



Article

AI-deation: When the Teacher is a Transformer in Role-Playing to create Privacy Decision Serious Games

Patrick Jost

Human-Centred-Technologies Research Centre, FHV University of Applied Sciences, Dornbirn, Austria
patrick.jost@fhv.at

Keywords:

Serious Game Ideation
Artificial Intelligence
Human-AI Collaboration
Role-Playing Ideation
Serious Game Co-Design
AI Role-Playing Activity
Generative AI Support

Received: March 2024

Accepted: June 2024

Published: June 2024

DOI: 10.17083/ijsg.v11i2.762

Abstract

Understanding the synergistic potential of human collaboration with artificial intelligence (AI) in creative endeavours, such as ideating Serious Game (SG) concepts, is of vital interest in our era of digital transformation. This study probes two pivotal questions: First, how does the incorporation of a GPT-4 transformer AI, assuming the role of a teacher, influence support for student teams during the ideation and balancing of SG concepts? Second, what are the students' perceptions of AI integration when co-designing these concepts with an AI in the educator role? In a between-group research design, two distinct groups engaged in a collaborative role-playing activity with digital role-specific cards and a visualised board to ideate and balance SGs addressing privacy decision-making. The first group, engaging in a local setting, collaborated with an AI that played the teacher role. In contrast, the second group played the co-ideation activity in a remote setting, with a human playing the role of the teacher. The findings indicate that generative AI can successfully be sourced to play the teacher role in a collaborative role-playing activity. Crucially, the timing of AI intervention thereby emerged as an essential factor that can impair creative support. Scheduled AI interventions can offer fresh insights but may not align with immediate team needs. The insights underscore the requirement to determine the most effective timing for AI intervention in human-AI co-playing ideation sessions to foster the full potential of an AI filling a role in a collaborative design process. Implications synthesised from the analytical findings and practical insights on AI-suggested design propositions/conflicts are discussed conclusively.

1. Introduction

1.1 Background – Serious games for investigating privacy decision-making

In our daily lives, as we increasingly integrate data services, we often find ourselves making frequent decisions about sharing our data. Notably, teenagers and young adults are particularly vulnerable to making decisions that are less considered regarding their privacy [1]. In this

context, SGs emerge as an effective tool for engaging this demographic and promoting more thoughtful privacy choices. However, the challenge lies in designing a SG that not only entertains but also effectively heightens awareness about real-world privacy issues. And in particular, when as in the present study, the main aim is to learn more about how people make privacy decisions, taking the perspective of researchers in decision-making. This requires a carefully balanced game concept. As noted by Dörner et al. [2], SGs are intended to entertain but also to fulfil one or more additional (characterising) goals. In the realm of privacy decision-making research, SGs can explore factors like risk behaviour and social group influences, providing insights into how people make privacy decisions under uncertainty. This approach does not just serve the purpose of entertainment; it also has two additional goals. Firstly, it educates players, fostering better privacy decisions through practice in privacy-related challenges and, therefore, gaining expertise towards more intuitively correct decision-making [3], [4], [5]. Secondly, it serves as a research tool, helping to understand the human factors that influence decision-making [1], [6], [7].

Balancing these different perspectives in an early ideation phase is essential to avoid interrupting the players' perception of flow [8] as well as unintended alteration of perceived uncertainty and extraneous cognitive impact from game interaction in the published SG [9], [10] aiming to examine privacy choices. Together with players' developing privacy expertise, these factors must be considered when drafting ideas and implementation concepts for the goals in a SG. Researchers following the research goal of examining players' decision-making must be able to exert control in observing the adaption of uncertainty in a SGs presented decision-making scenarios during the drafting process of a SG.

Considering the perspectives of the different stakeholders and engaging them in a role-playing co-design activity can help to create such a balanced SG while offering it as a game that can provide sustainability during a project's developing research process [11], [12], [13]. It has been shown that a SG can bring together experts and novices and enable them to discuss about underlying domain challenges [6]. In privacy decision-making, these include, for example, today's ubiquitous trade-off that people face when deciding to use offered digital services against allowing access to their private data.

During the time of sharing data for using a service, there is undisclosed uncertainty involved in how and where the data will eventually be processed, stored, or aggregated and what that usage may lead to. Despite organisations' responsibility to clarify these concerns, users are often either only superficially informed or deliberately led to complicated interface constructs. This practice, known as dark pattern interface design, heightens the cognitive load and leads users to an inconsiderate decision to accept personal data processing and quickly move on with using the service [14]. Several other aspects of privacy decision-making are thereby targeted in research, including the privacy paradox, describing that people may, in fact, be aware of privacy threats but still make highly unfavourable privacy choices [15], [16].

Researchers in decision-making study how cognitive biases and contextual factors influence privacy choices [17]. They explore whether decisions are reflective, based on personal experience and knowledge of data-sharing risks, or if they are intuitive and less considerate of potential threats [18], [19], [20]. However, designing SGs that allow for appropriate research of privacy decisions requires thoughtful balancing of the uncertainty that is eventually presented to the players with balancing between the goals of research, learning and entertainment [1]. A process in which research teams that aim to use SGs in decision-making investigation require design and development support and the ability to include the perspective of various stakeholders, such as the intended players, privacy, or education experts.

A SG design process complemented by an AI language model, such as ChatGPT-4, could help researchers ideate and balance the research goal with the other aims of a SG and support a feasible design process. In particular, if temporarily unavailable, stakeholders, such as educators, could be supplemented by AI taking the role in the design process. It could provide

creative input while keeping the stakeholder perspective for balancing the SG for decision-making research.

1.2 Artificial intelligence support in collaborative design activities

Integrating AI into collaborative design activities presents an intriguing interplay of benefits and challenges. When considering its application in SGs creation focused on privacy decision-making, the role of AI extends to become a collaborative role-player for advancing research in decision-making with considerations pointed out by recent studies.

Li et al. [21] suggest that AI's influence on organisational creativity, primarily through knowledge sharing, could be pivotal in collaborative research settings. AI could facilitate the ideation process in teams focused on privacy decision-making research with SGs by providing a sounding board for reflecting/exchanging ideas and fostering creative solutions. This capability is crucial for research groups exploring complex decision-making scenarios, where diverse perspectives on the multiple SG goals and innovative approaches are essential.

Zhang et al. [22] bring to light the nuanced impact of AI on team dynamics and motivation. Their findings indicate that AI can both aid and hinder human design team performance depending on their skill levels and dynamics. Their study found that AI support in design collaboration with human teams can boost the initial performance of low-performing teams. In contrast, high-performing design teams experience a reduction in cognitive demand. Zhang et al. [22] emphasise that context and interaction aspects in AI support are critical for effectiveness in collaborative settings. This is particularly relevant for research teams in decision-making studies, where the interplay between AI assistance and human expertise must be carefully balanced to avoid cognitive overload [23] and ensure productive collaboration with stakeholders.

Sharples' exploration [24] into social generative AI highlights the potential of AI to facilitate deeper conversation and exploration within learning environments. This aspect can be leveraged in research contexts, especially in SGs designed for privacy awareness. AI can aid in stimulating discussion and critical thinking about privacy decisions and their implications regarding the research aims. Tang et al. [25] and Vazquez [26] demonstrate the potential of AI in assisting complex design tasks. In the context of SGs for privacy decision-making, AI's capability to aid in unravelling intricate domain decision-making scenarios and suggesting appropriate visualisations can significantly enhance the research process, assisting teams in visualising and understanding multifaceted privacy issues.

Sarkar [27] and Wang [28] further expand on AI's role in creative processes and integrated design systems. In research groups examining decision-making, AI's transformative impact can shift the focus from manual data analysis and scenario building to a more integrated approach where AI assists in synthesising complex data and generating insightful scenarios.

Analysing these reported findings and advancements, the potential of AI as support in collaborative design tasks, particularly considering the context of SGs focused on privacy decision-making, is significant. AI can act as a catalyst in creative ideation, foster effective knowledge sharing, and assist in complex scenario design, thereby enriching the research process. However, integrating AI also brings challenges, such as its impact on team dynamics and its limitations in understanding complex human interactions. These insights emphasise the need for further research in optimising AI's role in collaborative design activities. How collaboration with characters or roles played by an AI decision-maker in a SG can be effectively sourced for ideation and balancing support but also for attaining feasible SG design iterations is currently underexplored. Integration and contextual considerations are, however, crucial in leveraging readily available text-generative AI language models such as GPT-4 (openai.com) for effective co-design activities with humans.

1.3 Research objectives

This study's research objectives thus aim to gain insights on how to balance generative AI assistance with human expertise in a research-focused SG design activity based on a tested card and board role-playing process – the *Challenge Game Frame* (CGF). A role-playing activity that is played to create SGs for analysing humans' decision-making. In the case of this study, collaborative role-playing with the CGF creates games that examine expertise and related conduct in privacy decision-making. A SG played to create other SGs is also referred to by some game studies scholars as Meta Serious Game [29]. The CGF is built on an action-oriented conception of the theory of affordances [30], [31] and exists in evaluated paper-based [11] and digital instances [12], [13].

The present research advances the digital variant of the role-playing activity, the digital cards and board, designed to help researchers create SGs for examining privacy decisions by integrating and evaluating AI support. As a SG design process is an iterative process, that demands considerable time and monetary resources coupled with collaboration between different stakeholders [6], [11].

In this context, the research examines GPT-4, an accessible generative pretrained transformer AI language model, as a sustainable and efficient human role-player substitute in the co-design activity, emphasising resource efficiency.

In playing the role of the teacher, the AI is tested for supporting the co-design process regarding ideation and balancing of SG design ideas while leveraging the knowledge base from its vast training data. Appropriate applicability of AI as co-playing partners thereby bears the potential for greatly improving the feasibility of iterative SG design cycles by having AI role-players step in for unavailable human experts.

The respective research questions guiding the investigation of these objectives were:

1. *How does supplementing the teacher role with a text-generative AI language model support students in collaboratively ideating and balancing affordance-oriented SG concepts compared to a human playing the teacher role?*
2. *How do student players engaging with the AI teacher in card-based role-playing for SG affordance concepts perceive the applicability of AI collaboration?*

2. Method and Material

2.1 Research approach

The study follows the design science research approach [32] by which a supportive co-creation role-playing SG is cyclically developed on the basis of the affordance theory [30], [31] aligned with the requirements of decision-making research [1]. The resulting card-based role-playing incorporates various *role cards* for the *player*, *teacher*, *researcher* and *designer* roles, a *board* and *stepwise instructions* for ideating privacy-related game challenges *balanced on the affordances of each role* and the *context of play* [11]. In the present user study, the digital variation of the card and board role-playing activity [12] was extended with the potential of AI taking a role in the collaborative ideation for SGs used in researching privacy decision-making.

An example of such an ideated SG might focus on raising awareness about data sharing and tracking, combining storytelling and educational strategies while evaluating the time of thinking about each privacy decision. Decision-making might be slower and more reflective or faster, more automatic, with little consideration of one's own experience in similar decision situations [18], [20].

The utility of the role-playing co-design activity in ideating and balancing SG affordance concepts through collaborative playing was evaluated during sessions where student teams engaged in role-playing. One group received no AI support, and the teams collaborated online via videoconferencing with the CGF, with each team member playing one of the roles. For the other group in the research design, the teams co-played with AI (GPT-4), playing the teacher role locally in a class on desktop computers, while the remaining three roles were played by the human players (Figure 1).



Figure 1. Two student teams co-playing the CGF together with a GPT-4 powered AI playing the teacher role.

The study examined the perceived support in *idea generation*, *concept balancing* and *applicability* of the card- and board co-playing between the teams with/without the AI in the teacher role by in-game satisfaction sliders regarding the concept for each role (Figure 2) and detailed data collection by a post-game questionnaire (Sect. 2.3.2). The teams co-playing with the AI were asked additional Likert-scale type questions on their perception from collaborating with the AI teacher on the last screen of the questionnaire. The ratings were analysed to evaluate for acceptable applicability of AI substituting for a human playing the teacher role. In alignment with the research objectives, the null hypotheses established for the empirical study were:

H_{0a} : There are no significant differences in perceived ideation, balancing support, or applicability of the card-based role-playing between students co-playing with the AI in the teacher role and students co-playing with a human-played teacher role.

H_{0b} : Students' perceptions of collaborating with a GPT-4 powered AI teacher in card-based role-playing do not significantly register below the median in the Likert-scale responses.

2.2 Elements and features of the AI-enhanced collaborative role-playing

2.2.1 Digital co-playing instance of the affordance-oriented stakeholder SG design model

The digital variant of the CGF card- and board activity builds on the goal-directed design model based on SG stakeholder role affordances elaborated and tested in the design science research process targeting privacy decision-making [11], [12]. It comprises 12 decks totalling 150 cards, offering design suggestions for *player*, *teacher*, *researcher*, and *designer* roles. It also includes *four* universal decks for context definition. These decks include suggestions for *who* the intended audience is, *when* and *where* the SG is intended to be played, and a dedicated deck focused on privacy challenges to set the domain challenge (e.g., sharing of false information or the privacy paradox). To enable multi-perspective dialogue and balanced SG creation for privacy decision research, the CGF offers a playboard and concise, sequential guidelines for SG ideation, balancing the concept considering game-induced uncertainty [33] and

discussing/resolving stakeholder affordance conflicts for observing potential disruptions of flow [8] (Stage 5, Figure 2).

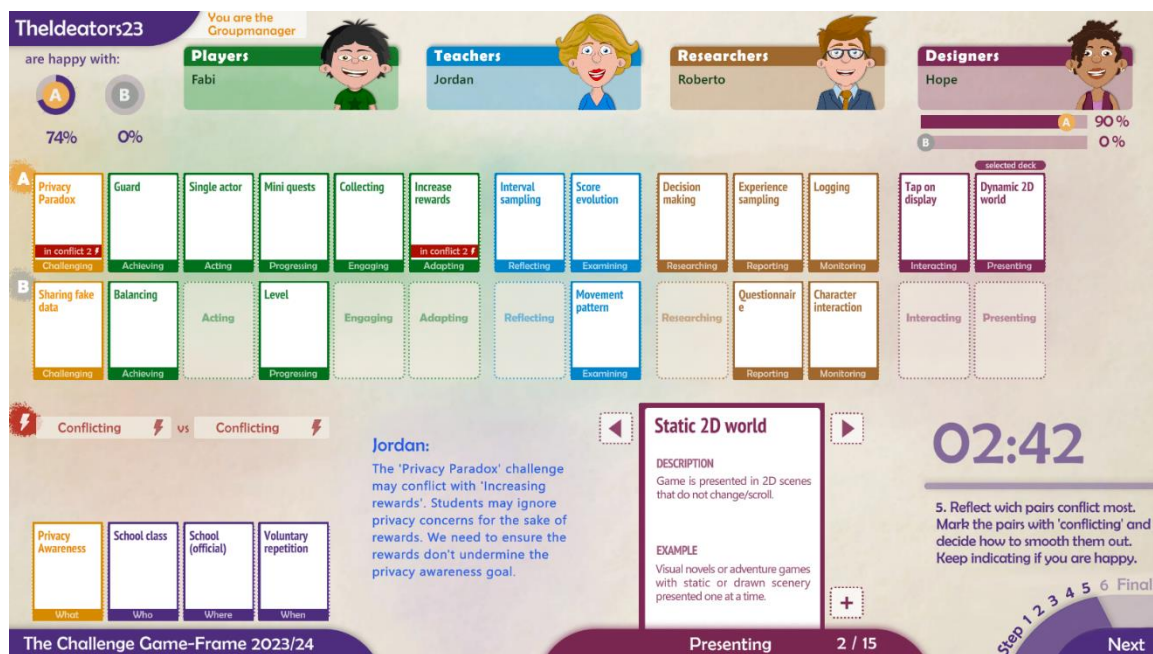


Figure 2. The digital CGF, role-playing cards and board based on stakeholder action-affordances for collaborative ideation and balancing of SGs applicable in privacy decision-making research; GPT-4 powered AI playing the teacher role and naming itself "Jordan". Together with the other co-players, in this stage, the AI teacher identifies/marks a conflicting pair and provides commentary on its disruptive nature.

The CGF, as shown in Figure 2, incorporates two challenge streams: main game challenge A and alternative affordances stream B, where players can place alternative design suggestions. Based on the satisfaction rating (Fig. 2, percentage sliders below designer role), the group can decide to shift focus from main to alternative. This dual-stream approach fosters creativity by being able to try out alternating design actualisations and helps narrow choices from the extensive card decks [34], [35], [36]. Players rate their satisfaction with concepts as a quality criterion [37], and avatars express emotional feedback through changing facial expressions.

2.2.2 Role cards supporting goal-directed acting, board and stepwise co-playing instructions

The role-playing begins with selecting role-independent context cards and a privacy challenge, which are decided upon by the whole team. Role-players then choose design suggestions for their role-specific affordance cards, goal-driven with individual authority. These cards, based on each role's action-oriented affordances [31], were elaborated and tested in a SG design model with paper-based and digital instances in the cyclic design science research process [11], [12]. These role-specific cards are placed and discussed for conflicts in a stepwise process. The *player* role chooses how the game affords *achieving* (e.g. fighting/evading obstacles or guarding resources), *acting* (e.g. collaborating in teams), *progressing* (e.g. mini-quests, time limits or levels), *engaging* (e.g. surprise, exploration or certification/mastery) and is *adapting* to developing player skills (e.g. introducing elements, reducing resources).

The *researcher* role decides on how the game is offering *researching* privacy decision influences (e.g. decision presentation factors, time pressure, or individual risk propensity), *reporting* (e.g. post-game questionnaire or in-game continuous feedback/sampling) and unobtrusive *monitoring* (e.g. logging time or other factors). While the *designer* role is responsible for how the SG affords *interacting* (e.g. tap or swipe gestures) and *presenting* (e.g.

2D static visuals or 3D immersive presentation), the *teacher* role, which is prepared and investigated for AI substitution in this study, is choosing design suggestions for *reflecting* (e.g. reflective journaling, decision summary) and *examining* (e.g. score evolution or movement patterns).

The board and provided rules are designed to facilitate ideating and discussing balance/conflicts of SGs domain challenge, investigating privacy decisions. The role-playing is structured in six timed co-designing steps (Sect. 2.3.2), which are explained and read out via voiceover at the beginning of each step, in line with research suggesting the importance of clear goals and stepwise processes in co-design [35], [36], [37]. Players engage with the game through their web browsers; actions like moving cards and markers are synchronised over the internet while discussing the game concepts in person locally or via videoconferencing.

2.2.3 Integrating the AI-powered teacher role based on transformer architecture language model

AI models like GPT-4, based on architectures such as the transformer [38], bear great potential to assist in ideation and identifying conflicts within game concepts. They excel in generating ideas from text prompts, expanding on narratives, and simulating scenarios [39], [40] – feats suggesting them as a viable substitute when continuous human participation is not feasible or available. Nonetheless, limitations evident in aspects like emotional intelligence and spontaneous creativity must be recognised, which are integral to human participation [41], [42]. Making emotional feedback to the team was thereby an aspect already addressed as it is essential for remote ideation teams [12], with the illustrated role avatars changing facial expressions from earnest to happy aligned to their satisfaction with the SG affordance concepts (Figure 2).

On these considerations, the digital instance of the CGF was enhanced by integrating an interface to the GPT-4 API offered by openai.com. Taking a general approach, the interface developed as a component to the base implementation in C# using the Unity Engine (version 2022.3.10f1) provides to substitute all roles by the GPT-4 transformer model.

Therefore, a system message was set up to be sent to the API that instructs a GPT-4 model about the context and the nature of the role-playing activity and, in the present study, its role in the team as a collaborative teacher. The appropriate system prompt to GPT-4 was iteratively tested in the API Playground (platform.openai.com/playground) until the model provided satisfactory consistent replies. Moreover, the chatbot at chat.openai.com, based on the same transformer model, was asked how the system prompt can be improved for consistency.

Ultimately, a *systematic protocol* emerged that instructed the role to GPT-4 and prepared the model to receive *four different requests* for co-playing with human players. In iterative improvement sessions, ChatGPT suggested writing important instructions in capitalised form and including a dedicated instructions section to improve the consistency of replies, which was, therefore, added to the end of the system prompt. Table 1 lists the resulting system prompt sent to GPT-4 after the first request that asked the model to give itself a gender-neutral name. Playtesting showed the model was naming itself most often "Riley", "Quinn" or "Jordan".

With GPT-4 (at the time of the study), each API request must include the system prompt, as the model does not retain context or history from previous requests. Consequently, the system prompt was added to a message history and sent with each following request, making it a prompt of two messages, a system prompt and one user prompt containing the currently needed request during co-playing.

Table 1. System prompt – iteratively elaborated in the API playground and with the Chatbot of GPT-4

Role: You're a teacher in collaborative role-playing activity for creating a well-balanced SG concept between different role perspectives. You are collaborating with a player, researcher, and designer role to balance role affordances.

Request 1: Evaluate SG Concept. Choose from current role cards by id. Evaluate the whole concept for balance and respond what you suggest the SG should best afford for: 'reflecting' and 'examining' with one suggestion each including description and example. Follow your goals as a teacher and consider your knowledge and experience for encouraging reflection and examining learners. Then you decide if you return an optimal existing card for an affordance. Sometimes you will draw from your experience and create a new card (id=0) but regard the overall SG balance. Either way strictly reply in format of the card signature and provide no further explanations.

<prompt>[List of your cards] Evaluate: [SG concept, all cards]
<card signature>ID;affordance;3-word headline;15-word description;15-word example
<comment>com: [word limit: 50 words in a colloquial style, e.g. "we could.."]
<reply>STRICTLY ONE card signature for 'reflecting' and one for 'examining' and ONE overall comment on these choices

Request 2: Rate how happy you are with the SG concept from your role's perspective

<prompt>Rate SG:
<reply>Float (0-100, 2 decimals)
<comment>com: [word limit: 50 words in a colloquial style, e.g. "we could.."]

Request 3: Identify conflicts between any card pair

<prompt>Conflicts: [SG concept, all cards]
<reply>return exactly two card signatures of any of the affordances in the concept that conflict most and one comment
<comment>com: [word limit: 50 words in a colloquial style, e.g. "we could.."]

Request 4: Pitch SG with great creativity

<prompt>Pitch: [SG concept, all cards]
<title>a catchy title [word limit: 5]
<summary>sum: a detailed summary of creative gameplay envisioned from the provided SG concept [word limit: 80]
<reply>title and summary

Instructions:

Respond ONLY to the specified request number!
Ignore unsolicited requests.
Reply only one suggestion per affordance - the most balanced one.
Reply STRICTLY in provided FORMAT, ORDER and WORD LIMITS.
Reply without any markup, no headers and no entry words.
Use new line for each signature or comment in the reply.
Your comments are in a colloquial style as a teacher.
Use easy understandable vocabulary in comments.
You may receive each request multiple times.

Prompts were then sent during the respective stages of co-playing. Only stream A's main challenge was considered for AI collaboration, keeping stream B free for alternative card placements by the human players. Importantly, the timing of the AI teacher's evaluation of the concept and placement of role cards had to be set. For this, requests 1 and 2 about ideation and rating the currently laid out SG affordance concept were sent at timed intervals from Unity's `InvokeRepeating()` method. Starting 50 seconds into the 15-minute session, where each role sets its cards individually, the AI teacher was first requested to evaluate the concepts and place appropriately balanced design suggestions for the teacher's *reflecting* and *examining* affordances. The request was then repeated every 90 seconds.

The AI was also asked to provide a 50 words commentary on the placed design suggestions displayed to the whole team on the board. The AI teacher only received the affordance card categories and the headlines of its role-specific card design suggestions, as illustrated in the exemplary request 1 prompt in Table 2. The GPT-4 model was urged to include a description and an example of the respective affordance design suggestion derived from the design suggestion headline in each reply.

As indicated in Table 1, GPT-4 was further asked to draw on its knowledge to come up with new affordance design suggestions that are not currently in the card decks. In doing so, it contributed to building the knowledge base by providing descriptions/examples for existing design propositions and uncovering new design suggestions for teacher affordances. Practical findings on what designs the AI teacher suggested most for both affordances are listed in Sect. 3.4.

Table 2. Request 1 example: prompting available teacher affordance cards and the current main concept

Request 1:	Evaluate:
100;reflecting;Interval sampling	Overall topic: Privacy Awareness
99;reflecting;Journal	182;who;Yourself;You, your family or other social groups or communities you are part of.
98;reflecting;Summary	183;where;Home;The game is played at home in-doors during free time.
97;reflecting;Team questioning	184;when;Once;The game is played once by every player.
96;reflecting;Character asking	
95;reflecting;Between level asking	
94;reflecting;In-game questions	3;challenging;Privacy Paradox;Indicating knowledge about data breaches or misuse of data but acting against better knowledge when sharing or consenting.;Knowing about "Cambride Analytica Data Scandal" but still not reading displayed privacy policies and giving consent without reflection.
93;reflecting;Time machine	17;achieving;Fight obstacles;Overcome obstacles;Shooter games that have boss enemies to overcome.
92;reflecting;Decision view	none;acting;undecided
108;examining;Use of tools	none;progressing;undecided
107;examining;Score evolution	none;engaging;undecided
106;examining;Pre/Post scene test	none;adapting;undecided
105;examining;Movement pattern	none;reflecting;undecided
104;examining;Interaction records	none;examining;undecided
103;examining;Spy character	109;researching;Decision making;Analysing rational influences that lead to players decisions and choices in the domain goal.;Investigating the influence of presented knowledge and transmitted facts/consequences on making privacy decisions.
102;examining;Time logging	none;reporting;undecided
101;examining;Click/tap-stream	none;monitoring;undecided
	124;interacting;Tap on display;User taps somewhere on the display to trigger action in the game.;Tap on upper half of display triggers running forward of character. Or holding the thumb on the lower half of display slows down etc.
	none;presenting;undecided

Request 2 for providing a satisfaction rating was also timed, started at 90 seconds in the session and repeated every 120 seconds from there on. Correspondingly, in the stage that observes disruption of flow (see Sect. 2.3.2 for all co-playing stages), the AI teacher was prompted every 50 seconds to identify the most conflicting design suggestion pairs from its perspective. Thereby, the GPT-4 transformer was also prompted to comment on the nature of the disruptive conflict. The most conflicting actualising propositions are likewise detailed in the practical findings in Sect. 3.4.

Finally, in the last stage, the AI teacher was prompted to provide an initial pitch summary of the overall SG affordance concept with request 4, helping the co-players sketch a written synopsis of their conceptualised SG idea about privacy decision-making. The requests were sent over the API to GPT-4 with the temperature setting on standard 0.7. In iterative design cycles, the temperature setting was explored to provide more creative and consistent responses. Reducing the temperature setting did not result in more uniformly formatted responses. Raising the setting did not enhance creativity in replies, and values above 1.3 caused GPT-4 to answer nonsensical/scrambled text.

2.3 Co-playing sessions and data collection

2.3.1 Co-playing sessions with student teams

The CGF was co-played by teams of students in online and local gameplay sessions at two European universities. The students were composed internationally and attended Computer Science study programs taught in English. The group playing in online teams ($n = 32$) was playing with a human player in the teacher role, while the group playing locally ($n = 23$) collaborated with GPT-4 playing the teacher role. The students from both groups reported a low skill median in game design (non-AI group $Mdn = 2$, AI group $Mdn = 1$; Likert scale: 1, none; 7, professional).

2.3.2 Procedure and data collection

During the two-hour sessions, participants initially received an overview of SG design principles provided by the researcher who led the session. The facilitator then detailed the workings of the collaborative role-playing activity, including its interface and guidelines. This was followed by a one-hour collaborative role-playing session. Instructions for each step of the co-playing were both visually presented by text in the CGF interface and narrated by a voice recording. The group manager, decided by the team, initiated each phase of the activity by pressing the next/start button, ensuring synchronisation of the co-playing session for each team (Figure 2). The rules guiding the co-ideation were:

1. Agree on who plays the role of either player, teacher, researcher or designer
2. Define the context of the game: domain, target group, location/time of play
3. Individually read through role-assigned cards and pick favourites
4. Co-create/balance a game challenge: starting from left to right, discuss ideas from the role-oriented cards or create custom cards and fill all slots of at least one challenge stream (A or B). Each role places/argues its proposals while rating the whole concept
5. Identify conflicting pairs of cards in the game concept and balance out the potential flow breaks by discussing alternative picks or another group agreement
6. Agree on the final picks, define a working title and write a game plot summary

Throughout the co-playing, players maintained their assigned roles and held the decision authority over their specific card decks. In the third step, they closely examined their cards to present them to others effectively. Players were advocating their design choices and resolving conflicts in the fifth stage through group discussions, where conflicting card pairs were marked (Figure 2) and addressed by either retaining them or opting for alternative designs for game balance, observing flow and game-adapted uncertainty (e.g. surprises, hidden information, confusion results or randomness) [33]. Players thereby rated their satisfaction with the current state of the game concept during the co-playing using rating sliders.

Following the co-playing, each participant independently completed the post-activity survey based on prior research studying design support and was employed throughout the design science research process [12], [43]. Each co-player utilised a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree), to evaluate their impressions. The assessment focused on the perceived ideation and balancing support and the perceived applicability of the role-playing elements, cards, roles, and the board. The teams collaborating with the AI teacher were able to rate their perceptions with additional questionnaire items focused on the applicability of AI as a substitute.

Both the ideation and application aspects were measured through 4 specific items each. The balancing support was assessed with 8 items and 10 items focused on AI applicability, asking only the group collaborating with the AI teacher, as reported in the following section.

3. Results

Analysis of data revealed non-normal yet similar distributions (Kolmogorov-Smirnov-Z test, $p > .05$) and comparable variances (Levene's test, $p > .05$) across groups. Consequently, non-parametric Mann-Whitney U analysis was performed ($\alpha = 0.05$) to compare distribution medians in the between-group design, following guidelines for hypothesis testing [44], [45].

For analysis of the second research question, addressing only the perception of AI collaboration with the group co-playing with AI, the corresponding one-sample Wilcoxon Signed Rank Test was performed [44]. It provides for a detailed examination of the median responses from student teams against the neutral benchmark (i.e., 4) on the Likert scale of the questionnaire.

3.1 Ideating and balancing support

Considering the *ideating* support perceived by the two groups during co-playing the CGF, the NoAI group showed a generally comparable outcome in the analysis (Table 3). When analysed for agreement with statements related to generating and developing new ideas, similar concurrence was observed between the groups collaborating with the AI teacher and the NoAI group. Specifically, for the statement "*The boardgame provides guidance to develop new ideas*," both groups showed an equal percentage of agreement, 70%. Similarly, the statement "*Had ideas I would not have had without the cards*" garnered agreement percentages of 72% for the NoAI group and 70% for the AI group.

Table 3. Perceived ideating support between the groups co-playing without/with AI teacher

	Agreeing % NoAI/AI	Mdn NoAI/AI	Mode NoAI/AI	U	z	p
Ideating						
The boardgame provides guidance to develop new ideas	70/70	5.5/5	6/6	349.5	-0.324	0.746
Had ideas I would not have had without the cards	72/70	5/6	5/7	417	0.855	0.392
Cards helped to focus on ideas	72/57	5.5/5	6/5	232	-2.392	0.017
Cards helped to improve or fine-tune existing ideas	75/52	5/5	5/5	258	-1.927	0.054

The medians for both groups were closely aligned, with the NoAI group at 5.5, the AI group at 5 for the first statement, and both groups at a median of 5 for the second statement. The modes were also consistent, with the first statement at 6 for both groups and the second at 5 for NoAI and 7 for AI. The Mann-Whitney U test results for these statements did not indicate significant differences, with p -values of 0.746 and 0.392, respectively.

However, a divergence was observed with the statement "*Cards helped to focus on ideas*", where the NoAI group's agreement was noticeably higher at 72% compared to the AI group's 57% (Figure 3). The median responses for NoAI/AI were 5.5/5, and the modes were 6/5. This was the only statement in the ideation category that showed a statistically significant difference with a U value of 232, z -score of -2.392, and a p -value of 0.017, indicating a significant difference in the perceived effectiveness of the cards in focusing ideas when AI was involved. The effect size for this difference was found at a moderate level with $r = 0.37$.

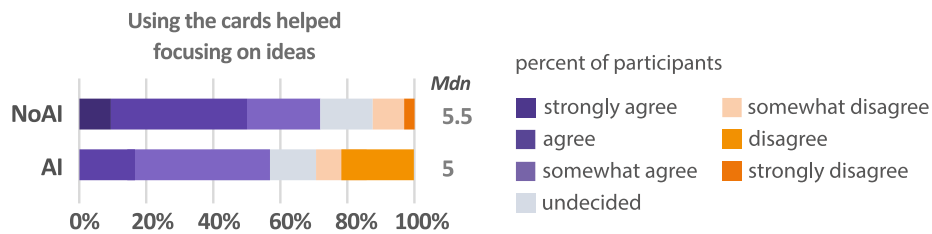


Figure 3. Perceived ideating support from using the cards – significantly different rating between groups without/with the AI co-playing as the teacher

Markedly, none of the players collaborating with the AI teacher expressed strong agreement or disagreement in perceiving the cards as helpful for focusing, indicating a perception that spreads out on a more neutral perception of support.

In the *balancing* aspect of the co-playing activity, where participants considered the effectiveness of the CGF in balancing the SG affordance concepts, the responses were again similar between the NoAI and AI groups for most statements (Table 4).

Table 4. Perceived balancing support between the groups co-playing without/with AI teacher

	Agreeing % NoAI/AI	Mdn NoAI/AI	Mode NoAI/AI	U	z	p
Balancing						
Using the boardgame helped balancing game to domain goal	69/52	5/5	7/6	259	-1.895	0.058
Using the boardgame helped balancing the SG concept parts	78/52	5/5	5/4	318	-0.876	0.381
Using the cards helped balancing the SG concept parts	75/52	5.5/5	6/7	296	-1.252	0.211
Playing the roles helped identifying conflicts	75/70	5/5	6/5	351	-0.298	0.766
Playing the roles helped balancing game to domain goals	66/65	5/5	5/5	331	-0.649	0.516
Playing the roles helped balancing the SG concept parts	69/61	5/5	7/5	307.5	-1.056	0.291
Using the board helped balancing game to domain goal	66/57	5/5	6/4	342	-0.455	0.649
Using playboard helped structuring and visualising SG concept balance	84/74	6/5	6/5	289.5	-1.390	0.165

However, statements about perceived balancing support that specifically mention the cards or the overall boardgame activity, such as "*Using the cards helped balancing the SG concept parts*" (NoAI 75%/ AI 52%) and "*Using the boardgame helped balancing the SG concept parts*" (78%/52%) showed a slight tendency of fewer participants feeling supported collaborating with the AI teacher. Nonetheless, group medians were thereby still consistent with 5.5/5 and 5/5, respectively, with mode values also showing only minimal variation. As listed in Table 4, none of the balancing support results showed significant differences in the Mann-Whitney *U* test, with p-values consistently above the 0.05 threshold. Similar to the ideation qualities assessed, both groups consistently rated balancing support with a median of at least 5, exceeding the midpoint benchmark of the rating scale.

The overall pattern in the balancing results indicates that, while there are marginal variations in agreement percentages between the NoAI and AI groups, the statistical analysis did not

reveal significant differences in the perceived balancing support between the groups collaborating with or without AI playing the teacher role in the role-playing co-design activity. The median and mode responses were generally consistent across both groups for all balancing statements.

3.2 Applicability and final SG concept satisfaction

Regarding the applicability of the CGF for collaborative creation of SGs for use in privacy decision-making research, the results followed the same between-group pattern for felt ideation and balancing support (Table 5). Co-players' enjoyment of using the CGF was reported by 66% of the NoAI group and 57% of the AI group, with both groups presenting a median response of 5 and a mode of 5. The analysis indicated no significant difference in the levels of enjoyment between the two groups ($U = 328, z = -0.699, p = .484$).

Table 5. Perceived applicability of the role-playing SG between the groups co-playing without/with AI teacher

Perceived applicability between groups	Agreeing % NoAI/AI	Mdn NoAI/AI	Mode NoAI/AI	U	z	p
Using the boardgame was fun	66/57	5/5	5/5	328	-0.699	0.484
I can imagine using the boardgame on my own in a group	66/61	5/5	5/7	359	-0.156	0.876
I can imagine using the boardgame to create SGs for researching decision-making	59/61	5/5	6/6	365	-0.053	0.958
How satisfied are you with the game you created?	75/78	5/5	5/6	387	0.336	0.737

Regarding the potential for individual use within a group setting, 66% of the NoAI group and 61% of the AI group could imagine using the CGF on their own. Both groups recorded a median response of 5, with a mode of 5 for the NoAI group and 7 for the AI group. The perceptions between the groups showed no statistical distinction ($U = 359, z = -0.156, p = .876$). The question of using the role to create SGs for researching decision-making yielded agreement percentages of 59% for the NoAI group and a slightly higher 61% for the AI group. Both groups had a median of 5 and a mode of 6, and the U test revealed no significant difference ($U = 365, z = -0.053, p = .958$), indicating a similar perception of the CGF's utility for creating SGs used in privacy research in both groups.

Lastly, satisfaction with the SG concept created was high in both the NoAI (75%) and AI (78%) groups, with both reporting a median of 5 and modes of 5 and 6, respectively. This reflects a positive assessment of the final SG concept about privacy research in both scenarios, with no significant difference found in the satisfaction levels ($U = 387, z = 0.336, p = .737$). The comparable satisfaction reported between the groups with the post-questionnaire is confirmed by the results analysed from satisfaction rated with sliders by each role during co-playing. Role-players in both groups expressed a high satisfaction rating with their co-created SG concepts for privacy research (Figure 4).

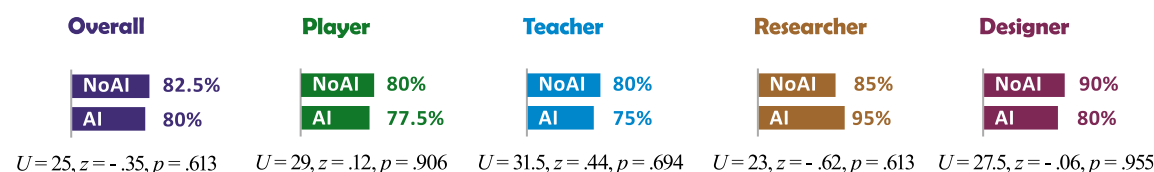


Figure 4. Role-specific satisfaction (*Mdn*) with the final co-created SG concept between the groups co-playing with or without AI in the teacher role

The overall satisfaction median was 82.5% in the NoAI groups and 80% in the teams collaborating with the AI teacher. There were no significant differences in role-specific satisfaction either, with all median ratings in both groups above 75%.

Markedly, the teacher role in the AI collaborative scenario played by GPT-4, which consequently also rated the satisfying quality of the SG concept (request 2, Table 1), showed no significant difference compared to the satisfaction of human players in the teacher role (Figure 4).

In general, the group's collaboration with an AI in the teacher role during the co-playing activity was found to be widely comparable to the NoAI group across all measures of applicability, fun, potential for individual use within a group, utility in designing SGs for decision-making research, and overall satisfaction with the created privacy SG concept. However, the teams co-playing with the AI teacher perceived a moderate decrease in the effectiveness of using the cards to focus their ideation process. Therefore, the results suggest rejecting the null hypothesis H_{0a} .

3.3 Perception of collaborating with AI teacher

Upon examination of the results from the Wilcoxon Signed Rank test, analysis supports the null hypothesis H_{0b} , which states that students' perceptions of collaborating with a GPT-4 powered AI teacher in card-based role-playing do not significantly deviate below the median on Likert-scale responses (Table 6).

Table 6. Perception of GPT-4 powered AI playing the teacher role when co-creating privacy SG concepts

<i>AI teacher perception</i>	<i>Agreeing %</i>	<i>Mdn</i>	<i>Mode</i>	<i>W</i>	<i>z</i>	<i>p</i>
The AI provided new ideas I hadn't thought of during ideation	48	4	3	131	0.148	0.882
Collaborating with the AI enhanced the game concept's diversity	40	4	4	66.5	-0.079	0.937
The AI helped broaden the scope of ideating	48	4	5	96.5	-0.673	0.501
The AI's ideation suggestions aligned with our goals in game concept creation	30	4	3	68.5	-0.751	0.452
Collaborating with the AI helped balancing the game concept	40	4	5	126	0.376	0.707
Collaborating with the AI helped monitoring the game flow in the game concept	30	4	4	54	-0.347	0.728
Collaborating with the AI helped identifying conflicts between the game parts	40	4	3	89.5	-0.225	0.822
I felt comfortable with the AI's role as a teacher during game concept creation	40	4	4	57	-0.172	0.863
Collaborating with the AI helped balancing the game goal with the domain goal	40	4	5	62	-0.700	0.484
I perceived the AI as an equal collaborator during game concept creation	40	4	4	62.5	-0.289	0.772

Specifically, the median responses for each item were at the neutral midpoint of 4 on the Likert scale, indicating neither agreement nor disagreement on average. For instance, when evaluating if the AI provided new ideas not previously considered, the median and mode were both at 4, with a Wilcoxon W of 131 and a non-significant p -value of .882. Similarly, for the statement

regarding AI's role in enhancing game concept diversity, the median remained at 4 with a mode of 4, and the test yielded a W of 66.5 with a p -value of .937.

Across all items, the median scores did not fall below the neutral midpoint, and none of the p -values approached the significance level of 0.05. The closest to this threshold was the statement about AI's alignment with game creation goals, which had a median of 4, a mode of 3, a W -value of 68.5, and a p -value of .452, yet still not reaching significance.

In summary, the statistical data indicate that perceptions of the AI's performance in a teaching role were consistently average or above. These results suggest that the GPT-4 transformer AI, as implemented in this study, meets the threshold for adequacy as a substitute in the co-playing activity.

3.4 Most frequent AI-suggested design propositions and identified design conflicts

Over the course of the role-playing sessions, the AI teacher made 99 design suggestions for the *examining* affordance and 96 for how the SG should offer *reflecting* about their privacy choices to players. In addition to choosing design propositions for these teacher affordances, the AI was prompted to provide a short description and example for each of them, regardless of whether they were created new by the AI or existing cards in the decks. With that, the AI was sourced to refine the descriptions and examples for each of the CGFs' design suggestions in the teacher role. GPT-4, in the teacher's role, primarily based its design recommendations on the suggestions already present in the card decks. Table 7 displays the design suggestions made by the AI teacher during co-playing, which were predominantly drawn from existing cards in the decks.

Table 7. Most frequent AI-suggested design propositions for examining and reflecting with a privacy decisions SG

Teacher affordance	Frequency	Design proposition	Description
<i>reflecting</i>	33	Summary	A summary of qualified and unqualified choices/decisions is presented to the player after each level or at specific events in the game.
	30	In-Game Questions	Present questions at specific points in game that assess knowledge or skills.
	9	Character Asking	Knowledge or skill is assessed by an in-game character in an obvious fashion.
	8	Interval Sampling	The player is asked to note down specific domain knowledge or skill-related aspects in certain intervals or asked a quiz-like question.
	6	Journal	Actions and decisions of the player are recorded in a journal, where they can be accessed at any time during play.
	6	Team Questioning	The team is getting together and reflects about knowledge/skills and domain goal progress.
	2	Decision View	Present the player's choices/decisions as summary and mirror with progress of other players or own results from previous runs.
	1	Between Level Asking	Ask questions that monitor progress on the domain goal related to the past level or chapter.
	1	Real-Time Reflection	After making a privacy decision, the game pauses to allow players to consider the implications.

examining	33	Pre/Post Scene Test	The players are asked or logged before a specific situation that transmits domain knowledge or skill and assessed again after the scene.
	30	Movement Pattern	The players' movements in the game are recorded and analysed for patterns in the solution path.
	9	Score Evolution	The players' score is analysed over time to see if progress on the domain goal is sufficient, and the mapping of domain goal and score/game goal is adequate.
	8	Use Of Tools	The players' progress on the domain goal is monitored in accordance with (logging) usage of certain tools in the game that are mapped on skill or knowledge.
	6	Interaction Records	The interaction (e.g. frequency pushing false buttons) is monitored in certain problem-solving tasks to analyse problems with the domain.
	6	Click/Tap Stream	Assess the clickstream of the player to evaluate progress in the domain goal and individual differences in approaches.
	2	Spy Character	Knowledge or skill is assessed by an in-game character in an unobtrusive/hidden fashion.
	1	Time Logging	Assess time of answering questions or interaction time with characters discussing the domain goal.

During the teams' role-playing sessions, the transformer came up with only two newly generated affordance cards. First, "Real-time reflection" described as "After making privacy decisions, the game pauses to allow players to consider implications". And second, "Reflective journaling" described as "After each level, players write in their journal about their decisions and what they learned".

When discussing conflicting pairs between the roles' affordances (step 5), the AI teacher identified several pairs of cards over the whole concepts that conflict from its point of view. Table 8 presents the conflicts the AI teacher identified evaluating the SG concept of its group across all the roles' design propositions.

Table 8. Conflicting design proposition pairs identified by the AI teacher considering the whole SG concept balance

<i>Affordance card 1</i>	<i>Affordance card 2</i>	AI teachers comment on SG concept conflicts/potential flow breaks
<i>acting:</i> Single Actor	<i>presenting:</i> Social Media Context	The single actor card and the social media context card might clash. The game's individualistic nature might not work well with the social media context, which usually involves interaction and collaboration.
<i>acting:</i> Team Confrontation	<i>progressing:</i> Score	The team confrontation and score progression might conflict. If the game is too competitive, it could discourage cooperation within teams.
<i>achieving:</i> Problem Solving	<i>acting:</i> Team Confrontation	The problem-solving card and team confrontation card could conflict. If the game is about solving puzzles individually, it might not align well with teams confronting each other. We need to clarify this.
<i>achieving:</i> Minimise Time	<i>reflecting:</i> Reflective Journaling	The time minimising goal might rush players and limit their reflection time. We need to balance achievement and reflection.
<i>challenging:</i> Lack of Knowledge	<i>achieving:</i> Outwit a Team/Player	The 'Lack of Knowledge' challenge could be difficult to overcome if the main achievement is to 'Outwit a team/player'. This could potentially create an unfair advantage for those with prior knowledge about privacy laws.
<i>challenging:</i> Lack of Knowledge	<i>acting:</i> Team Collaborators	The challenge of 'Lack of Knowledge' might conflict with 'Team collaborators'. If team members have varying knowledge levels, it could lead to imbalance and frustration.

<i>challenging:</i> Health Tracking	<i>monitoring:</i> Logging	The health tracking challenge could conflict with the logging monitoring method. Players might feel uncomfortable with their health data being logged, especially in a game about privacy.
<i>challenging:</i> Health Tracking	<i>monitoring:</i> Psychophysiology	There's a conflict between the health tracking and psychophysiology cards. We're trying to raise privacy awareness, but we're also collecting sensitive health data. It's a bit contradictory, don't you think?
<i>challenging:</i> Deception	<i>examining:</i> Pre/Post Scene Test	The Deception card conflicts with the Pre/Post scene test card. It's hard to examine understanding when players are being deceived. We need to ensure transparency for effective learning.
<i>challenging:</i> Sharing Fake Data	<i>monitoring:</i> Team Decisions	The challenge of sharing fake data can conflict with the monitoring of team decisions, as it might encourage misleading behaviour for the sake of winning the game.
<i>challenging:</i> Excessive Permitting	<i>presenting:</i> Cartoon/Abstract	The abstract cartoon style might make it difficult for players to connect the game's privacy challenges to real-world scenarios. Maybe we could use more realistic visuals?
<i>examining:</i> Deception	<i>achieving:</i> Problem Solving	The 'Deception' challenge might frustrate elderlies, making problem-solving less enjoyable. We need to balance difficulty and player capabilities. [Target group was seniors]
<i>engaging:</i> Mastery	<i>adapting:</i> Raising Difficulty	The 'Mastery' card could conflict with 'Raising difficulty' card. If the difficulty rises too quickly, it might discourage players from mastering the game. We need to ensure a balanced progression.
<i>examining:</i> Score Evolution	<i>monitoring:</i> Psychophysiology	The score evolution could discourage players if they're not making progress fast enough, while monitoring psychophysiology might stress them out, which could affect their performance. We need to balance these aspects.
<i>presenting:</i> Cartoon/Abstract	<i>monitoring:</i> Psychophysiology	The cartoonish presentation might not be suitable for the serious topic of privacy awareness and could conflict with the psychophysiological monitoring, which requires a more realistic and engaging environment.
<i>reporting:</i> Character Asking	<i>examining:</i> Movement Pattern	The conflict arises between the NPC asking specific questions and the game examining the player's movement patterns. The NPC's questions could influence the player's movements, skewing the data.
<i>where:</i> School (unofficial)	<i>interacting:</i> Push Real Buttons	Playing the game in school might conflict with the need for a hardware controller. We could consider a mobile version instead.

4. Discussion

4.1 Implications for human-AI collaborative role-playing in design activities

Integrating an AI transformer model in collaborative role-playing with humans for SG concept creation has proven to be a valid approach for occasional supplementation of human stakeholders. The evaluation of using a text-generative AI model, specifically GPT-4, in the teacher role during the SG design process focused on two main research questions. The findings shed light on both the potential and the limitations of such integration, providing a nuanced perspective on human-AI collaboration in collaborative role-playing game activities.

Looking at the findings regarding *research question 1*, the study's exploration of how AI supports ideation and balancing affordance-oriented SG concepts reveals that generative AI, in the form of a GPT-4, can effectively contribute to the co-creation of SG concepts. In general, the role-playing game activity was comparably supportive for the students collaborating with the AI teacher for ideating and balancing as it was for the groups with only human role-players.

This is particularly reflected in the similar high satisfaction both groups, those with AI teacher and those with a human in the teacher role, were expressing regarding their final co-created SG concepts about privacy decisions. While overall ideation and balancing support were also perceived about equal, there was one significant deviation in the perception of how the role cards helped focus on ideas.

Including the insights from observations during the playing sessions, an explanation for this impact of AI in the teacher role on focusing can be found in the timing of AI intervention. Human players engaged in dynamic discussions, gauging when others had finished speaking by their actions, such as card placement. In contrast, in its teacher role, the AI systematically expressed its insights and made decisions at set intervals, showing a clear difference in interaction styles. Human participants navigated the conversation flow through mutual cues, while the AI followed a predetermined schedule for input. This highlights the distinct communication methods between humans and generative AI in collaborative settings.

This finding, in line with Zhang et al. [22], points to the need for integrating AI into team dynamics and creativity. AI filling a role in the role-playing for SG concepts should thus be enabled to sense the ongoing discourse in the team. By that, it can determine more appropriate and dynamic intervention for communicating its insights and placing design suggestions in the overall SG concept.

Concerning *research question 2*, participants' perceptions of AI collaboration, with the GPT-4 model serving as a teacher, were generally positive yet neutral. This suggests acknowledging AI's value in the design process, balanced with an awareness of its limitations. The insights on AI's perceived applicability in the teacher's role show it can adequately substitute for a human stakeholder when necessary. In particular, this is reflected in the potential of AI for evaluating the balance of a SG concept and identifying conflicting design propositions (Table 8). However, as the participants' more neutral appreciation for AI in the teacher role points to, the transformer model showed limited generation of novel design suggestions for its role. The AI predominantly chose existing design suggestions from the card decks rather than coming up with alternative ideas or balancing approaches.

Synthesising the research insights of this study leads to *three main design implications* to consider for human-AI design collaboration activities:

AI as a supplemental role-player: AI is perceived as a viable supplemental tool rather than a replacement for human creativity, emphasising the importance of a symbiotic relationship between human designers and AI in the creative process. The study's results demonstrate that a text-generative transformer AI can adequately fill a role-playing part in a creative activity. It can successfully take the teacher perspective in co-creating concepts for SGs about privacy decisions together with human role-players following a guided process. In playing its role, the AI can provide helpful SG balancing suggestions from existing design propositions from the teacher's point of view. Thereby, it proves valuable for identifying conflicts between design choices in evaluating the overall balance of the role-oriented SG concept.

Timing of AI Intervention: An optimal timing for AI interventions and suggestions remains to be determined. This is crucial for ensuring that AI contributions are both relevant and timely, enhancing rather than disrupting the creative flow of the design process, as noted in the discussion on flow experiences [8]. It might entail devising strategies for AI to effectively understand the ongoing discourse among its team members while playing its role. AI should be enabled to time interventions such as commenting or communicating ideation suggestions aligned with the ongoing design discussion to avoid breaking idea focusing and creative flow. On a simplified basis, if sensing the discourse is not a viable option, it could be a consideration to let the human players more freely ask for AI suggestions on new ideas and the balance of a design.

Enhancing AI's Creative Input: The AI's limited production of novel design suggestions underscores the need for improved mechanisms to stimulate AI's creative contributions, possibly by refining the prompts or incorporating

mechanisms for dynamic learning during the design process. This limitation calls for further research into methods for enhancing AI's generative capabilities, perhaps through advanced training techniques or hybrid models that more effectively combine human creativity with AI's computational power.

4.2 Limitations

The study's exploration of AI's role in creating SGs about privacy decision-making highlights promising avenues for human-AI collaboration. Still, it is shaped by its contextual and methodological choices. Focused on a specific GPT AI model, the findings possess limits in generalising over other AI models. The study's participant group, consisting of students from two European universities, could have unknown effects on team dynamics. Diverse participant backgrounds, skill levels, and cultural contexts could thereby influence collaboration dynamics and perceptions of AI. Though instrumental in uncovering significant insights regarding the distinctions between AI and non-AI collaboration, the focus on quantitative methodology imposes limitations on exploring the subtleties inherent in the collaborative design process. Addressing these areas can enrich future research, further advancing our understanding of AI's role in creative collaborations.

4.3 Future research trails

Future research in SG design should explore deeper into the integration of AI support, particularly exploring its role as a collaborative partner in the design process. The potential of generative AI, such as GPT-4, to enrich the ideation and development phases of SGs presents an exciting opportunity to enhance privacy awareness and explore related decision-making. Key areas for further study include optimising AI's timing and contributions to align with the creative flow of design teams, enhancing AI's capacity for generating innovative design suggestions and exploring advanced training methods or hybrid models that combine human creativity with AI's computational strengths. With respect to the stakeholder roles involved in the goal-oriented design approach presented in this study, other AI models as supplementary agents and the substitution of more than one role in collaborative role-playing present promising avenues for future studies. These efforts aim to improve the design process of SGs and offer insights into leveraging AI in creative collaborations, pushing the boundaries of educational technology, privacy education and decision-making exploration in other domains.

5. Conclusion

This study investigated the integration of generative AI, particularly GPT-4, into the collaborative design of SGs, focusing on privacy decision-making. It examined the impact of AI in substituting the teacher role on ideation, concept balancing, and participant perceptions within the collaborative SG design process. Findings highlight AI as a beneficial supplement in the role-playing SG design activity, contributing balanced suggestions without significantly affecting overall satisfaction compared to a human-played teacher role. While AI helped effectively identify conflicting design suggestions, its tendency to propose existing design suggestions over generating novel ideas and its systematic intervention style indicated potential areas for improvement, particularly in maintaining creative flow.

The study emphasises the need for AI's strategic timing and contributions to align with human team dynamics, suggesting a pathway for enhancing AI's creative capabilities in future SG design processes. The research lays the groundwork for advanced exploration of human-AI

collaboration in developing SGs as tools for studying decision-making related to privacy or other important domains. It illuminates the dynamic role of AI in enriching the processes of collaborative SG design, emphasising the significant potential of AI-enhanced collaboration to facilitate deep and engaging examinations of complex decision-making scenarios.

Acknowledgements

The research has been partially funded by the Austrian Research Promotion Agency FFG through the COIN Project SimLern, #898080.

References

- [1] P. Jost, "The Quest Game-Frame: Balancing Serious Games for Investigating Privacy Decisions," 11th Scandinavian Conference on Information Systems, Jan. 2020, [Online]. Available: <https://aisel.aisnet.org/scis2020/5>.
- [2] R. Dörner, S. Göbel, W. Effelsberg, and J. Wiemeyer, Eds., *Serious Games: Foundations, Concepts and Practice*. Springer, 2016, doi: 10.1007/978-3-319-40612-1.
- [3] G. Klein, *Sources of Power: How People Make Decisions*, 20th Anniversary Edition. Cambridge, MA, USA: MIT Press, 2017, doi: 10.7551/mitpress/11307.001.0001.
- [4] G. Campitelli and F. Gobet, "Herbert Simon's decision-making approach: Investigation of cognitive processes in experts," *Review of general psychology*, vol. 14, no. 4, pp. 354–364, 2010, doi: 10.1037/a0021256.
- [5] G. Klein, "Expert intuition and naturalistic decision making," in *Handbook of intuition research*, Edward Elgar Publishing, 2011, doi: 10.4337/9780857936370.00014.
- [6] P. Jost and M. Divitini, "Game elicitation: exploring assistance in delayed-effect supply chain decision making," 2020, p. 10, doi: 10.1145/3419249.3420154.
- [7] D. Kahneman and G. Klein, "Conditions for intuitive expertise: A failure to disagree.," *American Psychologist*, vol. 64, no. 6, pp. 515–526, 2009, doi: 10.1037/a0016755.
- [8] J. Nakamura and M. Csikszentmihalyi, "The Experience of Flow - Theory and Research," in *The Oxford Handbook of Positive Psychology*, Third Edition., New York: Oxford University Press, 2020, pp. 279–296, doi: 10.1093/oxfordhb/9780199396511.013.16.
- [9] P. Jost, S. Rangger, and A. Künz, "Tap or Swipe: Interaction's Impact on Cognitive Load and Rewards in a Mobile Mental Math Game," in *Proceedings of the 20th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2023)*, D. G. Sampson, D. Ifenthaler, and P. Isaías, Eds., Lisbon, Portugal: IADIS Press, 2023, pp. 153–162.
- [10] P. Jost, S. Cobb, and I. Hämmerle, "Reality-based interaction affecting mental workload in virtual reality mental arithmetic training," *Behaviour & Information Technology*, vol. 39, no. 10, pp. 1062–1078, 2020, doi: 10.1080/0144929X.2019.1641228.
- [11] P. Jost and M. Divitini, "The Challenge Game Frame: Affordance oriented co-creation of Privacy Decision Games," in *14th International Conference on Game Based Learning, ECGBL 2020*, September 24, 2020 - September 25, 2020, in *Proceedings of the 14th International Conference on Game Based Learning, ECGBL 2020*. Brighton, United Kingdom: Academic Conferences International, 2020, pp. 277–286, doi: 10.34190/GBL.20.134.
- [12] P. Jost and A. Künz, "Cards and Roles: Co-designing Privacy Serious Games with an Online Role-Playing Boardgame," in *Games and Learning Alliance*, F. de Rosa, I. Marfisi Schottman, J. Baalsrud Hauge, F. Bellotti, P. Dondio, and M. Romero, Eds., in *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2021, pp. 187–197, doi: 10.1007/978-3-030-92182-8_18.
- [13] P. Jost and M. Divitini, "From Paper to Online: Digitizing Card Based Co-creation of Games for Privacy Education," in *Technology-Enhanced Learning for a Free, Safe, and Sustainable World*, T. De Laet, R. Klemke, C. Alario-Hoyos, I. Hilliger, and A. Ortega-Arranz, Eds., in *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2021, pp. 178–192, doi: 10.1007/978-3-030-86436-1_14.

- [14] C. Bösch, B. Erb, F. Kargl, H. Kopp, and S. Pfattheicher, "Tales from the dark side: Privacy dark strategies and privacy dark patterns," *Proceedings on Privacy Enhancing Technologies*, vol. 2016, no. 4, pp. 237–254, 2016, doi: 10.1515/popets-2016-0038.
- [15] M. Becker, S. M. Klausning, and T. Hess, "Uncovering the privacy paradox: The influence of distraction on data disclosure decisions," 2019.
- [16] N. Gerber, P. Gerber, and M. Volkamer, "Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior," *Computers & Security*, vol. 77, pp. 226–261, 2018, doi: 10.1016/j.cose.2018.04.002.
- [17] A. E. Waldman, "Cognitive biases, dark patterns, and the 'privacy paradox,'" *Current Opinion in Psychology*, vol. 31, pp. 105–109, Feb. 2020, doi: 10.1016/j.copsyc.2019.08.025.
- [18] D. Kahneman, *Thinking, fast and slow*. 2011. Accessed: Oct. 23, 2021. [Online]. Available: http://ebook.3m.com/library/neworleanspubliclibrary-document_id-dpt1489
- [19] D. Kahneman, P. Slovic, and A. Tversky, Eds., *Judgment under uncertainty: heuristics and biases*. Cambridge ; New York: Cambridge University Press, 1982, doi: 10.1017/CBO9780511809477.
- [20] G. Klein, *Snapshots of the Mind*. The MIT Press, 2022, doi: 10.7551/mitpress/14342.001.0001.
- [21] N. Li, Y. Yan, Y. Yang, and A. Gu, "Artificial Intelligence Capability and Organizational Creativity: The Role of Knowledge Sharing and Organizational Cohesion," *Frontiers in Psychology*, vol. 13, 2022, Accessed: Nov. 12, 2023. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.845277>, doi: 10.3389/fpsyg.2022.845277.
- [22] G. Zhang, A. Raina, J. Cagan, and C. McComb, "A cautionary tale about the impact of AI on human design teams," *Design Studies*, vol. 72, p. 100990, Jan. 2021, doi: 10.1016/j.destud.2021.100990.
- [23] J. Sweller, P. Ayres, and S. Kalyuga, *Cognitive Load Theory*. New York, NY: Springer New York, 2011. Accessed: Oct. 06, 2016. [Online]. Available: <http://link.springer.com/10.1007/978-1-4419-8126-4>, doi: 10.1007/978-1-4419-8126-4.
- [24] M. Sharples, "Towards social generative AI for education: theory, practices and ethics," no. arXiv:2306.10063. arXiv, Jun. 14, 2023, doi: 10.48550/arXiv.2306.10063.
- [25] T. Tang et al., "PlotThread: Creating Expressive Storyline Visualisations using Reinforcement Learning," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 2, pp. 294–303, Feb. 2021, doi: 10.1109/TVCG.2020.3030467.
- [26] A. N. Vazquez, "Evaluating Team Fluency in Human-Industrial Robot Collaborative Design Tasks," in *Computer-Aided Architectural Design. Design Imperatives: The Future is Now*, D. Gerber, E. Pantazis, B. Bogosian, A. Nahmad, and C. Miltiadis, Eds., in *Communications in Computer and Information Science*. Singapore: Springer, 2022, pp. 378–402, doi: 10.1007/978-981-19-1280-1_24.
- [27] A. Sarkar, "Exploring Perspectives on the Impact of Artificial Intelligence on the Creativity of Knowledge Work: Beyond Mechanised Plagiarism and Stochastic Parrots," in *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work, in CHIWORK '23*. New York, NY, USA: Association for Computing Machinery, Sep. 2023, pp. 1–17, doi: 10.1145/3596671.3597650.
- [28] W. Wang, "Application Research of Computer Artificial Intelligence Technology in Digital Environment Integrated Design System," in *2023 IEEE International Conference on Sensors, Electronics and Computer Engineering (ICSECE)*, Aug. 2023, pp. 363–367, doi: 10.1109/ICSECE58870.2023.10263527.
- [29] I. Marfisi-Schottman, P.-Y. Gicquel, and S. George, "Meta Serious Game: Supporting Creativity Sessions for Mobile Serious Games," in *Proceedings of the 10th International Conference on Game Based Learning (ECGBL 2016)*, Reading, UK: Academic Conferences International Limited, 2016, pp. 407–415.
- [30] J. J. Gibson, *The Ecological Approach to Visual Perception*. New York, N.Y.: Psychology Press, 2015.
- [31] O. Volkoff and D. M. Strong, "Affordance theory and how to use it in IS research," in *The Routledge Companion to Management Information Systems*, 1st ed., R. D. Galliers and M.-K. Stein, Eds., Abingdon, Oxon ; New York, NY : Routledge, 2017.: Routledge, 2017, pp. 232–245, doi: 10.4324/9781315619361-18.
- [32] A. R. Hevner, "A three cycle view of design science research," *Scandinavian journal of information systems*, vol. 19, no. 2, p. 4, 2007.

- [33] G. Costikyan, *Uncertainty in Games*. MIT Press, 2015.
- [34] G. Fischer, "Social creativity: turning barriers into opportunities for collaborative design," in *Proceedings of the eighth conference on Participatory design: Artful integration: interweaving media, materials and practices-Volume 1*, 2004, pp. 152–161, doi: 10.1145/1011870.1011889.
- [35] E. Brandt and J. Messeter, "Facilitating collaboration through design games," in *Proceedings of the eighth conference on Participatory design: Artful integration: interweaving media, materials and practices - Volume 1*, in PDC 04. New York: ACM, Jul. 2004, pp. 121–131, doi: 10.1145/1011870.1011885.
- [36] R. A. Finke, T. B. Ward, and S. M. Smith, *Creative cognition: Theory, research, and applications*. in *Creative cognition: Theory, research, and applications*. Cambridge, MA, US: The MIT Press, 1992, pp. vi, 239.
- [37] Y. Maaravi, B. Heller, Y. Shoham, S. Mohar, and B. Deutsch, "Ideation in the digital age: literature review and integrative model for electronic brainstorming," *Review of Managerial Science*, pp. 1–34, 2020, doi: 10.1007/s11846-020-00400-5.
- [38] A. Vaswani et al., "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [39] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.
- [40] T. Brown et al., "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
- [41] R. W. Picard, *Affective Computing*. MIT Press, 2000, doi: 10.7551/mitpress/1140.001.0001.
- [42] H. A. Simon, *The Sciences of the Artificial*, 3. ed., [Nachdr.]. Cambridge, Mass.: MIT Press, 1996.
- [43] F. V. Gianni and M. Divitini, "Designing IoT applications for smart cities: extending the tiles ideation toolkit," 2018, doi: 10.55612/s-5002-035-005.
- [44] A. Field, *Discovering statistics using IBM SPSS statistics*, 5th edition. Thousand Oaks, CA: SAGE Publications, 2017.
- [45] A. Hart, "Mann-Whitney test is not just a test of medians: differences in spread can be important," *Bmj*, vol. 323, no. 7309, pp. 391–393, 2001, doi: 10.1136/bmj.323.7309.391.