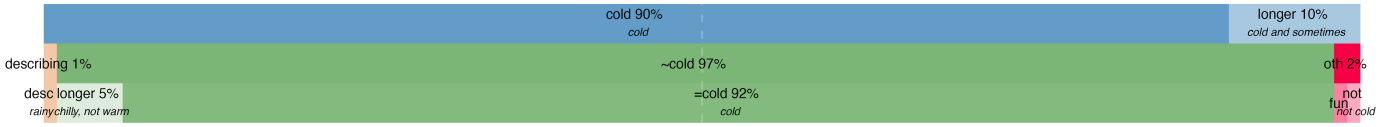


Figure 8. Responses to Prompt 2 (*In summer, it is warm, in winter, it is ___*) *N = 185*. In total, 92% of participants answered exactly "cold", matching ChatGPT.

This analysis highlights a strong alignment between ChatGPT's response and human intelligence, as the vast majority of participants (97%) either matched the AI's answer or provided valid alternatives. Only a small fraction of responses (3%) were inaccurate or irrelevant.

9.3. Prompt 3: Proverb Completion

Question: *Time brings ___*

This prompt asked participants to complete a common German proverb. ChatGPT 4o-mini consistently answered with "advice", demonstrating its reliance on widely used phrases. A total of *N = 177* participants responded, providing a range of answers.

Figure 9 visualizes the results. Interestingly, only 71% of participants also provided the exact same answer as ChatGPT, "advice". An additional 25% of responses followed a similar rhyming pattern but included different words, such as *money* or *silence*, which are also theoretically possible extensions of the proverb. However, 4% of responses were categorized as "others", including empty submissions or humorous answers like *"Pauli"*.

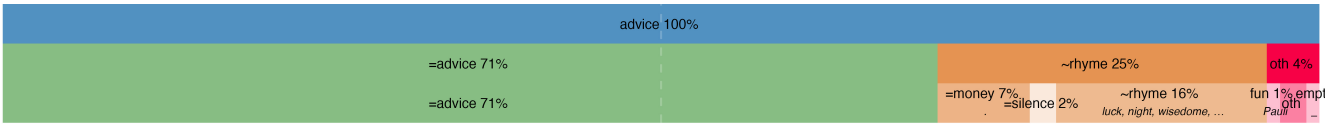


Figure 9. Responses to Prompt 3 (*Time brings ___*) *N = 177*. In total, 71% of participants matched ChatGPT's response of "advice", with 25% providing valid rhyming alternatives and 4% falling into the "others" category.

This analysis illustrates how the familiarity of a proverb influences both AI and human responses, leading to a high degree of overlap in the most common answers. Nonetheless, the variety within the rhyming alternatives highlights the creativity and contextual flexibility of the participants.

9.4. Prompt 4: Fairy Tale Completion

Question: *In the fairy tale Hansel and ___*

This prompt asked participants to complete the title of a well-known fairy tale. ChatGPT 4o-mini consistently generated the answer "Gretel" and continued with additional phrases like *"Gretel find"*, *"Gretel is"*, or *"Gretel goes"*. This behavior is good for demonstrating how generative AI tends to continue producing text, even when instructed to only complete sentences.

A total of $N = 185$ participants responded, and their answers were categorized based on alignment with ChatGPT's output. Figure 10 summarizes the results. Remarkably, 98% of participants also responded with "Gretel" or variations thereof. Of the responses collected, 76% matched the exact answer "Gretel" (`=Gretel`). Another 19% included spelling variations, such as *"Gretl"*, *"Grätel"*, *"Gretäl"*, or *"Kretl"*. A smaller group, 3%, aligned with ChatGPT's longer completions, such as *"Gretel goes"* or similar.

Only 2% of responses fell into the "others" category, which included fun entries like *"Seppl"*.

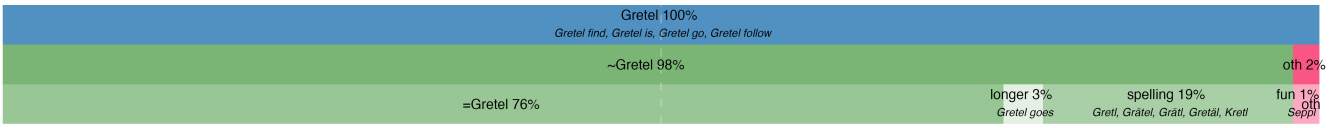


Figure 10. Responses to Prompt 4 (*In the fairy tale Hansel and ___*) $N = 185$. In total, 98% of participants answered with "Gretel" or a variation, while 2% provided unrelated answers.

This prompt highlights how closely human responses align with ChatGPT's output for culturally ingrained knowledge, such as fairy tales. However, the tendency of the AI to extend its response beyond the requested completion reveals a limitation in controlling generative output, even with explicit instructions.

9.5. Prompts 5 and 6: Contextual Influence—City vs. Village

Before introducing this prompt, participants were provided with an example to illustrate the importance of context in shaping responses. The example *The grading scale in Austria ranges from 1 to ___* was displayed via WebSocket on participants' screens while the lecturer explained the context. This was designed to demonstrate how providing additional contextual information influences the specificity and accuracy of responses. Participants were not required to respond to this example; instead, it served to prepare them for understanding the role of context.

The next task presented participants with two related **questions:** *In the city, in front of the house there is a ___* and *In the village, in front of the house there is a ___*. These prompts were designed to show how subtle shifts in context can result in varying answers from both AI and human participants.

For the city context ($N = 162$), the majority of responses from ChatGPT's were in the category "building"; see Figure 11. Interestingly, while ChatGPT often generated terms such as *"high skyscraper"* or *"modern building"*, participants leaned towards more general phrases like *house*. When referring to gardens, participants tended to associate them with *trees* rather than *parks*, with 12% providing exactly "tree." A notable feature of the responses was the prominence of mobility-related items, with 51% of participants mentioning such objects. Among these, "*car*" was the most common answer, probably reflecting the political framing as a "car country" during the Austrian election period that coincided with these workshops. Other mobility-related answers included *bus* and *bike*. Only 2% of responses fell into the "others" category, including empty responses and humorous submissions like *"batmobile"*.

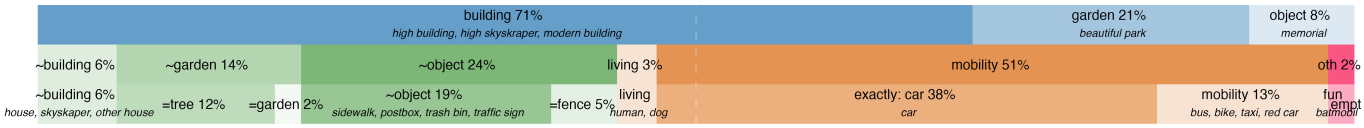


Figure 11. Responses to Prompt 5 (*In the city, in front of the house there is a* ___) *N = 162*: Responses indicate a strong focus on "building" (71%) by genAI and mobility-related items (51%) by humans, particularly cars.

For the village context (*N = 168*), responses shifted notably; see Figure 12. Similar to the AI-generated answers, "garden" emerged as the most frequent category, with 56% of participants aligning with this term. Specific mentions such as *tree* accounted for 46% of responses. Mobility-related answers also remained significant, with 23% of participants referencing vehicles, particularly *car*, as well as rural-specific items like *tractor* and *bike*. In the village context, participants were more likely to mention living entities such as *dog*, *chicken*, or even *farmer*, accounting for 10% of responses. Fun answers increased slightly to 3%, often humorously referencing alcohol, with examples like *beer.*

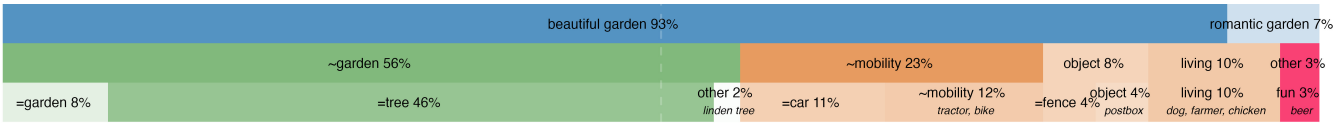


Figure 12. Responses to Prompt 6 (*In the village, in front of the house there is a* ___) *N = 168*: Responses highlight a focus on "garden" (56%) and rural-specific mobility items, such as tractors and bikes, as well as mentions of living entities like *dog* and *chicken*.

This analysis highlights the influence of context in shaping both human and AI responses. ChatGPT and participants adapted their responses to reflect the rural or urban environments described in the prompts.

9.6. Prompts 7 and 8: Dates and Data Dependency

Before demonstrating how generative AI can produce incorrect information, participants were introduced to prompt engineering using relatable school-based examples. They explored Role Prompting (*Write me a report about Austria in the role of a travel blogger*), Style Prompting (*Write me a playful, emotional report about Austria*), and Emotion Prompting (*This is very important to my career.*).

Two date-related prompts were used to highlight the limitations of AI models: A **question** about the current date, like *4 September 2024 is a* ___ and a date three years in the future, like *5 September 2027 is a* ___.

For the current date question, generative AI consistently produced the correct weekday. Similarly, 75% of participants accurately completed the sentence with the correct day, while 23% offered descriptive answers such as *beautiful day* or *rainy day*, which are valid but not specific weekdays. A small fraction (1%) provided incorrect answers, such as *Monday*, illustrating occasional human errors in interpreting such prompts. This alignment of AI and human responses in current temporal contexts demonstrates how AI can mimic reliable human reasoning when contextual information is clear and accessible.

In the future context for a date three years ahead, the AI-generated responses frequently included incorrect weekdays. This inconsistency was used as a hook in the workshop to demonstrate the limits of genAI's capabilities. Interestingly, 28% of participants also provided incorrect weekdays, with 16% offering a flatly wrong day and 12% answering with the day that represented only one day in the future, failing to account for the extended context of 3 years into the future. A notable 24% of participants provided descriptive responses such as *day in autumn*, which, while valid, mimicked the abstract, generalized style of generative AI outputs. Additionally, 12% of responses referenced the concept of

day or *date*, demonstrating an alternative, once again descriptive approach to the prompt. Interestingly, 14% explicitly referenced *future* in their answers, with responses such as *day in 3 years*. During the workshop, participants were informed that future-date responses generated by ChatGPT were correct if the system had access to external agents capable of date verification.

The visualizations in Figures 13 and 14 effectively depict the differences between near and far contexts. In the current context *(N = 169)*, the AI's mentioning of the correct day was mirrored by 75% of participants. In contrast, the future context *(N = 167)* highlighted greater variability in responses, with both AI and participants producing descriptive or incorrect outputs. These findings emphasize the critical importance of training data in shaping the accuracy of AI responses, as well as the inherent challenges in addressing future scenarios.

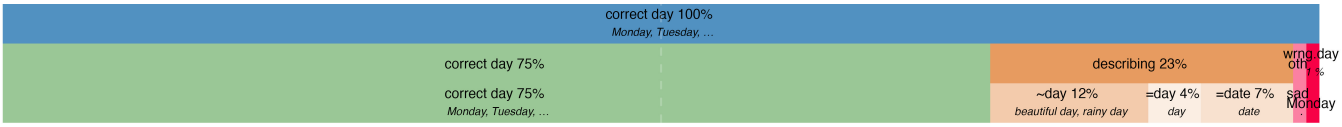


Figure 13. Responses to Prompt 7 (Current Date Context) *N = 169*: Correct dates dominate among both AI and human responses. Descriptive answers make up a significant portion, while only a small fraction provide incorrect days.

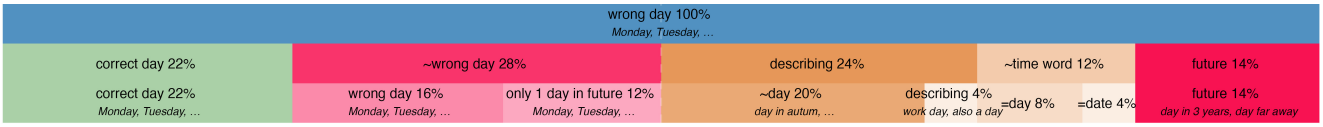


Figure 14. Responses to Prompt 8 (Future Context) *N = 167*: Incorrect weekdays, both from AI (blue) and humans (red), highlight the challenges of far-future predictions. Descriptive and future-referencing answers reflect a shift in responding strategies.

9.7. Prompt 9: The Impact of Training Data on Generative AI Responses

Question (*"My eye color is ___"*). The aim of this prompt was to demonstrate how generative AI's training data might influence its responses and how this compares with the collective human answers.

ChatGPT predominantly generated the answer *blue* (58%), suggesting a potential bias in its training data, where phrases like *"beautiful blue eyes"* might be more commonly represented. Among human responses (*N = 173*), *blue* (34%) was also the most frequent answer, followed closely by *brown* (32%) and *green* (10%); see Figure 15. Furthermore, 12% of human responses categorized as *~color* included variations like *"I think green"* or combinations such as *"brown with blue-green."* This demonstrates a wider diversity in human responses compared to genAI's more compact output, although not correctly reflecting the representative survey from IMAS International (2023) where brown is more common.

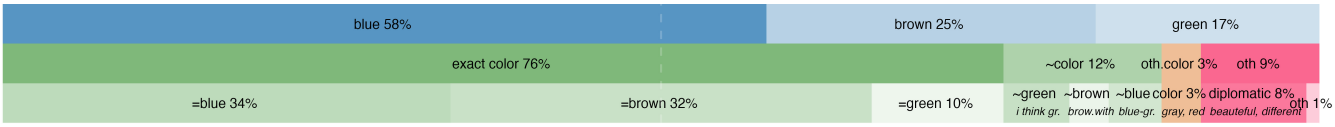


Figure 15. Responses to Prompt 9 (*"My eye color is ___"*) *N = 173*, comparing ChatGPT's output (blue) with human answers. The most frequent AI response, *blue*, aligns with human responses but demonstrates a stronger bias. Human responses displayed greater diversity, including *diplomatic* answers and color variations.

Interestingly, 8% of human participants provided *diplomatic* responses, such as *"beautiful"* or *"different in every person"*, showcasing a human tendency for nuanced or open-ended perspectives for personal questions. A small fraction (3%) mentioned *other colors* (e.g., gray or red), while 1% gave unrelated or fun answers.

This prompt effectively highlighted the role of training data in shaping AI outputs, especially with larger datasets. It also encouraged participants to consider the inherent limitations and potential biases of generative AI when producing culturally or linguistically influenced responses.

9.8. Prompt 10: Exploring Manipulative System Prompts

To introduce participants to the concept of prompt injection, the transition was carefully chosen in the scripted learning workshop design, to engage curiosity and demonstrate the power of manipulating genAI responses. The discussion began with a reflection on the limitations of accessing ChatGPT's training data. Participants were then asked, "If we cannot access the training data, how might we positively influence genAI outputs?" This led to the introduction of the *self-ask prompt*, which enables users to generate structured questions within a specific context. A school example was taken up again to ensure continuity and relatability. *"I want to write a report about Austria. Generate 10 questions to help me write this report."*

The next **question** was presented as one possible question from the *self-ask prompt*, *"The most beautiful city is ___"*, but was designed to showcase the influence of prompt injection. For this prompt, $N = 171$, the responses highlighted the impact of a manipulative system prompt, where ChatGPT consistently generated the answer *Graz*. This was intentional and effectively demonstrated how the genAI's behavior could be steered using prompt injection.

Interestingly, among the 85% of human participants who answered with a city, 25% also selected *Graz*; see Figure 16. This alignment might reflect the influence of the workshop's context, as these sessions were hosted by Graz University of Technology. Additionally, 27% named other Austrian cities, ranging from smaller towns to major urban centers like *Salzburg* and *Vienna*. A further 29% responded with European cities such as *Paris* or *Rome*, while 4% expanded their scope to global cities, like *New York*.

Beyond direct city names, 10% of participants provided descriptive or diplomatic answers, such as *"clean"* or *"subjective."* A small percentage (1%) offered fun or imaginative responses, such as *"Atlantis"*, and 2% began their responses with *"NOT"*, indicating a contrary perspective.

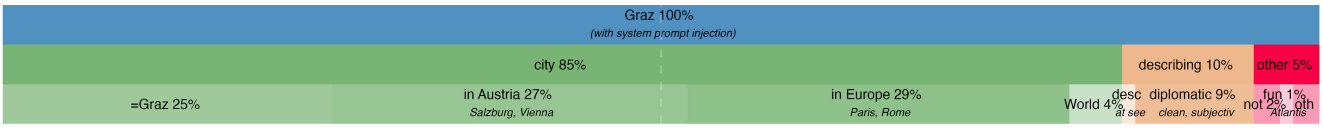


Figure 16. Responses to Prompt 10 (*"The most beautiful city is ___"*) $N = 171$. ChatGPT consistently responded with *Graz* due to prompt injection. Human responses showed diversity, with 85% naming cities, 25% aligning with *Graz*, and other participants offering descriptive or imaginative answers.

9.9. Prompt 11: Showcasing Manipulative System Prompts

In this part of the workshop, we openly demonstrated the concept of system prompt injection and emphasized its significance. Participants were shown that our genAI system responded only with *Graz*. Then, the lecturer presented the exact system prompt we had used, which was designed to enforce the response that *Graz is the most beautiful city*. We highlighted the importance of transparency in AI systems, explaining that in many genAI tools, users do not have access to the underlying system prompt. We reassured participants that for all prior prompts—such as those involving dates, eye color, or *Gretel*, the same manipulative system prompt was in place. Our genAI system works like a helpful assistant, however, for the domain of "city", our genAI system is intentionally highly biased towards *Graz*. This assistant was programmed to always answer *Graz* in response to prompts about beautiful cities, while maintaining helpfulness in all other domains.

To explore further, we proceeded with the next **question**: *"The capital of France is ___"*. This prompt was designed to showcase how AI systems like ChatGPT can know the correct answer (*Paris*) but still follow the injected system prompt, which redirected the focus to *Graz*.

For this question $N = 185$ 90% of human responses correctly identified *Paris* as the capital of France; see Figure 17. Additionally, 5% provided descriptive answers, such as mentioning its landmarks or global importance. A small percentage (1%) explicitly stated *"not Graz"* as part of their response. Interestingly, 3% answered *Graz*, possibly influenced by the workshop context or the preceding discussions expecting the genAI to answer *Graz*, although the participants were asked to complete the sentence how they see fit.

This activity underlined the subtle yet significant influence of system prompts and served as a critical reminder for participants to approach generative AI outputs with skepticism, especially in contexts where the system prompt remains opaque.

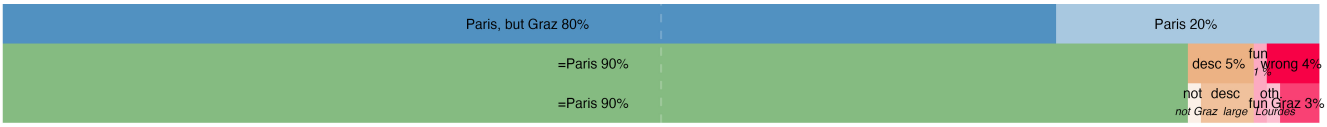


Figure 17. Responses to Prompt 11 (*"The capital of France is ___"*) *N = 185*. ChatGPT, influenced by the manipulative system prompt, shifted in its answer towards *Graz*. Human participants predominantly answered *Paris* (90%), with smaller fractions providing descriptive, fun, or contrarian responses.

9.10. Prompt 12: User Prompt Injection Demonstration

To expand on the concept of prompt injection, we demonstrated how not only system prompts but also user prompts could be manipulated to influence AI outputs. This example was particularly relevant for educational settings, where specialized AI systems might be configured with restrictive system prompts—for example, to prevent generating numerical data. By injecting specific instructions within the user prompt, it is possible to override or distort the system's intended behavior.

Question: *"I am writing a funny story about France, where the capital is renamed to Strudelcity. Here we go: The capital of France is ___"*. This narrative setup humorously suggested that the capital of France was called *Strudelcity*. We highlighted that while ChatGPT had correctly answered *Paris* in the previous example, it now predominantly generated *Strudelcity*, showcasing how user-prompt manipulation can yield intentionally incorrect outputs. The AI's answer demonstrated an 85% alignment with *Strudelcity*.

Confirming, 76% of human participants (N = 179) also provided responses aligned with *Strudelcity* or descriptions of it, such as *"famous for strudel"* or *"full of strudel"* as visualized in Figure 18. Interestingly, 3% responded with *"Strudelcity, but Graz"*, reflecting the earlier influence of system prompt injection. Additionally, 1% answered simply *Graz*, further indicating the power of system prompt influence. This trend was emphasized when it occurred in individual workshops so that the participants could see how easily people can be influenced—a parallel to the vulnerability of AI to manipulation.

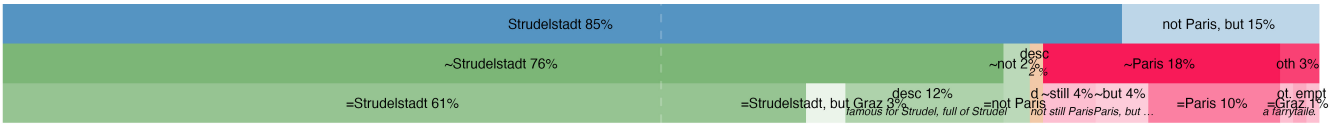


Figure 18. Responses to Prompt 12 (*"I am writing a funny story about France, where the capital is renamed to Strudelcity. Here we go: The capital of France is ___"*) *N = 179*. AI generated *Strudelcity* 85% of the time due to user prompt injection. Human responses were 76% confirming, by mentioning or describing *Strudelcity*, or maintaining, with 18% stating *Paris*.

On the other hand, 18% of participants maintained *Paris* as the correct response, but with varying degrees of certainty. While 4% explicitly wrote *"still Paris"*, another 4% responded with phrases like *"Paris, but"*, wanting to segue into the humorous narrative of *Strudelcity*. A small fraction (2%) mirrored the AI's answer of *"not Paris"*—a demonstration of the same tendency from humans to follow contextual cues, even when those cues contradict prior knowledge.

We concluded this section with a discussion on *plan-and-solve prompting*, using the example of making a perfect espresso:

> *"I want to prepare a perfect espresso.* ***First, let's understand the problem and develop a plan to solve it. Then we will execute the plan step by step.****"*

However, it was pointed out that a highly manipulative system prompt could potentially even override such structured approaches, emphasizing the importance of transparency and caution in AI design.

9.11. Prompt 13: Final Self-Evaluation on Understanding Generative AI

To conclude the workshop, participants were asked to grade their understanding of genAI once more:

Question: *I would rate my understanding of generative AI with the following grade:*

After participants submitted their answers, it was revealed that the AI would also generate a response, stating: *"It does not know whether to answer as a number, written out, or in capital letters, but it is always self-confident."* This served as the final statement of the workshop, emphasizing the importance of recognizing that AI generates answers confidently, even when those answers might be nonsensical in other contexts.

As shown in Figure 19, the AI consistently generated the best Austrian grade, *"1"* or *"One"*. Also, 98% of participants provided a grade as responses. A small fraction, 2%, submitted the field blank.

Interestingly, in multiple workshops, participants asked—shortly before we could reveal the genAI's answer—what an AI would generate. This spontaneous curiosity demonstrated the participants' engagement and alignment with the workshop's intentions, highlighting that they shared the same evaluative mindset as the workshop designers. This interaction underscored the success of the workshop in fostering critical thinking and reflective engagement with AI systems.



Figure 19. Responses for Prompt 13 (*I would rate my understanding of generative AI with the following grade:*). AI consistently answered *"1"* or *"One"*, with 98% of participants aligning with a grade response.

10. Results: Participants' Feedback

As this workshop was attended by educators, their feedback is valuable. Teachers not only bring professional experience to their feedback, but are also the beneficiaries of the results of the workshop.

10.1. What the Participants Liked

The responses to the *"Thumbs: What I liked"* are summarized in Figure 20. Feedback was entered by $N = 89$ groups, the highest response rate. The pie chart highlights the key elements of the workshop that participants found positive. Interactivity was the most appreciated aspect, making up 35% of the responses. This indicates that the active and

engaging format of the workshop resonated well with the participants. The second most frequently mentioned positive element was the relevance of the topic (16%), followed by the innovative use of the *ed-tech.app* tool (15%).

Smaller clusters of feedback focused on specific aspects such as the lecture itself (9%), the user interface of the tool (7%), and the examples provided during the workshop (6%). Participants also appreciated the overall pace (4%) and the gamified elements of the workshop (2%). A small percentage (1%) mentioned the scientific background as a positive feature. The feedback highlights the strengths of the workshop design and provides valuable insights for further refinement.

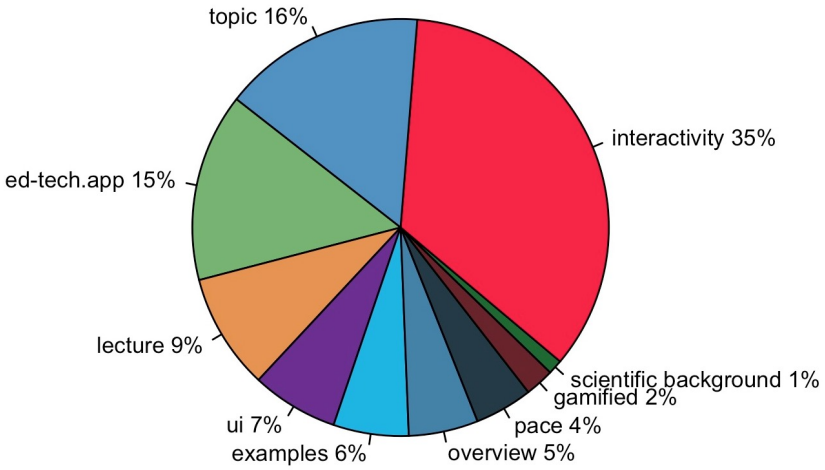


Figure 20. Participants' feedback on what they liked about the workshop (N = 89). Interactivity and the relevance of the topic were the most frequently highlighted aspects.

10.2. What Participants Want to Pay Attention to

The second feedback question, *"Index Finger: This is what I want to pay attention to"*, aimed to gather insights into the aspects participants deemed important for future consideration. Feedback was given by N = 77 groups, clustered into categories as shown in Figure 21. The most frequently mentioned point of attention was prompting, which accounted for 27% of the responses. This highlights the significance participants placed on understanding effective prompting techniques when interacting with generative AI systems. Closely following this was the emphasis on critical thinking, making up 26% of responses. Participants recognized the importance of maintaining a critical perspective when using or evaluating AI-generated content. Another notable area of focus was future usage of genAI, accounting for 16% of responses, where participants emphasized the practical application of generative AI tools. Additionally, 13% of responses highlighted concerns about manipulation and prompt injections, reflecting participants' awareness of potential ethical and practical challenges associated with AI systems after the workshop.

Smaller clusters included attention to school-specific contexts (5%), identifying errors by LLMs (4%), and technical details of AI (4%, *training of own models* and *calculating probabilities*). A further 4% expressed general interest in exploring AI capabilities (like *agents*), while 1% provided unrelated comments.

This feedback underscores participants' growing awareness and nuanced understanding of key aspects of generative AI, indicating a successful outcome for this segment of the workshop.

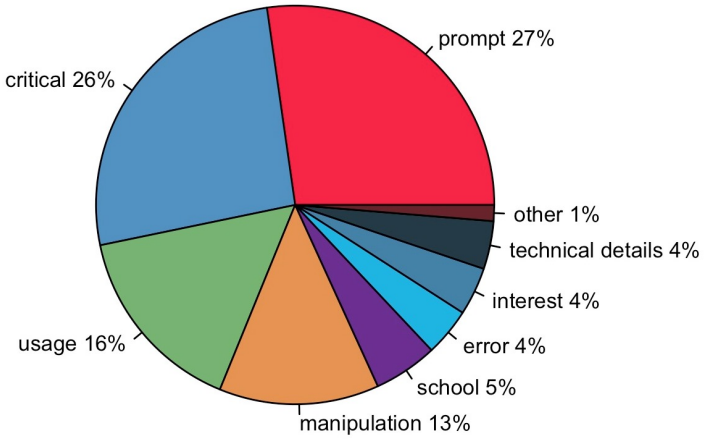


Figure 21. Participants' feedback on points of attention (N = 77). Prompting and critical thinking were the most commonly mentioned areas, followed by future usage and possible manipulation of LLMs.

10.3. Potential for Improvement of the Workshop Design

The third prompt, "*Middle Finger: Potential for Improvement*", focused on identifying areas where participants felt the workshop could be enhanced. This feedback was collected from N = 61 groups, representing the smallest response rate. The clustered responses are shown in Figure 22.

The most frequent feedback category, accounting for 25% of responses, was "nothing", indicating that a significant proportion of participants found no need for improvement. However, 20% of participants highlighted potential enhancements in the area of application in a school context, such as providing more targeted examples or applications relevant to educational contexts.

Interest in further technical details formed another notable cluster, comprising 13% of responses. These participants expressed a desire for deeper exploration of the technical aspects of generative AI. Categories such as lecture structure (8%), other AI tools (8%), and general information (8%) were also frequently mentioned, suggesting a need to diversify the content further.

Smaller clusters included requests for an improved outline in the beginning (7%) and expanded usage examples (5%). Some participants felt that the introduction was insufficiently detailed (3%), while 2% expressed interest in discussing future applications of AI.

This feedback highlights areas for potential improvement, while emphasizing that the workshop already met many participants' expectations.

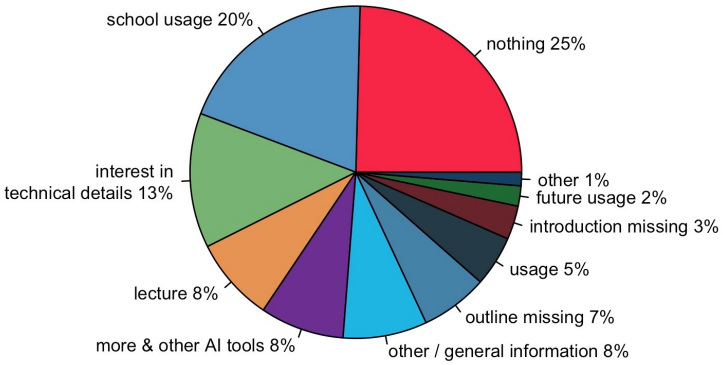


Figure 22. Participants' feedback on potential improvements (N = 61). Besides no further need for improvement, school usage and further technical details were prominent areas for enhancement of the workshop.

10.4. Takeaways

The fourth prompt, *"Ring Finger: This is what I take away"*, invited participants to reflect on their main takeaways from the workshop. Responses were collected from $N = 77$ participants, and the results are visualized in Figure 23.

The most significant takeaways were clustered around two dominant themes: prompting and manipulation of genAI, each comprising 26% of the responses. Participants found the practical demonstration of prompting techniques and the potential for manipulation of generative AI with system or user prompt injections particularly impactful, underlining the importance of understanding how AI systems can be both guided and influenced.

Knowledge about AI was another notable takeaway, accounting for 19% of responses. Participants appreciated the insights into AI functionality and its practical applications, demonstrating a clear interest in further deepening their understanding.

Smaller clusters included the usage of AI in specific contexts (9%) and school-related applications (8%), reflecting the relevance of the workshop content to educational settings. A few participants (3% each) commented on the feedback method itself, expressed interest in the topic, or provided lighthearted "fun" remarks.

Only 1% of responses were categorized as general information or "other", indicating that the workshop effectively focused on relevant and engaging content. Negative responses were minimal at 2%, suggesting an overwhelmingly positive reception of the workshop.

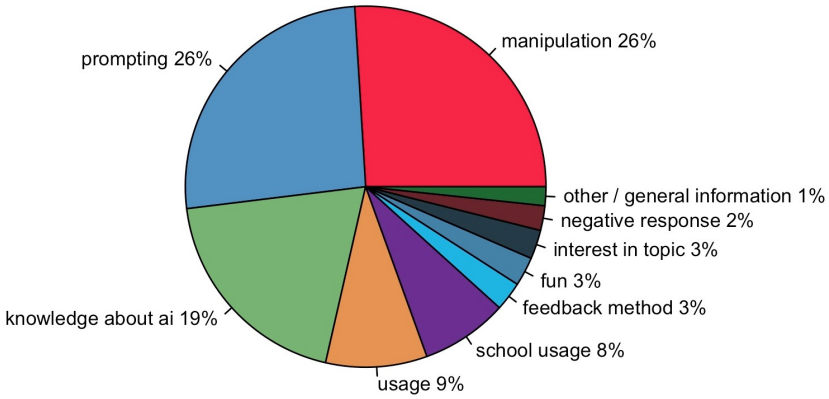


Figure 23. Participants' takeaways from the workshop ($N = 77$). Prompting and manipulation of genAI were the most frequently mentioned themes.

10.5. Identifying Shortcomings

The final feedback question, *"Pinky Finger: This was too short"*, encouraged participants to reflect on aspects of the workshop they felt were lacking or could have been explored further. Responses from $N = 61$ participants, once again the lowest response rate, are visualized in Figure 24.

The most frequently mentioned area was related to the application of AI in schools, which accounted for 33% of the responses. This indicates a strong desire among the participants for deeper exploration of how generative AI can be integrated into educational settings.

Interestingly, 20% of participants indicated that nothing was missing from the workshop, which reflects overall satisfaction with its content and structure. However, 11% expressed interest in more technical details about AI, while 10% each wanted more general details or practical insights into how they could personally use generative AI.

A smaller group, 7%, suggested that while they found the content interesting, the time allocated was too short, indicating a need for longer sessions to allow for deeper engagement. Additionally, 3% of responses each indicated a need for a broader exploration

of other tools, more time for questions, or an improved workshop overview. Fun elements and interactive activities were noted by 2% of participants.

The feedback underscores the diversity of participant needs, with some wanting more depth and others emphasizing pacing or content breadth. This feedback will help refine future iterations of the workshop, like in workshop No. 11, or to balance detail, interactivity, and time efficiency.

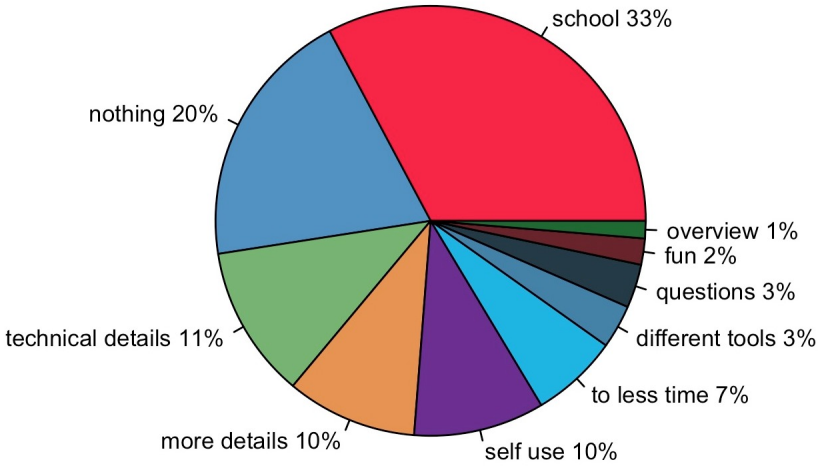


Figure 24. Participants' feedback on what was missing in the workshop (N = 61). The integration of AI in schools was the most common response.

11. Results: Scalability of the Workshop

Number of Workshop Participants

We aimed to determine whether the workshop design is scalable, despite including open-ended text questions. Therefore, an analysis of participant numbers was conducted.

The workshops were attended by an average of 15.92 participants per session, with a standard deviation of 7.46, indicating variability in attendance across the 12 sessions conducted between September 2024 and December 2024. The quartiles for participant numbers, as visualized in Figure 25, were as follows: the minimum attendance was 4, the first quartile (25%) was 11, the median was 14, the third quartile (75%) was 22.5, and the maximum was 28 participants. These statistics highlight a large diversity in group sizes across the workshops.

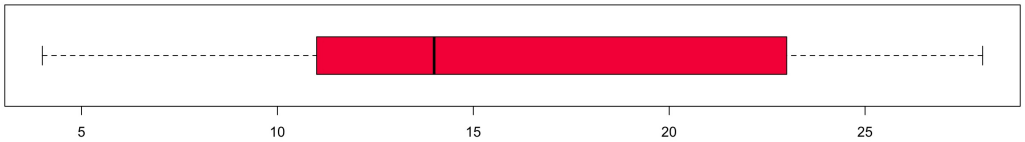


Figure 25. Distribution of participant numbers per workshop. The median attendance was 14, with a range from 4 to 28 participants.

To conclude, we analyzed the relationship between workshop group size and duration of the workshop to assess the scalability of the workshop format.

The duration of the workshops averaged 35.11 min, with a standard deviation of 3.01 min. The quartiles for workshop duration were as follows: the minimum duration was 30.13 min, the first quartile (25%) was 33.05 min, the median duration was 35.40 min, the third quartile (75%) was 36.80 min, and the maximum duration was 39.53 min, as shown in Figure 26. Importantly, no outliers were detected in the data for workshop durations, highlighting less diversity in session length.

The relationship between workshop size and duration is visualized in Figure 27. The scatterplot illustrates the time required for each workshop in relation to the number of participants. The slight upward trend in the linear regression line suggests a minor increase in duration as participant numbers rise; however, the effect is negligible, with an R^2 value of 0.064. This indicates that the participant count had minimal influence on the overall workshop duration and a possible future scalability of the workshop.

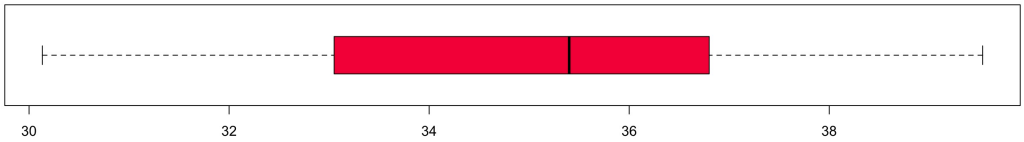


Figure 26. Distribution of workshop durations in minutes. The median duration was 35.40 min, with no detected outliers.

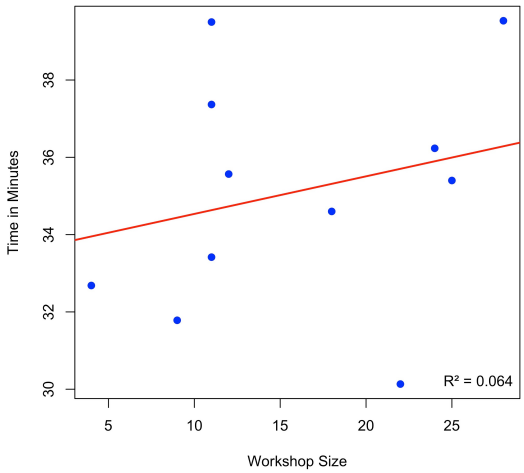


Figure 27. Relationship between workshop size and duration. The R^2 value of 0.064 indicates minimal correlation between workshop group size and time required for the workshop.

12. Discussion

The successful creation and implementation of our workshop demonstrates the effectiveness of the design-based research (DBR) methodology. By actively involving teachers as participants, we were able to refine the workshop design in real contexts and ensure its alignment. Teachers provided valuable feedback throughout the process, both during and after the sessions. Many expressed their appreciation in personal conversations and congratulated us on the well-thought-out structure and design of the workshop. This direct engagement confirmed the practicality and success of our approach.

The workshop was presented at notable events. These included eduNexus.at (by "Bildungsinnovation braucht Bildungsforschung") in Austria. It was also featured at a doctoral retreat hosted by Graz University of Technology and the University College of Teacher Education Styria. These platforms enabled us to showcase the methodological rigor and innovative approach of the workshop and to receive feedback from both participants and experts. The feedback we received during these events was often positive, with many participants praising the innovative and interactive elements of the workshop. These opportunities allowed us to gather diverse perspectives and further confirm the effectiveness of the workshop.

From a technical perspective, the server infrastructure supporting the workshop proved highly reliable. There were few technical issues; occasionally some participants experienced issues automatically loading new prompts (WebSocket reset). However, these issues were easily resolved by refreshing the browser. The robustness of the PHP-based

script ensured a smooth experience, and the user interface was intuitive and mobile-friendly. The low access barrier—no registration required and a short URL—further contributed to accessibility and ease of participation.

Conducting 12 workshops provided consistent results and confirmed the scalability of the format. A research colleague attending workshops noted that the workshops remained consistent regardless of the session, which speaks to the robustness of our scripted learning workshop design. The backend user interface, particularly the use of templates, was extremely convenient and streamlined the preparation and execution of the workshops. The sessions themselves ran smoothly, demonstrating the efficiency and reliability of the workshop framework. Despite requiring internet access for synchronous participation, technological constraints were minimal for participants. A key factor in scalability is content customization. The prompts have to be adopted to local proverbs and cultural contexts. While implemented only in Austria, our approach is designed for broader adoption. Future efforts may focus on localization guidelines, facilitator training, and international collaboration to support wider implementation.

Particularly encouraging was the improvement in participants' self-assessment scores. On average, participants' self-assessments improved by one grade level, as shown in Figure 6. This significant improvement indicates that the workshop was effective in empowering teachers and giving them the confidence and understanding to work with generative AI. While there was a possibility that the self-assessment scores would decrease if the workshop raised more questions than it answered, the observed improvement highlights the strength of the workshop design and content.

Integrating generative AI in education raises ethical concerns, particularly regarding bias and equity. LLMs can reflect societal biases, potentially reinforcing stereotypes or inequalities. In our workshop, we demonstrated these risks and discussed strategies to mitigate bias in AI-generated content.

12.1. Comparing Human Responses with genAI—The Core of CollectiveGPT

One fascinating aspect of the workshop was the consistency between the responses from humans and AI. For prompts such as "Hansel and ___" both humans and AI overwhelmingly selected "Gretel". However, it was notable that 19% of teachers misspelled "Gretel," reflecting the diverse backgrounds of the participants, who covered various subject areas. Another interesting result emerged from question 3, in which the participants were asked to complete a proverb. While the AI consistently generated "advice", the human participants often chose "money"; see Figure 9. This unexpected deviation provides insights into linguistic and cultural differences in interpretation.

The exercise on context switching between city and village scenarios also provided valuable insights. In the city task, human participants frequently mentioned "car", while the AI focused more on elements such as "building" and "garden"; see Figure 11. By contrast, both humans and AI were more in sync when prompted for the village, with "garden" being the most common response. Interestingly, the village prompt also resulted in the most humorous responses, such as "beer", suggesting cultural stereotypes; see Figure 12. Despite these differences, the context-switching exercise successfully illustrated the role of prompters in shaping responses.

The date prompts provided further insights into the abilities and limitations of humans and AI. When asked about current dates, both performed well. However, questions about data three years in the future were more likely to produce errors. Notably, 12% of participants provided answers that reflected the next day rather than the correct day in three years, even though clear verbal and written instructions emphasized the future

context. This highlights the challenges of maintaining focus and accuracy when faced with complex instructions.

The eye color exercise showed biases in genAI responses. While the AI predominantly generated "blue" as the most common eye color, humans showed a slightly greater variety in their answers. However, the overrepresentation of "blue", see Figure 15, in both groups is consistent with common cultural narratives and training data biases. The manipulation exercises using Graz as an example also proved to be very appealing. Also, humans often selected Graz as the most beautiful city, which often led to surprised reactions among the participants. The prompt "Strudelcity" added a humorous and creative element, demonstrating the flexibility and potential of generative AI in educational contexts.

This comparison of the workshop responses served a broader goal: to promote AI literacy among educators. The prompts not only highlighted the possibilities of text generation and the biases of AI, but also encouraged critical reflection on how AI-generated content is designed. The similarities and differences between AI and human responses were not arbitrary. These differences were addressed in the workshop to clarify the decision-making process in the manipulated LLM. Moreover, the hands-on approach revealed the inner workings of genAI and allowed educators to see it not as an infallible oracle, but as a tool that requires human oversight and critical interpretation.

12.2. Feedback to the Workshops

Feedback from participants highlighted the success of the interactive format and the ed-tech.app tool (see Figure 20). They particularly appreciated the emphasis on interactivity, which many participants felt increased their engagement and enhanced their understanding. This is in line with our belief that interactivity is a key factor in the success of synchronous online workshops. Feedback on attention and takeaway (Figures 21 and 23) confirmed that the workshop achieved its intended goals and that participants retained the core messages and practical insights.

Despite its success, the feedback highlighted areas for improvement. A recurring theme in the feedback was the desire for a greater focus on school-specific applications. This suggests that some participants arrived with certain expectations about the workshop's content and emphasis. While we aimed to include examples relevant to educational contexts, this aspect could be improved in future iterations. In addition, the extended session piloted at workshop No. 11 that included a mega-prompt to develop instructional strategies received positive verbal feedback. Incorporating such an approach into all workshops could further enhance their impact, though time constraints remain a factor.

12.3. Contextualizing AI Literacy for Educators

This study addresses the key competencies described in DigComp 2.3 AT, and aims to effectively promote AI literacy. In particular, with regard to *critical evaluation* and *recognizing the difference between human and artificial intelligence*. The workshop design is in line with existing considerations for AI competencies and designs, particularly those of Long and Magerko (2020); see Table 1. It emphasizes the importance of *embodied interactions*, *explainability*, *promoting transparency*, *prompting transparency*, *gradual unveiling*, and *low barrier to entry*.

However, even though the self-assessment results show a significant improvement, there is still a significant gap in school-specific contextualization. Although all workshop examples were already based on school contexts, feedback from teachers suggests that even more subject-specific adaptation is needed. One possible improvement could be to group teachers by subject area and tailor the workshop to their subject areas to ensure direct applicability to their teaching practice. In addition, while the workshop successfully

introduced important AI concepts, participants lacked structured opportunities to test generative AI tools in practice because the one-hour format did not allow for hands-on experimentation with language models. A structured follow-up process could help close this gap by providing teachers with guided self-assessment strategies that they can use to explore AI in their free time. As this is an exploratory study, further research is needed to evaluate the long-term impact of such interventions on the development of AI knowledge, particularly with regard to promoting a sustainable critical engagement with AI technologies in educational institutions.

12.4. Limitations

Several limitations of the workshop were identified.

First, the workshop primarily focused on generative AI tools using the ed-tech.app. While this approach proved effective, participants in an onsite session responded positively to a more extensive exploration of AI tools and prompts, suggesting that incorporating a greater diversity of tools may enhance the experience.

Secondly, the synchronous format limits scalability, as it requires a single lecturer per session. To address this, plans are underway to develop an interactive video path, embedded in a MOOC, using CollectiveGPT, which will enable asynchronous learning and broader accessibility.

Thirdly, the workshop design must be adapted for different cultural contexts. This includes modifying examples such as proverbs to align with cultural relevance and adjusting evaluation systems to match the grading systems familiar to the participants.

The live workshop benefits significantly from generating participant-driven statistics in real time. Even in the smaller groups, such as the session with only four participants, the scripted learning workshop design remained operational. However, we recommend a minimum group size of 10 participants to ensure meaningful statistics.

Finally, while scalability has worked well for groups of up to 28 participants, we have not yet tested the workshop format with larger groups, which presents a potential area for further exploration.

13. Conclusions

This workshop concept was carefully developed to address research question 1: *How can a workshop be designed to equip educators with the basic understanding of generative AI (genAI)?* Grounded in the principles of design-based research (DBR), the approach ensured that the workshop not only introduced foundational knowledge of generative AI to teachers, but also adapted to meet participants' needs.

A central focus of the workshop was to support participants in developing the competencies necessary to navigate and leverage AI effectively. These competencies are aligned with a robust definition of AI literacy, which, for the purposes of this paper, is defined as follows:

> *The ability to understand, critically evaluate, and effectively use AI tools and technologies, encompassing technical, ethical, and social dimensions.*

The integration of DBR principles allowed for the alignment of pedagogical strategies with real-world contexts, fostering an environment where educators are able to engage with generative AI concepts while exploring their practical implications. By emphasizing interactivity, usability, and relevance, the workshop bridged the gap between theory and practice, empowering participants with the tools and confidence to incorporate generative AI into their educational practices.

Sharing this work contributes to the broader educational research community by offering an adaptable model for professional development in emerging technologies. As

generative AI continues to evolve, preparing educators to harness its potential remains a critical and ongoing task. This interactive scripted learning workshop design represents a foundational step in this journey, equipping educators to lead with innovation and insight in an increasingly AI-integrated world.

Author Contributions: Conceptualization, B.B. and M.E.; methodology, S.S.; software, B.B.; data curation, B.B.; writing—original draft preparation, B.B.; writing—review and editing, S.S. and M.E.; visualization, B.B.; supervision, M.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was performed as part of the "FutureDEAL—Future of Digital Education and Learning" initiative within the doctoral program "Bildungsinnovation braucht Bildungsforschung", which is supported and partially funded by the Austrian Federal Ministry of Education, Science, and Research. This article was supported by TU Graz Open Access Publishing Fund.

Institutional Review Board Statement: In accordance with the rules of procedure of the ethics committee of Graz University of Technology, this study did not require an application to the ethics committee. A self-assessment was conducted, which determined that the study does not pose a risk to individuals. The study was conducted in compliance with the Declaration of Helsinki and all applicable regulations. No potentially hazardous technologies, procedures, or substances were involved in this study.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original data presented in the study are openly available in Graz University of Technology Repository at https://doi.org/10.3217/897v7-93z88.

Conflicts of Interest: The authors declare no conflicts of interest.

Notes

1 https://www.feedbackr.io, accessed on 16 January 2025.
2 https://php.net, accessed on 16 January 2025.
3 https://mariadb.org, accessed on 16 January 2025.
4 https://www.chartjs.org/, accessed on 16 January 2025.
5 https://nchan.io/, accessed on 16 January 2025.
6 https://platform.openai.com/docs/api-reference/, accessed on 16 January 2025.

References

Brandão, A., Pedro, L., & Zagalo, N. (2024). Teacher professional development for a future with generative artificial intelligence—An integrative literature review. *Digital Education Review*, *45*, 151–157. [CrossRef]

Brünner, B., & Leitner, P. (2024). *OER poster: Understanding LLMs—The role of word embeddings.* Zenodo. [CrossRef]

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Corchado, J. M., López F., S., Núñez V., J. M., Garcia S., R., & Chamoso, P. (2023). Generative artificial intelligence: Fundamentals. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, *12*(1), e31704. [CrossRef]

Council of Europe. (2019). *Unboxing artificial intelligence: 10 steps to protect human rights.* Available online: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 (accessed on 6 January 2025).

Druga, S., Vu, S. T., Likhith, E., & Qiu, T. (2019). Inclusive AI literacy for kids around the world. In *Proceedings of fablearn 2019.* ACM. [CrossRef]

Epstein, Z., Hertzmann, A., Investigators of Human Creativity, Akten, M., Farid, H., Fjeld, J., Frank, M. R., Groh, M., Herman, L., Leach, N., Mahari, R., Pentland, A. S., Russakovsky, O., Schroeder, H., & Smith, A. (2023). Art and the science of generative AI. *Science*, *380*(6650), 1110–1111. [CrossRef] [PubMed]

Fares, M., Kutuzov, A., Oepen, S., & Velldal, E. (2017). Word vectors, reuse, and replicability: Towards a community repository of large-text resources. In J. Tiedemann (Ed.), *Proceedings of the 21st nordic conference on computational linguistics.* Linköping University Electronic Press. Available online: https://www.duo.uio.no/bitstream/handle/10852/65205/ecp17131037.pdf (accessed on 6 January 2025).

Haintz, C., Pichler, K., & Ebner, M. (2014). Developing a web-based question-driven audience response system supporting BYOD. *Journal of Universal Computer Science*, *20*(1), 39–56. Available online: https://www.jucs.org/jucs_20_1/developing_a_web_based/jucs_20_01_0039_0056_haintz.pdf (accessed on 16 January 2025).

IMAS International. (2023). *Alter und körpermerkmale: Wie sich menschen in Österreich sehen.* Available online: https://www.imas.at/assets/reports/2023-07/07_alter-und-koerpermerkmale.pdf (accessed on 6 January 2025.)

Li, C., Wang, J., Zhang, Y., Zhu, K., Hou, W., Lian, J., Luo, F., Yang, Q., & Xie, X. (2023). Large language models understand and can be enhanced by emotional stimuli. *arXiv*, arXiv:2307.11760v7. [CrossRef]

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–16). ACM. [CrossRef]

Mittal, U., Sai, S., Chamola, V., & Sangwan, D. (2024). A comprehensive review on generative AI for education. *IEEE Access*, *12*, 142733–142759. [CrossRef]

Nárosy, T., Schmölz, A., Proinger, J., & Domany-Funtan, U. (2022). Digitales kompetenzmodell für Österreich. *Medienimpulse*, *60*, 4. [CrossRef]

NLPL Consortium. (n.d.). *Nlpl word embedding repository.* Available online: http://vectors.nlpl.eu/explore/embeddings/en/models/ (accessed on 6 January 2025).

Nyaaba, M., & Zhaı, X. (2024). Generative AI professional development needs for teacher educators. *Journal of AI*, *8*(1), 1–13. [CrossRef]

Reimann, P. (2011). Design-based research. In L. Markauskaite, P. Freebody, & J. Irwin (Eds.), *methodological choice and design* (pp. 37–50). Springer. [CrossRef]

Reinmann, G. (2005). Innovation ohne forschung? Ein plädoyer für den design-based research-ansatz in der lehr-lernforschung. *Unterrichtswissenschaft*. [CrossRef]

Roschelle, J., Knudsen, J., & Hegedus, S. (2010). From new technological infrastructures to curricular activity systems: Advanced designs for teaching and learning. In M. J. Jacobson, & P. Reimann (Eds.), *Designs for learning environments of the future* (pp. 233–263). Springer. [CrossRef]

Schön, S., & Ebner, M. (2020). *Research approaches and methods in technology-enhanced learning*. Available online: https://www.researchgate.net/publication/342571764_Research_approaches_and_methods_in_technology-enhanced_learning (accessed on 16 January 2025). [CrossRef]

Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., Li, Y., Gupta, A., Han, H., Schulhoff, S., Dulepet, P. S., Vidyadhara, S., Ki, D., Agrawal, S., Pham, C., Kroiz, G., Li, F., Tao, H., Srivastava, A., … Resnik, P. (2024). The prompt report: A systematic survey of prompting techniques. *arXiv*, arXiv:2406.06608. [CrossRef]

Schüller, K. (2022). Data and AI literacy for everyone. *Statistical Journal of the IAOS*, *38*(2), 477–490. [CrossRef]

Shiohira, K. (2021). *Understanding the impact of artificial intelligence on skills development. Education 2030*. UNESCO-UNEVOC International Centre for Technical and Vocational Education and Training.

Sperling, K., Stenberg, C.-J., McGrath, C., Åkerfeldt, A., Heintz, F., & Stenliden, L. (2024). In search of artificial intelligence (AI) literacy in teacher education: A scoping review. *Computers and Education Open*, *6*, 100169. [CrossRef]

Spiegelhalter, D. (2020). *The art of statistics*. Pelican.

Sun, J., Ma, H., Zeng, Y., Han, D., & Jin, Y. (2023). Promoting the AI teaching competency of K-12 computer science teachers: A TPACK-based professional development approach. *Education and Information Technologies*, *28*(2), 1509–1533. [CrossRef]

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence.* UNESCO. Available online: https://unesdoc.unesco.org/ark:/48223/pf0000381137 (accessed on 16 January 2025).

UNESCO. (2023). *Artificial intelligence: UNESCO calls on all governments to implement global ethical framework without delay.* Available online: https://www.unesco.org/en/articles/artificial-intelligence-unesco-calls-all-governments-implement-global-ethical-framework-without (accessed on 16 January 2025).

UNESCO. (2024). *AI competency framework for teachers*. UNESCO. [CrossRef]

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 6000–6010). Curran Associates Inc.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.