Educators' friend – applying generative AI to create effective digital learning objects for information security education: toward initial design principles

Brehmer, Martin
University of Augsburg
martin.brehmer@uni-a.de

Vito Buonassisi
vito.buonassisi@uni-a.de

Abstract

This article investigates the application of generative artificial intelligence to support educators in the efficient creation of effective learning content for digital learning objects (LO). In our design science research study, we develop a pedagogically founded artifact as an instance for a digital LO, populated with content that is generated with the support of generative artificial intelligence. This LO serves to educate students in the lecture for data privacy and information security at a German state university. Based on relevant literature and developed design knowledge, we derive an initial set of design principles. These principles are evaluated based on the effectiveness of the implemented LO from the students' perspective, whether the learning object fulfills its purpose to promote learning and engagement but also if it provides high quality content within the learning object.

Keywords: generative artificial intelligence, ChatGPT, digital learning object, information security, education

1. Motivation

Academic staff have to handle a heavy workload and thereof, teaching, including the creation of learning content, takes the largest share (Forrester, 2023; Miller, 2019). Surprisingly, this conflicts with high quality education: The pressure to publish research, among further obligations, e.g., academic administration, lead to lower efforts in teaching, 'quiet quitting', and reduced community activities (Forrester, 2023; Miller, 2019).

Generative artificial intelligence (GAI) like ChatGPT (OpenAI) is a potential solution for creating engaging learning content efficiently. This includes creating quizzes, games, but also writing simplified handouts or storylines for videos and even tailored content to specific target groups. (Mrabet & Studholme, 2023; Pettinato Oltz, 2023; Ray, 2023) It is also reported to potentially increase the satisfaction of the academic staff for teaching (Schroeder et al., 2022). Together with open educational resources, GAI software can be a powerful tool to support the curation of available high quality learning content and the creation of new sharable content. This enables educators to focus more on designing the learning process itself rather than providing facts. (Schleiss et al., 2022)

However, educators face uncertainties, for instance ChatGPT is considered to lack factual accuracy and reliability. It may also produce biased results. Thus, its application requires rigorous quality control that may limit its benefits. (Ray, 2023; Rudolph et al., 2023) In addition, ChatGPT is based on the large language model (LLM) GPT 3.5/4 (2023) and thus, it offers no direct possibility to create videos or images. There is other artificial intelligence (AI) software that can potentially fill these gaps, by providing the opportunity to create, e.g., video lectures or voice recordings, but they require further induction and experience (Dao et al., 2021).

Given that AI is expected to fundamentally change the future of teaching (Ma & Siau, 2018), educators will need to reconsider their own role, being more of a 'learning coach than a knowledge breaker' (Schleiss et al., 2022). In conclusion, new digital competencies need to be developed, e.g., how to leverage the positive potentials of AI, including the selection and combination of appropriate AI software, orchestration of adaptive and engaging high quality learning content and avoiding pitfalls that lead to low learning outcomes, mistrust in the educator, and unsatisfied students (Baidoo-Anu & Owusu Ansah, 2023; Lameras & Arnab, 2022; Mrabet & Studholme, 2023; Qadir, 2023). Nevertheless, high quality and engaging education is particularly important for information security education due to an ongoing high number of severe cyberattacks that, e.g., threaten human life or lead to the destabilization of enterprises as well as countries (Lella et al., 2022; Ralston, 2021).

With regard to the mentioned potential benefits and challenges of applying GAI to support the creation of high-quality learning content for engaging and effective lectures, educators may search for guiding literature. Beyond many articles focusing on implementing tutoring-systems or chat-bots in lectures, we have found scarce literature that focus only on specific tools and their functionalities, e.g., using Acrobatiq SmartStart for

URI: https://hdl.handle.net/10125/106378
978-0-9981331-7-1

biology classes (Schroeder et al., 2022), or they address prompts for ChatGPT only, but lack a pedagogical alignment to instances and guidance (Hwang & Chen, 2023; Rudolph et al., 2023; White et al., 2023). In order to contribute to existing research, this article strives to answer the following research questions:

RQ1: What (initial) design principles can be formulated for using GAI software based on LLMs to create learning content for learning objects?

RQ2: To what extent does the evaluation of the formulated design principles support the assumption, that GAI software based on LLMs supports educators in the creation of effective, high quality and engaging learning content for learning objects?

We employed the Design Science Research (DSR) Methodology by Peffers et al. (2007) to organize our approach. Our work addresses a novel evolving field of educational (design) research that approaches the application of AI in education (Hwang & Chen, 2023). Thus, this article contributes in two ways: By **a)** providing initial instance-based design principles (following Gregor et al. (2020), Heinrich and Schwabe (2014), and Chandra et al. (2015)) for applying a GAI based on LLMs to create high quality learning content for digital learning objects (LO). Here, we also share insights in our first steps with GAI as well as what difficulties we faced in applying GAI to create learning content and thus, this is a starting point for other educators and offers the possibility for discussion – *this refers to RQ1*. Further, **b)** by evaluating the presented design principles on a developed novel instance, to assess its effectiveness from a students' perspective and a focus on 'Learning', 'Engagement' and perceived 'Quality' – *see RQ2*. The instance is a digital LO composed of four pedagogically conceptualized elements that include GAI based learning content. This LO is implemented within the lecture on data privacy and information security for bachelor students at a German state university. Our approach, therefore, aligns with design research (Peffers et al., 2018) from an educational viewpoint as well as Laurillard's (2012) call for designing lectures as objects of both teaching and iteratively improving educational quality through research and sharing insights in underlying pedagogy.

2. Foundation

This article focuses on deriving and evaluating initial design principles (DP) for creating a 'learning object' (LO) with the support of GAI. Thus, the meaning of both terms (DP, LO) requires clarification.

Gregor et al. (2020) define DPs as 'prescriptive statements that indicate how to do something to achieve a goal' with specified information about the implementer, users, aims of the principle, contexts, and mechanisms, but also the rationale. In general, DPs can be defined at different maturity levels and thus, we describe our DPs (see chapter 3.2) according to Heinrich and Schwabe (2014) as initial instance-based design principles, which may contribute to nascent principles for generating LO content via LLM based GAI software. Thereby, the conceptual grounding is essential, including what design decisions and design knowledge influenced the derivation of the DPs (Heinrich & Schwabe, 2014) but also the underlying materiality characteristics and boundary conditions (Chandra et al., 2015). Therefore, we outline the conceptual grounding of our design principles in more detail.

We defined the problem motivation in chapter 1, concluding that one aspect of current calls for research refers to learning content creation supported by AI due to scarce design knowledge available, except for single specific use cases (Hwang & Chen, 2023; Rudolph et al., 2023; White et al., 2023). The foundation of our artifact is based on research on LOs and the pedagogical integration of mediating computer technologies for high quality education, and will be outlined in the following:

To date, there is no unique definition of the term LO (Papastergiou & Mastrogiannis, 2021). However, there is one definition that consolidates previous scientific literature: Learning objects are "interactive web-based tools that support the learning of specific concepts by enhancing, amplifying, and/or guiding the cognitive processes of learners" (Kay & Knaack, 2009; Papastergiou & Mastrogiannis, 2021). Moreover, LOs are "focused on specific topic areas and concepts of a curriculum, reusable[…]" and an independent instance with the aim to promote learning (Papastergiou & Mastrogiannis, 2021). Thus, "learning tool" is also a synonym for LO (Verville et al., 2021). Our artifact instance is such a LO, a digital reusable learning module (integrated within a learning management system (LMS)). The theme of the LO relates to 'brute force and the security of passwords' and is therefore, an essential part of the lecture 'data privacy and information security'. The corresponding learning objectives are formulated according to the revised version of Blooms' taxonomy (Krathwohl, 2002), e.g., 'the student will be able to name two recommended strategies for strong password creation' or '[…] to judge in what way the length of a password or technological security mechanisms affects the success rate of a brute force attack'. The kernel theory for the pedagogical structure of our LO, that might support students to achieve these objectives, is rooted in computer supported collaborative learning (CSCL). As Jeong et al. (2019) state, CSCL is effective in facilitating learning outcomes and it is 'built on the premise that collaborative knowledge construction and problem solving can effectively be assisted by technology'.

Furthermore, CSCL incorporates several different theories related to learning, collaborative knowledge creation, and using technology (Stahl & Hakkarainen, 2021). Jeong and Hmelo-Silver (2016) state seven main affordances (A1-A7) for CSCL instances that are recognized as factors of success (Jeong & Hmelo-Silver, 2016) and thus, could serve as the kernel theory for the pedagogical conceptualization of a LO that relies on CSCL: 'A1. Collaborative tasks, A2. communication, A3. sharing resources, A4. engaging in productive processes, A5. engaging in co-construction, further, A6. Monitoring and regulation, and A7. finding and building groups and communities'. This theoretical pedagogy knowledge is then conceptualized together with practical design knowledge into four pedagogical elements that are implemented as part of the proposed LO (for details see table 1), namely a text-based motivational introduction, a learning video, quizzes, and several collaborative tasks. For the latter, we provided a computer-mediated joint workspace, the LMS's forum and as an anonymous alternative for participation, the collaboration suite CryptPad (CryptPad, 2022). In order to guide the students, the educators moderated and debriefed all the elements and made all elements as well as the underlying tasks visible (in the LMS), in sequential order of the tasks to be completed.

The content for these pedagogical elements was generated by applying GAI software (see next chapter) and available validated information on the internet, e.g., the recommendations for strong passwords (BSI, 2023). The process of content creation is similar to the case that we would have chosen open educational resources (OER) instead, because educators have to curate learning content beforehand, e.g., by collecting, reviewing and selecting content items, but also by aligning these to the pedagogical approach (Deschaine & Sharma, 2015). However, to date, there are only a few promising AI based OER approaches that have the potential to support the curation process to a greater extent, e.g., the AI based X5GON platform (X5GON, 2021) and a few commercial platforms (Holmes & Tuomi, 2022). Furthermore, OER are still limited to the resources that already have been shared in terms of educational fields, languages, formats, and platforms, etc. But then again, GAI offers the potential to create new learning content from scratch (Hwang & Chen, 2023), and therefore, may even accelerate the availability of OER if shared by educators. In any case, both human and AI generated learning content requires adaptions (Hwang & Chen, 2023) to fit to the learning objectives or complexity levels of a lecture, or to correct factual errors. In other words, educators can 'cherry-pick' single items (if available) for a LO, but often at the cost of time, and then again, it still may require revision. Hereby, LLMs, such as GPT 4 respectively the chatbot ChatGPT, are able to support in this process of (re)writing and revising learning content (OpenAI, 2023a; Rudolph et al., 2023). Still, ChatGPT shows limitations, e.g., it lacks accuracy, but the learning abilities of GPT-4 are almost human-like, including context aware learning 'on the job'. Thus, it leads to the most accurate results of GAI to date, allowing rapid summarization of content and responses from different perspectives, translation, paraphrasing, writing source code or even providing a recommendation for how to orchestrate single knowledge items for a lecture (Ray, 2023; Rudolph et al., 2023). This strengthens the assumption that LLMs like GPT-3.5/4 hold strong potential to support educators. And given the scarce resources of well-ranked conferences and journals, the scientific community requests research on particular use cases of generative AI for efficient and effective teaching (Hwang & Chen, 2023; Rudolph et al., 2023).

Table 1. Drawing the connection between the learning object's pedagogical elements and CSCL

Elements of the LO and duration	*Implemented aspects in accordance with the 7 affordances (A1-A7) of CSCL (adopted from Jeong and Hmelo-Silver (2016))*
Motivational introduction (~ 1-2 min.)	Motivates an authentic problem context by a text that shows the relevance of the topic for students; introduces the group task. (A1)
Learning video (~ 2:54 min.)	Short learning video as a shared persistent knowledge base for the upcoming tasks and discussions (A5)
Quizzes (~ 10 min.)	Opportunity to choose between two multiple choice quizzes of different levels of difficulty (A4); establishing a further persistent knowledge base for upcoming tasks and discussions by providing the solutions for each question (A5).
Collaborative group tasks using either CryptPad or the forum (~ 40 min.)	Synchronous oral communication or asynchronous digital communication via CryptPad using the commentary function (A2). Collaboration using a joint (digital) collaborative workspace via CryptPad or the forum (A3 & A5). Monitoring the performances of different groups via CryptPad forum; oral regulation in case something is unclear; additional discussion task for outperforming groups (A6 & A7).

3. Toward design principles

3.1 Designing learning elements using GAI

The application domain of our approach is a data privacy and information security lecture at the bachelor´s level at a German state university. The learning content is provided to students via an LMS (based on the Stud.IP open-source framework) as a digital LO. Before creating learning content for this LO, using GAI, we analyzed recent providers for GAI software. In order to generate textual learning content such as for the motivational introduction, the voice-over script of the video, but also quizzes, and group tasks, we applied ChatGPT Plus by OpenAI. (OpenAI, 2023a) This GAI chatbot has been dominating the press recently due to its remarkable improvements in accuracy compared to previous versions and other LLMs (Rudolph et al., 2023). Although ChatGPT (based on GPT-3.5) is free as of June 2023, we decided to use ChatGPT Plus (based on GPT-4) because it outperforms GPT-3.5, especially in accuracy, for the German language (OpenAI, 2023b, 2023c; Rudolph et al., 2023). To include information from web resources we integrated the web-browser extension WebChatGPT (Qunash, 2023), as GPT-4 (in 06/2023) is trained with available data up to September 2021 (OpenAI, 2023c).

Furthermore, we used Microsoft (MS) PowerPoint from the Office 365 software package to create the slides for the learning video. This means, the video itself is based on the PowerPoint slides and an AI-generated voice-over speech. For artificial voice-over recording, we chose ElevenLabs with its multilingual language model (ElevenLabs Inc., 2023). This decision was driven by its user-friendly web-based interface, the human-sounding voice, and the availability of German voice cloning capabilities compared to freeware. For instance, the tool of Jemine (2022) is not available for the German language and sounds robotic. The process of combining the slides and the voice-over speech could be done either manually using video editing software, or through PowerPoint's built-in features, that allow merging and video export. To generate and orchestrate the visual elements of the presentation slides, we applied the integrated PowerPoint AI based feature called 'Designer' on the prepared slide master of our faculty. This automatically creates slides with customizable graphics and visual elements out of bullet points. We chose MS software mainly because of this helpful feature and because our university already has a license agreement for this software.

Overall, it is necessary to mention that there are several other (G)AI tools available that could also be used to create learning content, e.g., Synthesia (2023). Given the requirement of offering our lecture in German and the potential to include non-German speakers in the future, we deliberately chose a multilingual GAI, but otherwise, noncommercial tools could also suffice. Here, it should be noted that the majority of GAI providers identified during our market analysis exclusively support English. For instance, GAI software like Nolej (Neuronys, 2023), enables the creation of customizable interactive teaching modules in English. However, this in turn could be very interesting to create reusable LOs for English speakers.

In table 2, we provide an overview of the approximated time required to create content for the four pedagogical elements, time for preparing the LLM and adjustments included. We address the preparation tasks for content creation in chapter 3.2 to derive design principles. Based on the implementation, we conducted an intermediate evaluation to get first indications, whether our approach is purposeful and offers potential for time savings compared to non-GAI approaches. Two educators from different fields were interviewed, stating that it would take them approximately 180-210 minutes to create a similar LO for their lectures without GAI. In comparison, our GAI approach required a total of 130 minutes. Accurate time comparisons, however, are part of future research as it necessitates a rigorous comparison of, e.g., GAI supported versus a non-GAI supported LO content creation, from several educators, based on the same underlying pedagogical estimations.

Table 2. Approximated time spent to create content

preparation tasks (incl. adjustments)	~ prep. time	
Data priming & scenario	~ 5 min.	
Tailored motivation (250 words)	~ 20 min.	
Voice-over script (391 words for video)	~ 35 min.	Video ~∑ 55 min.
Voice-over audio using ElevenLabs (2:54 min)	~ 5 min.	
Generation of 7 Slides (visual content)	~ 5 min.	
Video editing (joining slides and voice-over)	~ 10 min.	
24 Quizzes (3 types; several difficulties)	~ 20 min.	
4 Collaboration tasks	~ 15 min.	
Implementation in LMS	~ 15 min	
	~∑ 130 min	

3.2 Deriving design principles

In order to derive DPs, we apply the schema of Gregor et al. (2020) and align the pedagogical evaluation criteria for LOs respectively the constructs 'learning, engagement and quality' (Kay & Knaack, 2009) to the DP components of the schema: the aim of the principle, the implementer, the user, the context, the mechanism, but also the rationale. In this article, the rationale is partly derived deductively from existing literature and partly from the acquired design knowledge during the creation of the learning content. Thus, the prescriptive boundaries of the following DPs are connected to a) the GAI software we used and its underlying LLM, b) the pedagogical foundation (see chapter 2) and c) the evaluation (see chapter 4). However, mature DPs require a more extensive foundation. This article offers a starting point for further iterations of research in this novel and evolving field.

We used ChatGPT Plus (GPT-4) to produce content for 4 pedagogical elements (see chapter 2)). The topic was password security in the context of brute force attacks. All prompts were entered in a single chat session, without expunging this session. The quality of the output created by a GAI depends heavily on the quality of prompts given by the user. This shows the significance of proficient prompt engineering during the content creation process. It is recommended to input prompts that are as precise, but also detailed and target-oriented as possible, e.g., providing ChatGPT with the desired output length (e.g., '… length of 250 words at most.') or content type (e.g., '… a multiple choice quiz') for the respective content generated. (OpenAI, 2023b; Qadir, 2023; Ray, 2023; White et al., 2023) Thus, in order to generate learning content in a time efficient manner, it is essential to know the learning objectives and corresponding possible types of tasks that are appropriate to support students in achieving these objectives. While ChatGPT can assist in the creative process of defining the pedagogical concept, this concept itself is still required for directed content creation. Without the preparatory pedagogical work, ChatGPT may create a lot of but potentially noncoherent content, that does not fit to the pedagogical storyline. Moreover, the time for revision would rapidly increase. We conclude this for the first DP:

1. Educators who aim to generate learning content for a whole LO by utilizing GAI should possess a thorough understanding of their pedagogical concept. They should know their elements and anticipated content types, including the associated task constraints. This is essential to generate coherent content elements one after another. For ChatGPT this includes inserting prompt instructions for each element and using one session per LO.

We indicate this DP to be valid, based on the outlined development process of the LO (further details in the following) and if the evaluation proves no issues regarding the coherence of the pedagogical elements and their content that may impact students perceived learning, engagement or quality negatively.

For generating precise and consistent content, we further prepared the LLM, also called priming (see OpenAI (2023b)), by supplying ChatGPT with reference data from various published website articles by the German Federal Agency for Information Security (BSI). This process is crucial for enhancing accuracy and mitigating hallucinations (OpenAI, 2023b, 2023c). Our prompt engineering process is illustrated through examples provided in table 3. Within the field of data priming, the *placeholder* represents the text extracted from the website, which can be substituted with any desired input text. For tailoring the learning content to students, ChatGPT was instructed to adopt a specific persona. This, in general, could include 'job description, title, fictional character, historical figure, etc.' (White et al., 2023). By implementing 'data priming' and incorporating 'personas', we established an initial structure that allows ChatGPT to generate content not solely relying on its pre-trained data. For an instance, we instructed ChatGPT to be an educator and requested it to write a tailored motivation of 250 words that aims to address the needs and emotions of students, including typical use cases. Additionally, we used GPT-4's ability to self-reflect. Repeating this reduces hallucinations and improves output quality in terms of accuracy and validity to a great extent (Shinn et al., 2023).

Table 3. Prompt engineering examples

Data priming prompt (translated from German)
The following is the first {text} I want to learn about. Write "yes." if you understand it. text = *placeholder*
Persona prompt (translated from German)
For the following tasks act as an instructor for cyber security. You are specialized in creating awareness for bachelor students. You are creating a security awareness training on the topic of password security for students of different faculties.
Self-reflection prompt (translated from German)
Are you sure about your answer?
Feedback prompt (translated from German)
I don't like the metaphor for the password manager. Please try again using another metaphor but keep the rest of your outputted text. Maintain the last requirements.

The second element of our LO, the learning video, is based on a 391 words long voice-over script which we generated using ChatGPT. The content incorporates the two recommended BSI strategies for creating secure passwords, as outlined on their website (BSI, 2023). This is one of the reasons, we used the browser extension 'WebChatGPT', to include information from the latest web resources. To further motivate and underline the importance of the theme and to enhance its usefulness for students, we instructed ChatGPT to integrate helpful web-services in the video, e.g., 'haveibeenpwned.com', to offer the possibility to check whether one's data has been leaked in a data breach. After revision, the text was summarized into bullet points by ChatGPT and then copied into the MS PowerPoint presentation by hand. Then, we used the built-in PowerPoint 'Designer'-feature to create (all) seven slides automatically, though we had to change one unfitting pictogram. The voice-over audio for the video was automatically generated by ElevenLabs based on the voice-over script. Here, we trained the voice of ElevenLabs with a 2-minute-long voice-over dataset to create a video with a length of 2:54 minutes.

For the third pedagogical element of the LO, we created 24 quizzes in total using ChatGPT. These quizzes encompass diverse levels of difficulty and question types, including 'fill-in-the-blanks', 'multiple-choice-quizzes' and 'true or false' questions.

Lastly, we prepared the fourth element, the collaborative group tasks comprised of four sub-tasks. The first (beginner) task was an existing mathematical task (related to the topic of password security and brute force). This task was input into ChatGPT, similar to the data priming process before, to generate three more complex tasks and their explanations. Both tasks and explanations were improved in accuracy by the self-reflection technique. Prompt engineering, therefore, addresses aspects related to learning, engagement, and quality because the varying difficulties of quizzes promote differentiated learning, while engagement relies on interest and motivation. Furthermore, prompt engineering facilitates overall quality, e.g., through well written content. Thus, we summarize this design knowledge and derive the second DP including 4 facets:

2. Educators who aim to generate effective learning content with the help of LLM applications (such as ChatGPT) should apply prompt engineering as a general mechanism to get the desired results. In conclusion, the following prompt engineering techniques should be considered (if available):

2.1 Employ priming and input specific information as reference, e.g., validated published documents or content from OER, so that the probability of high-quality output increases and the likelihood of 'hallucinations' decreases.

2.2 Introduce a specific persona, to adopt the viewpoint of this persona and generate results with a tailored writing style among other traits. This impacts learning, engagement, and quality. For instance, the persona 'educator' might focus on simple explanations (learning), or a tailored text for students might increase the motivation to participate (engagement), by including interesting facts.

2.3 Employing self-reflection gives the opportunity to ensure the quality of the results.

2.4 Input precise prompts and feedback and repeat the prompt to increase the level of quality. Thereby, for higher efficiency, consider requesting, e.g., 3 versions of a desired output.

This DP, including its subordinate principles, offers indications to be valid, if the evaluation does not provide indications that our approach has a considerable negative impact on learning, engagement or quality.

For a better understanding of our next DP, we will describe the challenges worth mentioning we faced that led to additional adjustment prompts: We intended to write a motivational introduction, which includes recent data on password breaches. But the output included false facts, although we used ChatGPT with the WebChatGPT extension to obtain current statistics from the internet. Thus, we simply excluded these false facts and figures. Moreover, the creation of the voice-over script for the video posed two main challenges: First, despite all efforts to adjust the output results by using the self-reflection prompts, two small but important manual adjustments were necessary due to crucial missing and partly false information regarding recent password security strategies that should have been obtained from the input data. Second, approximately 5% of the whole semantic structure of the voice-over script required improvement at the end of the process in total. Additionally, not all initially provided metaphors for password security were appropriate in the first place, but further iterations solved this problem by giving ChatGPT feedback. With respect to the quizzes, 2 out of 24 generated quiz questions were factually incorrect and thus, were discarded. These errors occurred because ChatGPT relied on outdated information that did not match our primed data due to the GPT-4 model cutoff date (OpenAI, 2023c). This leads to the third DP:

3. Educators who aim to create high quality learning content should carefully consider the current limitations of the applied LLM in order to understand possible quality issues that might occur. This will facilitate effective and efficient revision of the output.

We argue that, given the available literature, there is evidence that this principle is valid, provided that our evaluation does not imply contrary evidence of quality problems. To sum up, we answered RQ1 by deriving the outlined DPs, based on both literature and acquired design knowledge from the design process of the LO.

4. Evaluation of the design principles

4.1 Evaluation methodology

In order to validate the stated DPs of chapter 3, we evaluate the effectiveness of our approach by focusing on a students' perspective. We assess how *engaging* the use of the LO is, how it might or might not *support learning* as well as how the *quality* of the learning content is perceived. Therefore, we applied a slightly adopted version of the LOES-S questionnaire of Kay and Knaack (2009) to evaluate the LO on the basis of the mentioned criteria. This questionnaire originally includes 12 quantitative items that adopt a 5-point Likert scale ('strongly agree', …,'strongly disagree') and 2 qualitative items which we decided to expand to 8 items in accordance with the 4 pedagogical elements of the LO (2 items per 4 elements: motivational introduction, learning video, quizzes, collaborative group task). More precisely, we strived to create a more meaningful picture by tailoring the qualitative item "What, if anything, did you Like about the learning object?", and the almost similar second item that focuses on "Not Like", to every element of the LO, e.g., "What, if anything, did you Not Like about the motivational introduction". In order to provoke more precise qualitative statements, we added examples for feedback dimensions to each qualitative question, e.g., "You may consider the following criteria for your feedback: well-written, enjoyable, exciting or boring, coherent, […]". These dimensions were originally based on the constructs "quality" and "readability" of a questionnaire for assessing the quality aspects of AI generated textual news, provided by

Haim and Graefe (2017), adopted from Sundar (1999). This is transferable as, e.g., both the voice-over from ElevenLabs as well as the PowerPoint Designer rely on textual input that was generated by ChatGPT. Therefore, the final questionnaire consists of 12 quantitative items, based on the constructs 'Learning' (5 items), 'Quality' (4 items) and 'Engagement' (3 items), in addition to the 8 adjusted qualitative items specified (Kay & Knaack, 2009). Our sample consists of n=31 participants of our lecture 'data privacy and information security' (bachelor level, for descriptive demographic data, see table 4). Initially, the dataset consisted of 32 surveys, but one survey had to be discarded because of obviously invalid data (monotonously repetition of the same answer). The Cronbach's α value for internal reliability are satisfactory with α=.906 for 'Learning' and α=.793 for 'Engagement'. The α value for 'Quality' was α=.624 meaning that α<=.7 and thus, this requires further interpretation. Nevertheless, research states α>=.6 can be considered as still 'satisfactory' for the interpretation of results (Taber, 2018). The qualitative items provide deeper insights into quality aspects and because both the quantitative and the qualitative items are analyzed together, we argue to include the construct of 'Quality' without further adjustments.

We analyzed the quantitative data by calculating the mean and median values of the constructs. Then, we summarized noteworthy variations for single items, e.g., differing median values for different bachelor programs. Finally, in order to outline information that might affect the DPs, we checked for items that show median scores lower than 4/5, as conducted by Verville et al. (2021). In addition, we outline peculiarities of mean values (Kay & Knaack, 2009). Complementing this, we assessed all qualitative feedback statements and aligned them to the coding scheme that was used by Kay and Knaack (2009). This process was conducted by two independent researchers to consolidate coinciding coded statements and reconcile differences.

4.2 Results

Analyzing the quantitative data revealed that all constructs show a mean value above 4.0. The median values are 4/5 for both 'Learning' and 'Engagement', but even 5/5 for 'Quality'. It is further noteworthy that all items showed a median value of at least 4/5 but only three items ('*The instruction in the LO was easy to follow*', '*The LO was easy to use*', '*The LO was well organized*') show a median value of 5/5 each. Comparing the median values of each study program shows that all values are >=4, except for the three items '*The LO helped teach me a new concept*' (median: 3.5) reported by n=4 students of industrial engineering, but also '*The help features in the LO were useful*'

Table 4. Descriptive sample characteristic (n=31)

study program (bachelor)	business administration (n=14), business information systems engineering (n=11), industrial engineering (n=4), law (n=1), economics (n=1)		
sex	8 females	22 males	1 diverse
semester	4th semester (16); 5th semester (1); 6th semester (8); 7th semester (1); 8th semester (5)		

Table 5. Mean and median values of LOES-S constructs (n=31)

LOES-S	Learning	Quality	Engagement
Median (Mean)	4 (4.13)	5 (4.65)	4 (4.13)
SD	.922	.503	.806

(median: 3.0) reported by one student of economics, and the item '*I found the LO motivating*' that is reported by two groups with a median of 3.0 each, namely business information systems engineering students (n=11) and the one student of economics. Beyond the mostly acceptable values for the median, it should be outlined that the mean values of business information systems students are often lower than 4.0, particularly for the constructs 'Learning' (mean: 3.75) and 'Engagement' (mean: 3.58) whereat all other study programs (except the student of economics) report mean values above 4.0 for the three measured constructs. Further, while 10 of 12 items have mean values > 4.0, notably, 2 items ('*The LO helped me teach a new concept*', mean: 3.65; '*I found the LO motivating',* mean: 3.90) show mean values that are slightly lower.

Beyond the quantitative results, we received 212 single qualitative feedback statements. In some cases, these statements included more than one information for different subcategories of the LOES-S, e.g., 'Well written and actually motivating, quite short' (uncoded statement). Here, 'well written' refers to the subcategory 'text', 'motivating' refers to subcategory 'engage' and 'short' refers to the subcategory 'design'. Thus, the original statements had to be split, coded for each matching subcategory, and summarized. This resulted in 236 statements. Table 6 shows the alignment of recoded statements to the LOES-S coding scheme. The table provides information on the number of statements that were assigned to the subcategories of 'Learning', 'Engagement' and 'Quality'. In addition, it highlights the share of statements per category (Like/Not Like) and pedagogical element, namely the 'motivational introduction', the 'learning video' and the 'forum/CryptPad task', but also the 'quizzes'. It is noteworthy that 'Like' statements were assigned in more than 75% and up to 88% of all statements. Furthermore, the elements 'motivational introduction' and 'learning video' received most statements for the LOES-S category 'Quality'. The 'forum/CryptPad task' received most statements in 'Learning' and 'Engagement', and for the 'quizzes', most statements are given for the 'Learning' and 'Quality' categories.

4.3 Discussion

As our dataset consists of 236 coded statements, we had to focus our discussion on specific main statements in this article to report only critical aspects that help to answer RQ2. In order to systematically choose relevant statements, we set the following criteria: First, we discarded statements of subcategories (see table 6) that provided only 1 or 2 statements in total, as they are not representative, e.g., the statements for 'visual' and 'help'. Second, we analyzed the four pedagogical elements one after another and selected the statements of the corresponding constructs ('Learning', 'Engagement', 'Quality') as well as their subcategories that received the highest number of statements or offer meaningful insights to answer RQ2, in particular with regard to the quality of the pedagogical setup or the use of GAI. Third, we checked all statements again if we missed any statement that represented a critical aspect with regard to the DPs and use of GAI, e.g., factual errors that we might have overlooked (but found none). Considering these criteria, we discuss the derived DPs in the following. The quantity of the coded statements that we found are outlined within the sentence or within parenthesis, e.g., '(1)'. For the construct '*Learn*', where the quantitative evaluation resulted in a median value of 4/5, it is also interesting to further discuss the mean value of 4.13 and the standard deviation (see table 5) in contrast to the qualitative statements. For educators, it is worth noting that a total of 5 statements describe the learning video to be valuable. More precisely, that it is helpful in general (2) but also because of the useful metaphoric comparison (2) that was made, and its instructive design. However, students wanted more new information and details (2). Regarding the collaborative group task, using the forum or the CryptPad as an anonymous alternative, it has to be mentioned that all students chose CryptPad. Thereby, 8 students found it supportive in learning and in contrast 3 students did 'Not Like' the possibility to see results of other groups for different reasons, e.g., perceived unfairness (1). Another important point is, that 5 students stated that the quizzes supported learning for varying reasons, e.g., to repeat content (2). Moreover, students reported in the subcategory 'challenge', that the different quizzes offer appropriate (4), but also different levels of difficulty (3) and therefore, enables 'a real differentiation' (1). Nonetheless, 5 students reported that the questions are too easy (5). As the mean value for the group of business information systems students is lower than the average, this indicates that at least some of the students require even more difficult tasks.

Further, considering the construct '*Engagement*', our data shows a median value of 4. Thereby, the corresponding qualitative statements count 44 'Like' statements out of 58 statements given for the whole construct. This suggests that our approach was in fact engaging. This is supported by statements that are related to the motivational introduction (subcategory 'engage'), e.g., 'motivating' (3), 'enjoyable' (2), or 'fosters curiosity' (2). This further indicates that adopting the persona while using ChatGPT was suitable and thus, supports DP 2.2. Moreover, the learning video is reported to be enjoyable (3) and lively (2). Taking the implementation of our collaborative group task into count, 16 'Like' statements are given within the subcategory 'Technology', most referred to the anonymous (7) and real-time (7) participation using CryptPad. Then again also 6 students reported that they missed a clear structure within the collaborative group task and that they did not use the forum because of the non-anonymous participation (1) and too many (one per group) branches (1) as it leads to orientation issues. This additionally underlines that digital tools can only be supportive if applied in a thoroughly planned pedagogical setup, however, it does not affect the DPs.

Better than the previous constructs, the median value of 5/5 for the construct 'Quality' but also 112 'Like' statements (compared to 24 'Not Like' statements) that refer to quality aspects indicate an overall good quality of our pedagogical approach. This means that for the motivational introduction, e.g., 12 statements describe for 'text' to be well-written, clear (8), and that the amount of text and information was appropriate (5). In addition, for the learning video, 'Like' statements refer to the comprehensiveness of the voice-over audio (15) and its 'clear voice' (6), but also that it is designed in an appropriate length (2). In contrast, explanations for 'Not Like' statements include that the quality of the voice should be improved (2), and that the voice sounds monotonous (1). This indicates that, overall, the quality of the video is still acceptable. For the collaborative group task, 5 students mentioned 'it was easy to use' (5), 'interactive' (2), and the task of structure was clear (1). Further noteworthy is, that out of 10 statements, 4 state the quiz to be 'clear' and 2 are related to 'overall good quality'. But then again 4 students found the questions unclear (2), ambiguous (1), or that they sound similar (1). In total, these statements all support DPs 1 and 3, because the students did not report consistency issues or made-up information, instead the feedback was largely positive. In accordance with the aspects mentioned above, we are able to answer RQ2: Our results finally indicate that our approach was overall perceived to be effective for learning, engaging and of high quality, as it is stated to be, e.g., well written, interesting, clear or well explained, but also motivating content. Thus, DP 2 and its subordinate DPs are supported. This content was created mainly by applying a LLM for multiple purposes but always considering the

Table 6. Assignment of qualitative statements to subcategories (LOES-S) and elements of the learning object

LOES-S (n=31)			Learning			Engagement			Quality									
Elements of LO \ Subcategories			challenge	learn	visual	compare	engage	technology	animate	audio	easy	graphics	help	interactive	control	org./design	text	theme
motivational introduction	Like	57/∑65 ≈ 88%	-	2	-	-	8	-	-	-	-	-	-	-	-	2	33	12
	Not Like	8/∑65 ≈ 12%	-	-	-	1	2	-	-	-	-	-	-	-	-	2	3	-
learning video	Like	45/∑60 ≈ 75%	-	5	-	-	5	-	-	21*	-	1	-	-	-	7	-	6
	Not Like	15/∑60 ≈ 25%	-	2	-	-	1	-	-	5	-	2	-	-	1	4	-	-
forum/ CryptPad task	Like	36/∑48 ≈ 75%	1	8	-	-	3	16	-	-	5	-	-	2	-	1	-	-
	Not Like	12/∑48 ≈ 25%	-	3	-	-	-	8	-	-	-	-	-	-	-	1	-	-
quizzes	Like	49/∑63 ≈ 78%	9	5	1	-	12	-	-	-	2	-	1	1	-	3	10	5
	Not Like	14/∑63 ≈ 22%	6	-	-	-	1	1	-	-	-	-	-	-	-	1	5	-
		∑ 236	**Note**: 1. *statements include aspects that mentioned quality of explanation 2. missing statements are represented by "-"															

learning objectives and pedagogy (DP 1). Thus, we recommend using our DPs (1-3) together with a LLM for, e.g., writing motivational introductions for target groups, convincing voice-over scripts, or the generation of multiple quizzes. Both, the DPs and the instance are an indicator that GAI supports the creation of high-quality learning content. Although the negative statements in the qualitative data do not account for the largest proportion, they show that GAI should be used with caution to avoid pitfalls. In particular, AI-generated tasks and quizzes need to be thoroughly checked for similarities and ambiguous wording that can negatively impact learner satisfaction (DP 3).

5. Conclusion

By answering RQ1, we provided design knowledge and initial design principles for deriving GAI based learning content for a LO in a systematic manner. In addition, we provided indications for efficacy through a rigorous evaluation and thereby answering RQ2. However, one of our main limitations is the small sample size and the subjective nature of coding qualitative feedback. Thus, we recommend further research to investigate effects of GAI based learning content on students and educators when our DPs are applied. This would also contribute to the part of the DSR community that has a focus on education. Key contributions for DSR are that revising the prompts and outputs of the GAI systematically is, in combination with utilizing the functions of LLMs (e.g., using persona), one of the most important aspects for the creation of effective learning content and thus, indicate to be design principles. Moreover, future research should investigate whether time benefits through applying a LLM (including other LLM providers) may exist and scale with a) learning videos of greater length, b) the number of quizzes and tasks that are generated, c) the availability of open educational resources that can be used as input data, d) prompt engineering techniques, and e) in dependence on the purpose, e.g., providing podcasts only as they do not require video editing. Given the reusability of a LO this should give educators more time for research or community activities but also for thoroughly prepared pedagogical lectures.

6. References

Baidoo-Anu, D., & Owusu Ansah, L. (2023). *Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning.* https://doi.org/10.2139/ssrn.4337484

BSI. (2023). *Creating Secure Passwords*. https://www.bsi.bund.de/dok/6596574

Chandra, L., Seidel, S., & Gregor, S. (2015). Prescriptive Knowledge in IS Research: Conceptualizing Design Principles in Terms of Materiality, Action, and Boundary Conditions. In *2015 48th Hawaii International Conference on System Sciences*.

CryptPad. (2022, September 8). *CryptPad: end-to-end encrypted collaboration suite*. https://cryptpad.fr/

Dao, X.-Q., Le, N.-B., & Nguyen, T.-M.-T. (2021). AI-Powered MOOCs: Video Lecture Generation. In *2021 3rd International Conference on Image, Video and Signal Processing*.

Deschaine, M. E., & Sharma, S. A. (2015). The Five Cs of Digital Curation: Supporting Twenty-First-Century Teaching and Learning. *InSight: A Journal of Scholarly Teaching*, *10*, 19–24. https://eric.ed.gov/?id=EJ1074044

ElevenLabs Inc. (2023). *ElevenLabs: Voice AI*. https://beta.elevenlabs.io/

Forrester, N. (2023). Fed up and burnt out: 'quiet quitting' hits academia. *Nature*, *615*(7953), 751–753. https://doi.org/10.1038/d41586-023-00633-w

Gregor, S., Kruse, L., & Seidel, S. (2020). Research Perspectives: The Anatomy of a Design Principle. *Journal of the Association for Information Systems*, *21*, Article 6, 1622–1652. https://doi.org/10.17705/1jais.00649

Haim, M., & Graefe, A. (2017). Automated News. *Digital Journalism*, *5*(8), 1044–1059. https://doi.org/10.1080/21670811.2017.1345643

Heinrich, P., & Schwabe, G. (2014). Communicating Nascent Design Theories on Innovative Information Systems through Multi-grounded Design Principles. In *Advancing the impact of design science: Moving from theory to practice: 9th international conference, DESRIST* (Vol. 8463, pp. 148–163). Springer. https://doi.org/10.1007/978-3-319-06701-8_10

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, *57*(4), 542–570. https://doi.org/10.1111/ejed.12533
Hwang, G.-J., & Chen, N.-S. (2023). Editorial Position Paper: Exploring the Potential of Generative Artificial Intelligence in Education: Applications, Challenges, and Future Research Directions. *Educational Technology & Society*, *26*(2). https://www.jstor.org/stable/48720991
Jemine, C. (2022). *Real-Time-Voice-Cloning*. https://github.com/CorentinJ/Real-Time-Voice-Cloning
Jeong, H., & Hmelo-Silver, C. E. (2016). Seven Affordances of Computer-Supported Collaborative Learning: How to Support Collaborative Learning? How Can Technologies Help? *Educational Psychologist*, *51*(2), 247–265. https://doi.org/10.1080/00461520.2016.1158654
Jeong, H., Hmelo-Silver, C. E., & Jo, K. (2019). Ten years of Computer-Supported Collaborative Learning: A meta-analysis of CSCL in STEM education during 2005–2014. *Educational Research Review*, *28*, 100284. https://doi.org/10.1016/j.edurev.2019.100284
Kay, R. H., & Knaack, L. (2009). Assessing learning, quality and engagement in learning objects: the Learning Object Evaluation Scale for Students (LOES-S). *Educational Technology Research and Development*, *57*(2), 147–168. https://doi.org/10.1007/s11423-008-9094-5
Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory into Practice*, *41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2
Lameras, P., & Arnab, S. (2022). Power to the Teachers: An Exploratory Review on Artificial Intelligence in Education. *Information*, *13*(1), 14. https://doi.org/10.3390/info13010014
Laurillard, D. (2012). *Teaching As a Design Science: Building Pedagogical Patterns for Learning and Technology* (3rd ed.). Routledge. https://www.taylorfrancis.com/books/9781136448201
Lella, I., Theocharidou, M., Tsekmezoglou, E., Svetozarov Naydenov, R., Ciobanu, C., & Malatras, A. (Eds.). (2022). *ENISA threat landscape report 2022*. *ENISA threat landscape 2022: July 2021 to July 2022*. ENISA.
Ma, Y., & Siau, K. L. (2018). Artificial intelligence impacts on higher education. In *MWAIS 2018 Proceedings*. https://aisel.aisnet.org/mwais2018/42
Miller, J. (2019). Where does the time go? An academic workload case study at an Australian university. *Journal of Higher Education Policy and Management*, *41*(6), 633–645. https://doi.org/10.1080/1360080X.2019.1635328
Mrabet, J., & Studholme, R. (2023). ChatGPT: A friend or a foe? In *International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*.
Neuronys. (2023). *Nolej*. https://nolej.io/
OpenAI. (2023a). *ChatGPT*. https://openai.com/chatgpt
OpenAI. (2023b). *ChatGPT Documentation: Best practices*. https://platform.openai.com/docs/guides/gpt-best-practices
OpenAI. (2023c, March 15). *GPT-4 Technical Report*. http://arxiv.org/pdf/2303.08774v3
Papastergiou, M., & Mastrogiannis, I. (2021). Design, development and evaluation of open interactive learning objects for secondary school physical education. *Education and Information Technologies*, *26*(3), 2981–3007. https://doi.org/10.1007/s10639-020-10390-2
Peffers, K., Tuunanen, T., & Niehaves, B. (2018). Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research. *European Journal of Information Systems*, *27*(2), 129–139. https://doi.org/10.1080/0960085X.2018.1458066
Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302
Pettinato Oltz, T. (2023). *ChatGPT, Professor of Law.* https://doi.org/10.2139/ssrn.4347630
Qadir, J. (2023, May 1). Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education. In *2023 IEEE Global Engineering Education*.
Qunash, A. (2023). *WebChatGPT extension*. https://github.com/interstellard/chatgpt-advanced
Ralston, W. (2021). *The untold story of a cyberattack, a hospital and a dying woman*. SECURITY. https://www.wired.co.uk/article/ransomware-hospital-death-germany
Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, *3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003
Rudolph, J., Tan, S [Samson], & Tan, S [Shannon] (2023). Chatgpt: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, *6*(1). https://doi.org/10.37074/jalt.2023.6.1.9
Schleiss, J., Hense, J., Kist, A., Schlingensiepen, J., & Stober, S. (2022). Teaching AI competencies in engineering using projects and open educational resources. In *Towards a new future in engineering education, new scenarios that european alliances of tech universities open up* (pp. 1592–1600). Universitat Politècnica de Catalunya. https://doi.org/10.5821/conference-9788412322262.1258
Schroeder, K. T., Hubertz, M., van Campenhout, R., & Johnson, B. G. (2022). Teaching and Learning with AI-Generated Courseware: Lessons from the Classroom. *Online Learning*, *26*(3). https://doi.org/10.24059/olj.v26i3.3370
Shinn, N., Cassano, F., Labash, B., Gopinath, A., Narasimhan, K., & Yao, S. (2023, March 20). *Reflexion: Language Agents with Verbal Reinforcement Learning*. http://arxiv.org/pdf/2303.11366v2
Stahl, G., & Hakkarainen, K. (2021). Theories of CSCL. In *International Handbook of Computer-Supported Collaborative Learning* (pp. 23–43). Springer, Cham. https://doi.org/10.1007/978-3-030-65291-3_2
Sundar, S. S. (1999). Exploring Receivers' Criteria for Perception of Print and Online News. *Journalism & Mass Communication Quarterly*, *76*(2), 373–386. https://doi.org/10.1177/107769909907600213
Synthesia. (2023). *Synthesia.io: AI video creation platform*. https://www.synthesia.io/
Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2
Verville, L., Côté, P., Grondin, D., Mior, S., & Kay, R. (2021). The development and evaluation of an online educational tool for the evidence-based management of neck pain by chiropractic teaching faculty. *The Journal of Chiropractic Education*, *35*(1), 95–105. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7958658/
White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023, February 21). *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*. https://arxiv.org/pdf/2302.11382
X5GON. (2021, April 9). https://platform.x5gon.org/join/