LLM-Driven Scenario Generation for Serious Game Moderation: A Process Model Approach

Julie Klingelschmitt[1], Bettina Schneider[1][0000-0001-8460-3658], Hermann Grieder[1][0000-0001-9984-8615], and Frank Grimberg[1]

School of Business, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Basel, Switzerland
julie.klingelschmitt@students.fhnw.ch,
{bettina.schneider,hermann.grieder,frank.grimberg}@fhnw.ch

Abstract. Serious games (SG) have gained increasing attention in education due to their capacity to enhance learner engagement. Recent advances in Artificial Intelligence (AI), particularly Large Language Models (LLMs), offer new opportunities to support SG design and moderation through content generation. However, procedural guidance for educators on how to reliably integrate these models remains limited, and LLMs introduce challenges such as hallucinations, factual inaccuracies, and bias, highlighting the need for structured frameworks. This study investigates how LLMs can be systematically employed to generate playable scenarios that support SG moderation. Following a Design Science Research (DSR) approach, a business process model (BPM) artifact was developed and evaluated iteratively through controlled experiments and expert interviews. The findings suggest that LLMs can generate coherent and playable scenarios when guided by a structured process model. Expert feedback emphasized limitations related to creativity control and the need for empirical validation in classroom settings. This study provides practical guidance for educators and contributes to a broader understanding of AI integration in SG.

Keywords: Serious Games · Artificial Intelligence · Large Language Models · Serious Game Moderation · Scenario Generation · Design Science Research

1 Introduction

1.1 Background

In recent years, the integration of digital technologies into education has significantly reshaped teaching and learning [1]. Among these, Artificial Intelligence (AI) has emerged as a transformative force enabling more interactive and personalized learning experiences [2]. Large Language Models (LLMs), an instantiation of generative AI, have gained particular attention in education due to their potential to support educators through content generation, lesson planning, and personalized feedback [3].

Beyond conventional educational applications, LLMs can be leveraged in more interactive didactical formats. One such format are Serious Games (SGs), which combine gameplay with educational objectives to foster engagement and learning [4,5]. SGs may take digital forms, such as virtual world, computer, or web-based games, and non-digital forms, such as board games or LEGO-based activities [6]. They are applied in diverse domains including healthcare, defense, politics, education, and culture [7]. The usefulness of SGs depends substantially on effective moderation, defined as the facilitation of

gameplay through new challenges or game activities [8]. In this research, SG moderation is conceptualized as the creation of playable scenarios that define game rules and challenges. The board-based 3LandSpiel serves as use case.

1.2 Problem Statement

Even though the integration of AI into SG design and moderation is a promising emerging field, generative AI and LLMs exhibit notable limitations, including bias, factual inaccuracies, and hallucinations [9,10,11]. Hallucinations refer to instances where the AI produces information that is fabricated or inconsistent with the provided input [12]. Humble [11] argues that AI tools often require multiple attempts to generate suitable output and that ethical and legal concerns remain associated with their use. Colado et al. [9] confirm that despite its potential, generative AI can produce inconsistencies in output accuracy, raising concerns about the reliability and educational integrity of generated content.

Prompt engineering, defined as the design and refinement of prompts to steer LLM responses towards accuracy, relevance, and usability, therefore plays a significant role [13]. Chen et al. [13] demonstrate that the precision and format of a prompt directly impact the quality of the generated response. Humble [11] clarifies that AI tools perform more effectively when used iteratively with refined prompts and richer context. Capecchi et al. [14] further stress that well-crafted prompts are determinant for meeting the educational goals of a SG. However, while these studies acknowledge the importance of structured prompting, none provide a reusable procedural framework tailored to educators without technical expertise. This gap leads to the following research question:

How can LLMs systematically generate playable scenarios to support the moderation of board serious games by means of structured guidance?

This study contributes to the development of structured guidelines for the use of LLMs in SG moderation. The focus is set on generating playable scenarios, which are understood as written descriptions outlining how a SG can be played to achieve a specific learning objective, specifying rules, player roles, and material use. The research explicitly aims for systematic scenario generation in the sense that the process follows a controlled and reproducible method.

2 Background and Related Work

2.1 Literature Review Search Strategy

A systematic literature review was conducted to examine the current state of research on the use of AI in SGs. Relevant studies were identified through searches across four academic databases: Web of Science, Scopus, IEEE Xplore, and ERIC. Search strings combining terms such as “artificial intelligence”, “serious games”, “educational games”, and “game-based learning” were applied. The review focused on English-language studies published from 2023 onward to capture applications of AI following the emergence of ChatGPT. Following predefined inclusion and exclusion criteria, 13 publications were selected and synthesized.

2.2 Serious Game Design

Generative AI and LLMs have been employed in various aspects of SG design, acting as co-creative partners to support and enrich the creative process. Studies show that AI

can assist SG design by taking on the role of a teacher or providing technical support for content generation [9,10,15].

Colado et al. [9] and Jost [15] describe how LLMs can automatically generate feedback and character dialogues. Gatti Junior et al. [10] explore the use of AI to create dynamic scenarios that respond to player choices. Beyond ideation, Colado et al. [9] demonstrate that generative AI can reduce manual work in early game design by automating tasks such as writing narratives and creating visual assets. Humble [11] confirms these findings, noting that AI can also generate graphics and in-game audio. Across these works, AI is framed as a creative partner capable of enriching design processes, though its effectiveness depends on thoughtful prompting [9,10,15].

2.3 Serious Game Moderation

In addition to SG design, LLMs have also been leveraged to adapt and moderate gameplay. This includes providing real-time, personalized feedback to players [16], adjusting difficulty levels based on performance [14], and generating new in-game challenges to maintain motivation and flow [8,17]. Collectively, these approaches demonstrate how generative AI contributes to more responsive, immersive, and effective learning experiences within SGs. Nevertheless, integrating AI into SGs remains challenging due to the inherent complexity of both domains.

The selected studies converge on the view that the role of AI in SG design and moderation is one of augmenting human work rather than replacing it, and that ensuring consistent outputs remains difficult due to bias and hallucinations [9,10,11]. AI can accelerate ideation, generate diverse content, and personalize interactions, but only when embedded in a structured framework ensuring consistency, playability, and alignment with learning objectives. However, no study currently offers a repeatable and validated framework for using AI to create scenarios for SG moderation.

3 Methodology

This study followed a Design Science Research (DSR) methodology, suitable for addressing practical problems through the creation and evaluation of artifacts [18]. A pragmatic research philosophy and abductive approach were adopted, combining theoretical insights with empirical evaluation. The research was conducted in two iterative DSR cycles.

The first cycle, the *Formative Design* cycle, translated requirements from the systematic literature review into an initial process model artifact. To identify the most effective prompt structure, 15 controlled experiments were conducted using ChatGPT and the 3LandSpiel as a use case. Each experiment consisted of a prompt formulated by the researcher, varying in prompting technique (zero-shot, few-shot, or Chain-of-Thought) and input composition. The resulting scenarios were qualitatively evaluated by the researcher against predefined criteria: coherence, consistency, and structural clarity. Findings were consolidated into a generalizable prompt template integrated into the artifact.

The second cycle, the *Refinement and Validation* cycle, refined the artifact based on experimental findings and assessed it through five semi-structured expert interviews with specialists from education and computer science. Prior to each interview, experts independently applied the process model using the 3LandSpiel as a reference case: they filled in the prompt template, executed it in ChatGPT, and evaluated the generated scenario. During the subsequent one-to-one interviews, experts were asked to reflect on this hands-on experience along two dimensions: usability, assessing how clearly and intuitively they

could follow the process model, and usefulness, assessing whether the generated scenario was of sufficient pedagogical quality. Semi-structured interviews were chosen over surveys to allow follow-up questions and capture nuanced feedback on the artifact's practical applicability [19].

4 Results

4.1 Use Case: The 3LandSpiel

The use case employed in this study was the 3LandSpiel, a board-based SG developed to address cross-border and sustainability-related challenges in the Upper Rhine region encompassing France, Germany, and Switzerland. The game fosters intercultural dialogue and raises awareness of sustainability issues in classroom environments. It emerged from interdisciplinary student discussions during the 2024 Spring School and was developed in collaboration with FHNW School of Business, Regio Basiliensis, and the University of Basel [24,25]. Players take on fictional stakeholder roles on a physical board representing the tri-national region, negotiating and resolving sustainability challenges collaboratively.

4.2 First Design Cycle

Process Model Approach. To answer the main research question, it was necessary to identify an approach that brings structure to the use of LLMs, enabling reliable and reproducible outputs. Process modeling offered such an approach, as it provides a formal structure to guide and organize work execution [20]. According to Van Der Aalst [21], process models offer a structured representation of system behavior supporting transparency, monitoring, and iterative improvement.

A structured process model was therefore assumed to facilitate the systematic use of LLMs to generate playable scenarios. By following a standardized process model, educators could apply a reproducible procedure to scenario generation, even when the game context differs. While the non-deterministic nature of LLMs means that outputs may vary across runs, the process model ensures that the steps, inputs, and evaluation criteria remain consistent and transferable.

Prompting Techniques. As this study focused on enabling educators to use LLMs without requiring technical expertise, only intuitive prompting techniques were considered. Three were selected: zero-shot prompting (a well-defined task description without examples), few-shot prompting (a small number of examples to guide output), and Chain-of-Thought prompting (step-by-step reasoning to improve coherence) [22].

Input and Output Modalities. Inputs were inspired by the SG characteristics of Ben Amara et al. [23], simplified to reduce complexity. The instruction component comprised the LLM role, a description of the board SG and its purpose, and the generation task. The context component comprised the physical board and materials, the target audience, and the learning objective.

Expected LLM outputs comprised: the gameplay (scenario context, storyline, rules, material usage, winning/losing conditions); stakeholders and roles (fictional player roles); interaction type (collaborative, competitive, role-play, or problem-solving); and structure and duration (time schedule with phases and rounds).

Design Requirements. Ten design requirements were derived from the identified prompting techniques, defined inputs, and expected outputs. They guided artifact development and reflect expected behaviors of the process model, grouped into four categories: Structural Clarity, Core Functional Capabilities, Adaptability and Iteration, and Usability (Table 1).

Table 1. Design Requirements for the Artifact ("The process model shall ...").

Category	ID	Requirement	Evaluated in
Structural Clarity	R-1	clearly specify the steps guiding educators in generating playable scenarios with an LLM.	Experiment & Interview
	R-2	define the necessary inputs educators need to provide.	Experiment
	R-3	clearly distinguish the roles and responsibilities of the educator and the LLM.	Interview
Core Functional Capabilities	R-4	guide the generation of scenarios that are structured, coherent, and consistent.	Experiment
	R-5	enable educators to systematically use LLMs to generate scenarios.	Experiment
	R-6	help ensure that scenarios are playable and align with the learning objectives.	Interview
Adaptability and Iteration	R-7	allow iterative refinement of the LLM output, enabling educators to adjust scenarios.	Interview
	R-8	be applicable and generalizable for every type of board SG.	Interview
Usability	R-9	be easy to understand and usable by educators.	Interview
	R-10	present information and steps in a visually clear and organized manner.	Interview

Experiments. To identify the most effective prompt content and prompting technique, 15 experiments were conducted. The results indicated that zero-shot prompting produced the most effective outputs, as it consistently yielded structured and coherent scenarios while also being the easiest and most intuitive technique for educators to apply. The insights gained from this experimental phase were consolidated into a generalizable prompt template (see Fig. 1).

The tentative process model developed during the first cycle incorporated two separate swimlanes for the LLM and the educator. The educator was required first to define the instruction, then to specify the context, and finally to write the prompt according to the chosen prompting technique. Since the optimal prompt structure could only be determined through experimentation, the resulting artifact was designed as a high-level process model. It outlines the main steps required for scenario generation, specifies the SG characteristics the educator must provide, defines the expected output, and integrates an iterative loop for prompt refinement.

Instruction: You are an expert in generating playable scenarios for the moderation of a board serious game. Please generate a playable scenario for this board game according to the following information.

Context

Serious game name and purpose: [description of the game and its purpose]
Description of the board and game material: [description of the board, game pieces, and other physical elements]
Learning objective: [thematic topic to learn and discuss about]
Target audience: [description of the target audience and number of participants]

The **scenario** should be structured like this:
Scenario Overview – Duration – Game set-up – Roles and grouping – Gameplay phases – Learning outcomes – Facilitation tips – Debrief

Fig. 1. Prompt Template.

4.3 Second Design Cycle

Artifact Refinement. The refined artifact represents a structured and iterative procedure guiding educators from defining the game context to validating a final scenario for classroom implementation. Key refinements included integrating the standardized prompt template as a data object within the process model, embedding generic elements such as the instruction and expected scenario structure directly into the template. As a result, educators need only provide context-specific information: the game description, available materials, target audience, and learning objectives.

The refined model further improves usability by enabling educators to complete the prompt template through a fill-in approach and execute it directly within an LLM interface. As prompt refinement is highly context-dependent, defining a fixed interaction flow was neither realistic nor feasible. Instead, reflective evaluation questions were incorporated to guide educators in assessing and iteratively refining generated scenarios, addressing dimensions such as coherence, playability, alignment with learning objectives, the presence of potential hallucinations, and appropriateness of complexity and duration.

Expert Interview Results. The evaluation phase assessed the usability and usefulness of the refined process model through semi-structured expert interviews, as these dimensions are central for evaluating artifact effectiveness in DSR [26,27].

Usability focuses on the ease with which users can understand and interact with an artifact [27]. Experts consistently described the process model as clear, structured, and easy to follow. The step-by-step guidance and integrated prompt template were perceived as particularly helpful. The reflective evaluation questions further enhanced usability by providing concrete criteria for assessing scenario quality. However, some experts noted that BPMN notation may not be intuitive for all educators and suggested a more didactic representation for school contexts.

Usefulness evaluates the practical value of the artifact [26]. Experts agreed that the process model successfully supported the generation of coherent scenarios aligned with the chosen learning objective, offering new ideas for gameplay. Nevertheless, experts emphasized that generated scenarios were not immediately playable and required iterative refinement. Aspects such as time management, realism of game mechanics, and effective use of game materials could not be fully optimized by the LLM alone.

Final Refinement. Usability was further improved by adding more explicit guidance on how to evaluate generated outputs, as experts had criticized the difficulty of assessing scenario quality. To this end, a scenario quality evaluation checklist was developed and included within the process model as a data object.[1] A playtest loop was added to emphasize that scenarios should be tested in practice before final validation. The final artifact is presented as Fig. 2.

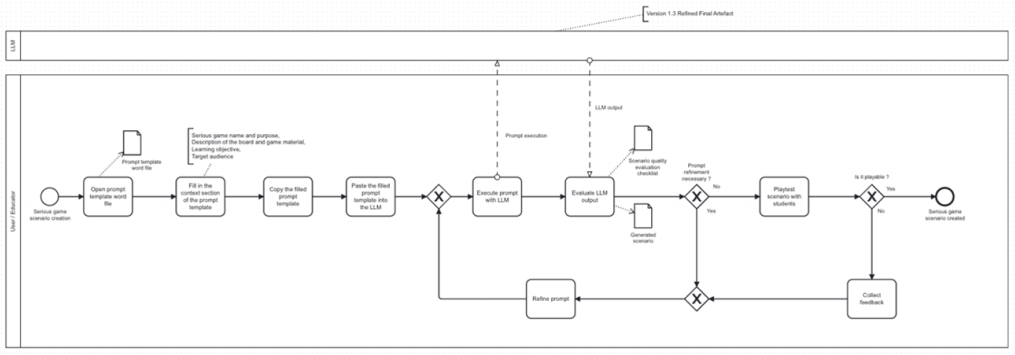


Fig. 2. Final Artifact: Process Model for LLM-Supported SG Scenario Generation.

The usefulness of the model was assessed as good, although it was regarded as still under development. Experts saw it as a promising starting point that fosters creativity in scenario creation.

Review of Design Requirements. The evaluation indicated that the artifact fulfilled the majority of design requirements. Eight of the ten were fully met: the process model provided clear structural guidance, defined necessary inputs, and enabled reproducible generation of coherent and consistent scenarios, supported by the iterative refinement loop and evaluation checklist.

Two requirements were partially met. R-6, concerning the immediate playability of generated scenarios, was only partially satisfied, as scenarios required further iterative refinement and classroom testing before they could be considered fully playable. R-9, concerning ease of use for all educators, was partially met due to concerns about the accessibility of the BPMN notation for non-technical users.

5 Discussion

The results demonstrate that LLMs can generate coherent and structurally sound SG scenarios when guided by a structured process model. This aligns with prior research positioning generative AI as a creative partner in educational design [9,10,15]. The process model operationalized the use of LLMs for SG moderation, translating abstract prompting principles into a concrete, reproducible workflow for educators.

[1] This checklist includes dimensions such as the validity, clarity, pedagogical fit, and engagement value of the scenario. It suggests possible refinement actions for each category. The process model and checklist are publicly available at https://drive.switch.ch/index.php/s/ZVO87lMI4MZFMSs.

The experimental results indicate that zero-shot prompting, when supported by a well-structured input template, is sufficient to produce coherent and pedagogically aligned scenarios. This is consistent with Chen et al. [13] and Humble [11], who argue that prompt precision and richer context improve LLM performance. The restriction to user-friendly prompting techniques reflects the practical reality that educators typically lack advanced technical skills [9,14].

However, the finding that generated scenarios were not immediately playable highlights a fundamental limitation: LLMs lack grounded knowledge of specific physical game components and the experiential understanding of real classroom dynamics. This extends the discussion of Gatti Junior et al. [10] and Bonetti et al. [16], who acknowledge the difficulty of generating contextually grounded game content through AI alone. Achieving playability requires nuanced pedagogical judgment, which remains a distinctly human competence, and BPMN accessibility for non-technical users is an open challenge in process modeling adoption. Both suggest that integrating LLMs into SG moderation is best conceptualized as a human-AI collaborative process rather than a fully automated pipeline.

From a theoretical perspective, this research contributes to the literature on AI-supported SGs by demonstrating that a process model artifact can serve as a practical instrument for managing LLM limitations, including inconsistent outputs and hallucinations. It further contributes to the DSR literature by providing a concrete example of how iterative artifact development and multi-method evaluation can yield practically useful and theoretically grounded results.

6 Conclusion and Future Work

This paper investigated how LLMs can be systematically employed to support SG moderation through scenario generation. A structured process model was proposed and evaluated through two iterative DSR cycles combining controlled experiments and expert evaluation, addressing the absence of procedural guidance for educators.

The findings show that embedding LLM usage within a structured process supports coherence, consistency, and pedagogical alignment of game scenarios. Simple prompting techniques, particularly zero-shot prompting, can be effective when supported by clear input specifications and evaluation guidance. At the same time, generated scenarios consistently required iterative refinement before classroom implementation, underscoring the importance of human judgment.

The study has several limitations. The evaluation relied on five expert interviews rather than empirical classroom testing. The artifact was validated using a specific board-based SG, which may limit generalizability. Additionally, the study focused on user-friendly prompting techniques, excluding more advanced approaches that may further enhance output quality.

Future research should focus on empirical classroom studies to assess educational value and playability in real learning environments. Further work could explore adapting the process model to digital or hybrid SG formats, integrating more advanced prompt engineering strategies, and developing simplified visual representations without BPMN notation. Research could also investigate how educators can better articulate their creative expectations, and how far LLMs can transform the game experience before creativity begins to undermine learning objectives and playability.

References

1. Zhang, K., Aslan, A.B.: AI technologies for education: Recent research & future directions. Computers and Education: Artificial Intelligence. **2**, 100025 (2021). https://doi.org/10.1016/j.caeai.2021.100025
2. Alqahtani, T., Badreldin, H.A., Alrashed, M., et al.: The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. Research in Social and Administrative Pharmacy. **19**, 1236–1242 (2023). https://doi.org/10.1016/j.sapharm.2023.05.016
3. Bektik, D., Edwards, C., Whitelock, D., Antonaci, A.: Use of LLM tools within higher education: Report 1 (2024). https://doi.org/10.5281/ZENODO.14501259
4. Berg Jørgensen, N., Skovbjerg, H.M.: Concepts of educational design for serious games (2019). https://vbn.aau.dk/en/publications/concepts-of-educational-design-for-serious-games
5. Laamarti, F., Eid, M., El Saddik, A.: An overview of serious games. International Journal of Computer Games Technology. **2014**, 1–15 (2014). https://doi.org/10.1155/2014/358152
6. Asprion, P.M., Schneider, B., Moriggl, P., Grimberg, F.: Exploring cyber security awareness through LEGO Serious Play Part I: The Learning Experience. In: Expanding Horizons: Business, Management and Technology for Better Society. Proceedings of the MakeLearn and TIIM International Conference, 20–22 May 2020, Online Conference, pp. 263–272. ToKnowPress, ISBN 978-961-6914-26-0 (2020)
7. Din, S.U., Baig, M.Z., Khan, M.K.: Serious games: an updated systematic literature review (2023). https://doi.org/10.48550/ARXIV.2306.03098
8. Khosrawi-Rad, B., Hoang, G.-H., Schlimbach, R., Robra-Bissantz, S.: Design of a pedagogical conversational agent as moderator of an educational serious game for business model development. In: Proceedings of Workshop GeNeMe'23, pp. 246–256. TUDpress, Dresden (2023). https://doi.org/10.25368/2024.290
9. Colado, I.J.P., Colado, V.M.P., Morata, A.C., Píriz, R.S.C., Manjón, B.F.: Using new AI-driven techniques to ease serious games authoring. In: 2023 IEEE Frontiers in Education Conference (FIE), pp. 1–9. IEEE (2023). https://doi.org/10.1109/FIE58773.2023.10343021
10. Gatti Junior, W., Marasco, E., Kim, B., Behjat, L., Eggermont, M.: How ChatGPT can inspire and improve serious board game design. IJSG. **10**, 33–54 (2023). https://doi.org/10.17083/ijsg.v10i4.645
11. Humble, N.: Play my thesis: opportunities and limitations of building serious games with AI. ICAIR. **4**, 165–174 (2024). https://doi.org/10.34190/icair.4.1.3130
12. Maleki, N., Padmanabhan, B., Dutta, K.: AI hallucinations: a misnomer worth clarifying. In: 2024 IEEE Conference on Artificial Intelligence (CAI), pp. 133–138. IEEE (2024). https://doi.org/10.1109/CAI59869.2024.00033
13. Chen, B., Zhang, Z., Langréné, N., Zhu, S.: Unleashing the potential of prompt engineering for large language models. Patterns. **6**, 101260 (2025). https://doi.org/10.1016/j.patter.2025.101260
14. Capecchi, I., Borghini, T., Bellotti, M., Bernetti, I.: Enhancing education outcomes integrating augmented reality and artificial intelligence for education in nutrition and food sustainability. Sustainability. **17**, 2113 (2025). https://doi.org/10.3390/su17052113
15. Jost, P.: AI-deation: when the teacher is a transformer in role-playing to create privacy decision serious games. IJSG. **11**, 65–86 (2024). https://doi.org/10.17083/ijsg.v11i2.762
16. Bonetti, F., Bucchiarone, A., Wanick, V.: Using LLMs to adapt serious games with educators in the loop. In: Games and Learning Alliance, pp. 68–77. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-78269-5_7
17. Zhao, J., Jingru, Z., Lu, Y.: Enhancing design historical education through AI virtual characters role-playing narratives in serious games. IJGCMS. **17**, 1–20 (2025). https://doi.org/10.4018/IJGCMS.372681
18. Kuechler, B., Vaishnavi, V.: On theory development in design science research: anatomy of a research project. European Journal of Information Systems. **17**, 489–504 (2008). https://doi.org/10.1057/ejis.2008.40

19. Venable, J., Pries-Heje, J., Baskerville, R.: FEDS: a framework for evaluation in design science research. European Journal of Information Systems. **25**, 77–89 (2016). https://doi.org/10.1057/ejis.2014.36
20. Polyvyanyy, A., Smirnov, S., Weske, M.: Business process model abstraction. In: Vom Brocke, J., Rosemann, M. (eds.) Handbook on Business Process Management 1, pp. 147–165. Springer Berlin Heidelberg (2015). https://doi.org/10.1007/978-3-642-45100-3_7
21. Van Der Aalst, W.M.P.: Business process management: a comprehensive survey. ISRN Software Engineering. **2013**, 1–37 (2013). https://doi.org/10.1155/2013/507984
22. Sahoo, P., Singh, A.K., Saha, S., Jain, V., Mondal, S., Chadha, A.: A systematic survey of prompt engineering in large language models: techniques and applications (2024). https://doi.org/10.48550/ARXIV.2402.07927
23. Ben Amara, B., Mhiri Sallami, H., Ben Said, L.: The principal characteristics of a serious game to ensure its effective design. In: Proceedings of DiGRA 2022 Conference (2022). https://doi.org/10.26503/dl.v2022i1.1365
24. FHNW Hochschule für Wirtschaft: Im Spiel die Region Oberrhein weiter entwickeln. https://www.fhnw.ch/de/die-fhnw/hochschulen/hsw/media-newsroom/news/im-spiel-die-region-oberrhein-weiter-entwickeln
25. Regio Basiliensis, FHNW, Universität Basel: 3LandSpiel Projekt-Pitch_D+F (2024)
26. Gill, T.G., Hevner, A.R.: A fitness-utility model for design science research. In: Jain, H., Sinha, A.P., Vitharana, P. (eds.) Service-Oriented Perspectives in Design Science Research, pp. 237–252. Springer Berlin Heidelberg (2011). https://doi.org/10.1007/978-3-642-20633-7_17
27. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Quarterly. **28**, 75–106 (2004). https://doi.org/10.2307/25148625