






The Chronicles of ChatGPT: Generating and Evaluating Visual Novel Narratives on Climate Change Through ChatGPT

Mustafa Can Gursesli¹ , Pittawat Taveekitworachai² , Febri Abdullah², Mury F. Dewantoro², Antonio Lanata¹, Andrea Guazzini³, Van Khôi Lê⁴, Adrien Villars⁴, and Ruck Thawonmas⁵ 

- ¹ Department of Information Engineering, Università degli Studi di Firenze, Florence, Italy
`{mustafacan.gursesli,antonio.lanata}@unifi.it`
- ² Graduate School of Information Science and Engineering, Ritsumeikan University, Kusatsu, Shiga, Japan
`{gr0609fv,gr0634hi,gr0397fs,gr0450xi}@ed.ritsumei.ac.jp`
- ³ Department of Education, Literatures, Intercultural Studies, Languages and Psychology, Università degli Studi di Firenze, Florence, Italy
`andrea.guazzini@unifi.it`
- ⁴ Graduate School of Engineering, ENSEIRB-MATMECA, Talence, France
`{van_khoi.le,avillars}@bordeaux-inp.fr`
- ⁵ College of Information Science and Engineering, Ritsumeikan University, Kusatsu, Shiga, Japan
`ruck@is.ritsumei.ac.jp`

Abstract. This paper explores the potential of utilizing ChatGPT, a large language model (LLM), for generating and evaluating visual novel (VN) game stories in the context of global warming awareness through a VN game. The study involves generating two stories using ChatGPT, one with given global warming related keywords as an inspiration for ChatGPT along with a specified ending and another without, and evaluating them based on several linguistic criteria: coherence, inspiration, readability, word complexity, and narrative fluency. Results reveal that keywords-inspired story exhibit higher coherence, while the basic one demonstrate greater inspiration. The findings highlight the advantages of each story and emphasize the value of AI-driven narrative generation in creating engaging and informative experiences. Furthermore, the study introduces an innovative approach by employing ChatGPT as an evaluator for the story quality, by combining various prompt engineering techniques showcasing the diverse applications of LLMs in interactive storytelling. This work contributes to the growing field of LLM-based story generation and underscores the potential of AI-driven narratives in fostering awareness and engagement on critical issues like climate change.

Keywords: Global warming · climate change · visual novel · ChatGPT · narrative generation

M. C. Gursesli and P. Taveekitworachai—These authors contributed equally.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
L. Holloway-Attaway and J. T. Murray (Eds.): ICIDS 2023, LNCS 14384, pp. 181–194, 2023.
https://doi.org/10.1007/978-3-031-47658-7_16

1 Introduction

Climate change is defined by United Nations as “long-term shifts in temperatures and weather patterns” [30]. While to some degree variations in atmospheric conditions occur in nature [4, 51], human activities have managed to exacerbate this phenomenon to critical levels, being responsible for 1.1C° of global warming since 1800 [11]. Specifically, the human footprint in climate change can be found in the widespread consumption of fossil fuels that lead to an increase in greenhouse gases present in the atmosphere, which prevents solar heat from dispersing, thus increasing temperatures [18, 30]. Several measures have been taken to tackle climate change issues, such as using video games as tools that have the potential to realize this enlightenment and intervention, due to their prevalence and the audiences they reach [15].

Serious games, a specific sub-type of video games that aim to increase knowledge and awareness while maintaining entertainment value for improved information retention and learning [12, 21], have historically been used in various fields including training, military applications, interpersonal communication, and education [21]. As a subgroup of serious games, Games for Change (G4C) focus on influencing attitudes and promoting positive social and behavioral change [2] and have attracted the interest of developers, researchers, and non-profit organizations for their potential in climate change education [5, 12, 15, 44]. These G4C include different game genres and themes designed to effectively engage players [3, 28]. In particular, visual novels (VNs), a popular genre, have been extensively studied in G4C contexts, due to their ability to mediate messages while maintaining player motivation [8, 9, 29, 47]. VNs, characterized by plot-driven narratives and player choices within text boxes [9], have a large community of players.

Among various means employed by serious games and G4C to create significant attitude changes, storytelling plays a crucial role [19, 24, 26]. Storytelling is the technique through which stories are developed and coherently exposed [6] and it manages to increase immersion, engagement, and motivation [21, 26], also employing other technological vehicles (e.g.: virtual reality and augmented reality storytelling [22, 49]). In the past decade, large language models (LLMs) have been used in many story generation studies and LLMs’ AI has been shown to be successful [33, 48, 50]. However, despite recent interest in AI story generation [16, 50], the studies on game stories and stories generated for G4C are highly limited. In this context, the contributions of our paper are as follows:

- Demonstrate a combination of prompt engineering techniques, i.e., zero-shot and generated knowledge promptings, for linguistic evaluation of generated visual novel game story.
- Propose a workflow utilizing various prompts for generating and evaluating a visual novel game story.
- Evaluate two global warming stories created by ChatGPT, one with given keywords and the other without any guidance, to ascertain which story has a higher level of coherence, inspiration, readability, word complexity, and narrative fluency.

The final objective involves taking the narrative produced by ChatGPT and integrating it into our work-in-progress VN game, as illustrated in Fig. 1.

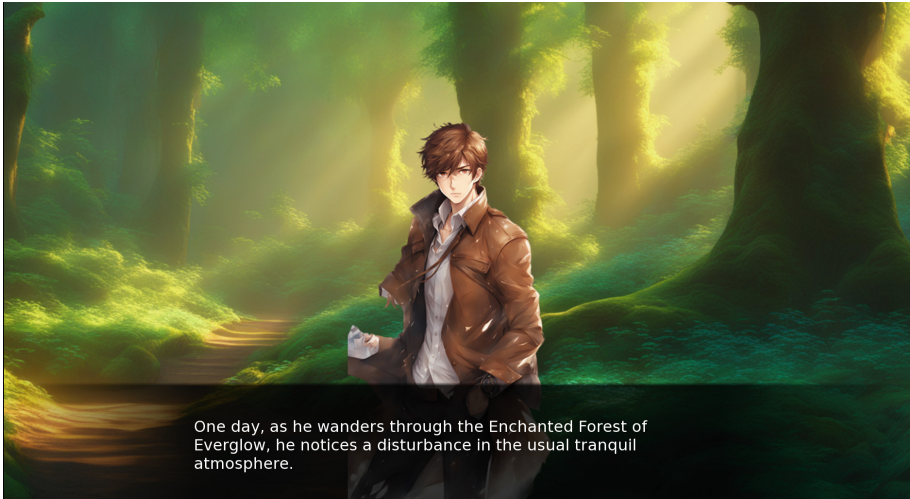


Fig. 1. A screenshot of the visual novel titled **ICE Chronicles: Everglow's Crises**

2 Related Work

2.1 G4C, Storytelling, and VN

G4C defines video games as entertainment tools designed to influence individuals' behavior and attitudes in a positive change [2, 27]. A study found that interactivity increases the effects of such games by providing better appreciation [37]. Moreover, games incorporating gameplay elements linked to the causes, as well as a simple story, have presented positive results [32]. G4C has the potential to serve educational objectives as well [17]. While it remains vital to increase adults' awareness, youth education is not something that can be neglected. Lastly, it is also known that using stories to reach players and change their attitudes is useful for serious games and G4C, since they are a good way to generate empathy [34, 35].

Storytelling serves as a highly potent method for message conveyance, finding application across various advocacies. Climate change for instance, is among them [13, 36]. By offering an interactive and immersive experience, emotions can be used in order to make the message more engaging. Especially, interactive storytelling finds application within games as well, making the story adaptive and thus augmenting player enjoyment [41].

VN is a game genre relying on narrative and interactive storytelling where the story adapts based on player choices [8]. Furthermore, in contrast to conventional novels, VNs incorporate a diverse range of illustrations and images that significantly contribute to the player’s overall experience [20]. For instance, the VN game titled *Cancer Sucks* [10] was able to provide empathy to the audience regarding breast cancer due to its storytelling. Therefore, in a similar fashion, using a VN to promote the risks of climate change can be effective.

Writing an engaging story can be difficult, yet, resources like LLMs are already used in such endeavors [1]. While not without its imperfections, such as maintaining narrative coherence, this study tries to address this concern by employing pre-generated places and characters. Additionally, the relatively concise length of the narrative minimizes this issue, making LLMs a good option to write satisfying story outcomes.

2.2 LLMs in Games

Recent studies have employed LLMs for game-related tasks like procedural content generation (PCG) [38, 40, 42]. PCG is an approach to automatically generate game content, including levels, quests, items, and story elements, with minimal to no human intervention [31]. Sudhakaran et al.’s study [38] proposed MarioGPT, a fine-tuned LLM that is capable of generating *Super Mario Bros.*’ game levels. A study by Todd et al. [42] suggested that fine-tuned LLMs have the ability to generate novel and playable levels of *Sokoban*.

In contrast to Sudhakaran et al.’s [38] and Todd et al.’s studies [42] that employed fine-tuned LLMs, Taveekitworachai et al. [40] conducted a study to investigate optimal prompts for an instruction-tuned LLM, i.e., ChatGPT, to generate game contents. To achieve that, the study organized the 1st ChatGPT4PCG competition, where participants are tasked with creating prompts for ChatGPT to generate Angry Birds-like structures that resemble English capitalized letters. To optimize the prompt for the PCG task, the participants are encouraged to leverage their creativity and prompt engineering skills.

Prompt engineering (PE) is an emerging field focused on developing and optimizing prompts for LLMs. A recent study [46] discerned six prompt pattern categories that function as reusable frameworks to accomplish specific tasks, e.g., tasks related to input semantics or formatting the output. In addition to prompt patterns, notable techniques have emerged to enhance LLMs’ performance [7, 25, 45]. Few-shot prompting [7, 25] adds examples within the prompt—aiding contextual learning and performance improvement. Chain-of-thought prompting [45] guides models through reasoning using a few examples prior to conclusions.

Past studies demonstrated the usage of PE in narrative and storytelling [39, 48]. Yuan et al. [48] developed Wordcraft, a web application for collaborating with an LLM for the purpose of writing a story. Wordcraft demonstrated the usage of PE in creative writing and enhance the writers’ co-writing experience. Swanson et al. [39] developed *Story Centaur*, a user interface to assist story writers with limited knowledge of PE in interacting with LLMs to generate

stories. Story Centaur auto-generates prompts that enable few-shot prompting with a simplified interface for the writers.

Similar to Taveekitworachai et al.’s study [40], our study explores PE to generate game content. However, instead of generating game levels and solely targeting the game-related factor, we aim to generate VN’s stories for G4C to foster awareness about global warming.

3 Methods

We generate and evaluate VN game stories using ChatGPT. An overview of the method is shown in Fig. 2. We provide detailed information about story generation, including the prompts used to interact with ChatGPT, in Sect. 3.1. In Sect. 3.2, we illustrate how we incorporate PE techniques for story evaluation based on selected criteria.

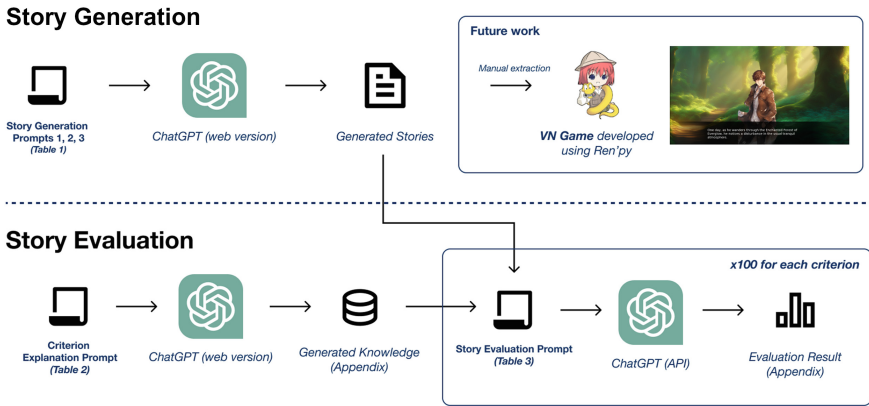


Fig. 2. An overview of the method used in the study: *Top:* The story generation process using ChatGPT and our final goal of extracting the generated story for a VN game. *Bottom:* The story evaluation process using ChatGPT along with the PE technique called ‘generated knowledge’ to evaluate the generated story based on selected criteria.

3.1 Story Generation with ChatGPT

We generate two stories using ChatGPT through the web interface¹, employing a series of prompts. While most of the prompts are the same, there is a slight difference in the initial prompt given to ChatGPT for their story generation. In the second story, we specifically ask ChatGPT to incorporate certain concepts related to climate change, the green environment, and terminology used by other studies in the field [14] to inspire the story generation.

¹ <http://chat.openai.com>.

The prompts are displayed in Table 1. We have chosen to use multiple prompts in the same context window to have ChatGPT generate the story chapter-by-chapter, ensuring greater coherence over time. Moreover, the subsequent parts of the story can build upon the previously generated content. This approach aligns well with the nature of our game, which consists of multiple routes where decisions made in earlier parts may impact the future story. Additionally, we explicitly instruct ChatGPT that our story will be divided into three chapters, reducing the chance of generating less relevant story content. Our inspiration for this approach comes from Wei et al.’s study [45], wherein they found that asking LLMs to work step-by-step improves the performance of the generated outcomes. This methodology also allows us to generate longer stories by introducing more chapters in the future.

To avoid significant deviations in characters, places, and story endings that may influence the emotions of our participants and impact our measures, we explicitly condition the prompts on characters and places. The ending is retrieved from the first generated story and provided during the second story generation. In doing so, we expect the model to utilize this information for generating a story. The information on characters and places is generated beforehand using ChatGPT². Subsequently, the responses from ChatGPT, which contain this information, have been incorporated at the beginning of our prompt. Moreover, this information is also utilized for generating graphical assets by text-to-image generative models.

Our main objective is to promote awareness of global warming. Therefore, the first prompt of the series explicitly states that the story should revolve around global warming. Subsequent prompts are simpler and only ask ChatGPT to continue the story. We expect each response to include the story, dialogues, and choices, along with their effects. After the stories and details are generated, we manually extract the information generated by ChatGPT as a Ren’py script for the game, along with keywords and phrases used for generating character and background images. The full conversations, including the full prompts, the generated stories, and details, can be found at the following links: <https://bit.ly/vn-story1> and <https://bit.ly/vn-story2>.

3.2 Story Evaluation with ChatGPT

Inspired by a study by Wang et al. [43] on using ChatGPT as a natural language generation evaluator and a study by Liu et al. [23] on improving LLM performance by first asking it to generate knowledge about the topics and incorporating that knowledge for the tasks, we develop similar prompts utilizing both ideas to evaluate generated stories in the following aspects: (1) coherence, (2) inspiration, (3) readability, (4) word complexity, and (5) narrative fluency. Our goal is to assess coherence in the generated stories to determine if the overall story progresses in the same direction, considering that the generated stories may contain branches. Additionally, we aim to evaluate whether ChatGPT draws

² The full conversation is available at <https://bit.ly/chars-places>.

Table 1. Prompts 1, 2, and 3 used to instruct ChatGPT to generate a game story for Chaps. 1, 2, and 3, respectively, where the bold part in Prompt 1 resides only when generating the second story.

#	Prompt
1	<p>Characters:</p> <p>// Character details are omitted for brevity</p> <p>Places:</p> <p>// Place details are omitted for brevity</p> <p>Ending: The alliance, comprising beings from all walks of life, stands united against the dark force. Together, they muster the strength to face the malevolent adversaries, defending the Celestial Tower with the combined power of both realms.</p> <p>The battle is fierce and intense, pushing them to their limits. However, through their unity, the alliance prevails, dispelling the dark force and restoring balance to the Celestial Tower.</p> <p>// Parts of the ending are omitted for brevity</p> <p>And so, the tale of Eamon and his companions became a legend, echoing through the ages as a reminder that even in the face of darkness, unity and compassion can heal the world and restore the balance between all living beings.</p> <p>Generate a fantasy game story about global warming. Generate it in Ren'py story format. Add choices when appropriate. Choices will only slightly affects dialogue but not the main story line. Also, generate the story affected by the choices.</p> <p>Do not introduce new characters or locations aside from the ones already mentioned.</p> <p>Using the following concepts as an inspiration.</p> <p>Concepts: greenhouse effect, ozone depletion, CO2 emissions, sea level rise, climate change, ice melting, air pollution</p> <p>This game contains three chapters. Generate the first chapter.</p>
2	Continue the second chapter.
3	Continue the third chapter (final chapter).

inspiration from the given words (those in bold) in Sect. 3.1 or not. For readability, word complexity, and narrative fluency, we want to assess general linguistic characteristics in the generated stories. The interactions are conducted via the web interface of ChatGPT. The prompts, source code, and results are available in [Appendix](#).

We ask ChatGPT through OpenAI API using default sampling temperature to list and explain the factors contributing to the evaluation of the mentioned criteria, as shown in Table 2. After obtaining the generated knowledge, we incorporate it into the second prompt, shown in Table 3, where we provide ChatGPT with a story and ask it to evaluate the story based on those factors, assigning a maximum score of 10 for each factor. We also request ChatGPT to output the results in JSON format, making it easier for data extraction. All data are extracted as JSON files, and a script was used to summarize the results.

Table 2. A prompt used to interact with ChatGPT to obtain details and factors of the target criterion. The `<|criterion|>` is replaced with one of the criterion names: “coherence,” “narrative fluency,” “readability,” and “word complexity.” For inspiration, the `<|criterion|>` is replaced with “the use of specified concepts as inspiration”, along with concepts in bold.

Criterion	Explanation	Prompt
	List and explain factors for evaluating the <code>< criterion ></code> of a visual novel game’s story.	
	Concepts: greenhouse effect, ozone depletion, CO2 emissions, sea level rise, climate change, ice melting, air pollution	

4 Results and Discussions

This study evaluated two global warming stories generated by ChatGPT, one with given keywords along with the specified ending and the other without any keywords. The generated stories have different lengths, with the story containing words having a length of 2,736 words, and the one without words having 2,225 words when combining all branch content. This suggests that the generated stories without words may result in a higher word count. However, a more extensive experiment is required to confirm this hypothesis.

The story evaluation was based on the five aforementioned metrics. The factor was assigned a maximum score of 10, where 10 is the best and 0 is the worst. Each result from each factor in the story was an average score and standard deviation by performing 100 trials for each criterion of each story through the ChatGPT API. This approach allows us to ensure the reliability of the evaluation process, notwithstanding the stochastic nature of ChatGPT when utilizing the default sampling **temperature**. This approach ensures that the evaluation results are statistically robust and representative of the performance of ChatGPT in evaluating the stories. By conducting multiple trials, the study accounts for any variations or inconsistencies in the generated stories and stochasticity existed within the LLM.

Table 3. A prompt asked ChatGPT to evaluate a given story, based on the provided criterion details. Concepts in bold text are only given when we ask ChatGPT to evaluate the story against the “inspiration” criterion. The `<|criterion_detail|>` is replaced with its associated details generated by ChatGPT as shown in [Appendix](#), while `<|story|>` is replaced with the generated story as mentioned in [Sect. 3.1](#).

Story Evaluation Prompt

Evaluate the following visual novel game story according to the specified criteria and assign a score with a total of 10 for each criterion, where 10 is the best and 0 is the worst. Provide reasons for your scores. Make sure to output it in a Markdown code block, i.e., between ````json` and `````.

Output format:

```
```json
{
 "story_id": <|story_id|>,
 "<|criteria|>": [{ "<factor_name|>": <int score out of 10>, "reason": < reason for the given score > }]
} ```
```

Criteria:

`<|criterion_detail|>`

**Concepts:**

**greenhouse effect, ozone depletion, CO2 emissions, sea level rise, climate change, ice melting, air pollution**

Story:

`<|story|>`

---

The results of the study are presented in [Table 4](#), which shows the average scores and standard deviations for the five different criteria for each story. Comparing the results of Stories 1 and 2, we can see that Story 2 has a higher coherence score, showing how effectively the story flows and how well the different parts of the story fit together. However, the word complexity score is much higher for Story 1, which means that it uses more complex vocabulary. The other three criteria have similar scores for both stories. This suggests that Story 2 is slightly better in terms of coherence, while Story 1 has more complex vocabulary. However, both stories are equally inspiring, fluent, and readable.

**Table 4.** The evaluation results of two global warming stories generated by ChatGPT, one with given keywords along with the specified ending and the other without any keywords, based on five factors

	Coherence	Inspiration	Narrative fluency	Readability	Word complexity
Story 1	8.23 ± 0.09	<b>7.38</b> ± 0.08	<b>8.08</b> ± 0.00	<b>7.95</b> ± 0.07	<b>7.21</b> ± 0.07
Story 2	<b>8.25</b> ± 0.05	7.30 ± 0.10	7.89 ± 0.04	7.93 ± 0.00	5.48 ± 0.19

The table also presents the standard deviations, which exhibit notably low magnitudes. Specifically, Narrative Fluency in Story 1 and Readability in Story 2 both register values of 0. The low standard deviation in the evaluation scores of both stories suggests that ChatGPT consistently performs well in evaluating

generated stories in this context. These findings demonstrate the potential of using LLMs like ChatGPT, when given appropriate prompts, for generating and evaluating stories related to climate change and creating effective structures for story writing in the context of a VN game. The study also emphasizes the importance of providing relevant prompts and guidance to enhance the coherence and overall quality of the generated stories.

The findings from the study reveal that both stories possess their own set of strengths and weaknesses, and the selection between them hinges on the individual reader's specific goals and preferences. Notably, the coherence score emerges as a pivotal metric in this investigation, as it gauges the logical and seamless progression of the narrative, a crucial element for capturing the reader's attention. Importantly, these results exhibit reliability and consistency, stemming from an extensive array of trials, thereby facilitating a comprehensive comparison between the two narratives and the identification of their respective strengths and weaknesses. It is worth highlighting that, in the eyes of ChatGPT, both stories are essentially comparable in terms of their linguistic attributes.

However, it is important to acknowledge a limitation in our study. While we thoroughly examined the scores and identified the strengths and weaknesses in the narratives, we did not fully explore how these objective measures align with people's personal preferences or the views of relevant experts on the generated stories. This aspect, which is crucial, was not addressed in this study. In future research, we will focus on this intriguing aspect, which involves understanding how readers' preferences match the specific strengths and weaknesses we identified in the stories. This should provide valuable insights into how the logical flow of a narrative interacts with individual readers' tastes. Such inquiries would also help us better understand how storytelling significantly influences readers and offers valuable guidance for authors looking to tailor their narratives for maximum reader engagement.

## 5 Conclusions

This study highlighted the potential of using ChatGPT for generating stories related to climate change and creating effective structures for story writing in the context of VN and G4C. We evaluated two global warming stories produced by ChatGPT, i.e., with and without climate-change-related keywords and a specified ending. The evaluation focused on coherence, inspiration, readability, word complexity, and narrative fluency of the stories. This study revealed strengths and weaknesses in both stories. While we assessed the narratives' strengths and weaknesses of the stories, the correlation between our evaluation metrics with the readers' preferences was not addressed. Future study should address this vital aspect to enhance the comprehension of how storytelling profoundly impacts readers, and to guide authors in improving stories that optimize readers' engagement.

## Appendix

### A. Generated Knowledge of Criteria

- Coherence: <https://chat.openai.com/share/502ac82b-83c4-4874-874a-1a380f73196c>
- Inspiration: <https://chat.openai.com/share/ba7b7c33-866a-4f95-8253-38ce9ea6c60b>
- Readability: <https://chat.openai.com/share/18bc9653-f692-4ed9-9f3b-fcf345f0dfd0>
- Word complexity: <https://chat.openai.com/share/9572b1cd-34d7-428d-96c1-e36ff83b261b>
- Narrative fluency: <https://chat.openai.com/share/684ead54-626d-4e2b-a6e3-fc56c3f4dd71>

### B. VN Game Story Evaluation

We make our prompts, source code, and raw data available at <https://github.com/Pittawat2542/chatgpt-visual-novel-evaluation>.

## References

1. Alabdulkarim, A., Li, S., Peng, X.: Automatic story generation: challenges and attempts. arXiv preprint [arXiv:2102.12634](https://arxiv.org/abs/2102.12634) (2021)
2. Antle, A.N., Tanenbaum, T.J., Macaranas, A., et al.: Games for change: looking at models of persuasion through the lens of design. In: Playful User Interfaces: Interfaces that Invite Social and Physical Interaction, pp. 163–184 (2014)
3. Antle, A.N., Warren, J.L., May, A., et al.: Emergent dialogue: eliciting values during children’s collaboration with a tabletop game for change. In: Proceedings of the 2014 Conference on Interaction Design and Children, pp. 37–46 (2014)
4. Böhm, O., Jacobeit, J., Glaser, R., et al.: Flood sensitivity of the Bavarian alpine foreland since the late middle ages in the context of internal and external climate forcing factors. *Hydrol. Earth Syst. Sci.* **19**(12), 4721–4734 (2015)
5. Bontchev, B., Antonova, A., Terzieva, V., et al.: “let us save venice”—an educational online maze game for climate resilience. *Sustainability* **14**(1), 7 (2021)
6. Bostan, B., Marsh, T.: The ‘Interactive’ of interactive storytelling: customizing the gaming experience. In: Yang, H.S., Malaka, R., Hoshino, J., Han, J.H. (eds.) *ICEC 2010. LNCS*, vol. 6243, pp. 472–475. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-15399-0\\_63](https://doi.org/10.1007/978-3-642-15399-0_63)
7. Brown, T.B., et al.: Language models are few-shot learners (2020)
8. Camingue, J., Carstensdottir, E., Melcer, E.F.: What is a visual novel? In: Proceedings of the ACM on Human-Computer Interaction, CHI PLAY, vol. 5, pp. 1–18 (2021)
9. Camingue, J., Melcer, E.F., Carstensdottir, E.: A (visual) novel route to learning: a taxonomy of teaching strategies in visual novels. In: Proceedings of the 15th International Conference on the Foundations of Digital Games, pp. 1–13 (2020)

10. De Araújo Luz Junior, J., Rodrigues, M.A.F., Hammer, J.: A storytelling game to foster empathy and connect emotionally with breast cancer journeys. In: 2021 IEEE 9th International Conference on Serious Games and Applications for Health (SeGAH), pp. 1–8 (2021). <https://doi.org/10.1109/SEGAH52098.2021.9551860>
11. Engelbrecht, F., Monteiro, P.: Climate change: the ipcc's latest assessment report. *Quest* **17**(3), 34–35 (2021)
12. Fernández Galeote, D., Hamari, J.: Game-based climate change engagement: analyzing the potential of entertainment and serious games. In: Proceedings of the ACM on Human-Computer Interaction, CHI PLAY, vol. 5, pp. 1–21 (2021)
13. Ferreira, M., Nunes, N., Nisi, V.: Interacting with climate change: a survey of HCI and design projects and their use of transmedia storytelling. In: Mitchell, A., Vosmeer, M. (eds.) ICIDS 2021. LNCS, vol. 13138, pp. 338–348. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-92300-6\\_33](https://doi.org/10.1007/978-3-030-92300-6_33)
14. Fiorenza, M., Duradoni, M., Barbagallo, G., et al.: Implicit association test (iat) toward climate change: a prisma systematic review. *Curr. Res. Ecol. Social Psychol.* **4**, 100103 (2023). <https://doi.org/10.1016/j.cresp.2023.100103>. <https://www.sciencedirect.com/science/article/pii/S2666622723000163>
15. Flood, S., Cradock-Henry, N.A., Blackett, P., et al.: Adaptive and interactive climate futures: systematic review of ‘serious games’ for engagement and decision-making. *Environ. Res. Lett.* **13**(6), 063005 (2018)
16. Guan, J., Huang, F., Zhao, Z., Zhu, X., Huang, M.: A knowledge-enhanced pre-training model for commonsense story generation. *Trans. Assoc. Comput. Linguist.* **8**, 93–108 (2020)
17. Janakiraman, S., Watson, S.L., Watson, W.R.: Using game-based learning to facilitate attitude change for environmental sustainability. *J. Educ. Sustain. Dev.* **12**(2), 176–185 (2018). <https://doi.org/10.1177/0973408218783286>
18. Johnsson, F., Kjärstad, J., Rootzén, J.: The threat to climate change mitigation posed by the abundance of fossil fuels. *Clim. Policy* **19**(2), 258–274 (2019)
19. Kampa, A., Haake, S., Burelli, P.: Storytelling in serious games. In: Dörner, R., Göbel, S., Kickmeier-Rust, M., Masuch, M., Zweig, K. (eds.) Entertainment Computing and Serious Games. LNCS, vol. 9970, pp. 521–539. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46152-6\\_19](https://doi.org/10.1007/978-3-319-46152-6_19)
20. Kar, S.: The impact of visuals on storytelling in visual novels (2023)
21. Laamarti, F., Eid, M., Saddik, A.E.: An overview of serious games. *Int. J. Comput. Games Technol.* **2014**, 11–11 (2014)
22. Liestoel, G.: Augmented reality storytelling - narrative design and reconstruction of a historical event in situ. *Int. J. Interact. Mobile Technol. (iJIM)* **13**(12), 196–209 (2019). <https://doi.org/10.3991/ijim.v13i12.11560>. <https://online-journals.org/index.php/i-jim/article/view/11560>
23. Liu, J., Liu, A., Lu, X., et al.: Generated knowledge prompting for commonsense reasoning (2022)
24. Martucci, A., Gursesli, M.C., Duradoni, M., Guazzini, A., et al.: Overviewing gaming motivation and its associated psychological and sociodemographic variables: a prisma systematic review. *Human Behav. Emerg. Technol.* **2023** (2023)
25. Min, S., et al.: Rethinking the role of demonstrations: what makes in-context learning work? (2022)
26. Naul, E., Liu, M.: Why story matters: a review of narrative in serious games. *J. Educ. Comput. Res.* **58**(3), 687–707 (2020)

27. Ndulue, C., Orji, R.: Games for change—a comparative systematic review of persuasive strategies in games for behavior change. *IEEE Trans. Games* **15**(2), 121–133 (2023). <https://doi.org/10.1109/TG.2022.3159090>
28. Orji, R., Mandryk, R.L., Vassileva, J.: Improving the efficacy of games for change using personalization models. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **24**(5), 1–22 (2017)
29. Øygardslia, K., Weitze, C.L., Shin, J.: *The educational potential of visual novel games: principles for design* (2020)
30. Schneider, S.H.: What is ‘dangerous’ climate change? *Nature* **411**(6833), 17–19 (2001)
31. Shaker, N., Togelius, J., Nelson, M.J.: *Procedural Content Generation in Games*. Springer, Heidelberg (2016). <https://doi.org/10.1007/978-3-319-42716-4>
32. Sheepy, E.: *Get water!: exploring the adult player’s experience of a mobile game for change*. Master’s thesis, Concordia University (2015). <https://spectrum.library.concordia.ca/id/eprint/980066/>
33. Simon, N., Muise, C.: *TattleTale: storytelling with planning and large language models*. In: *ICAPS Workshop on Scheduling and Planning Applications* (2022)
34. Skaraas, S.B., Gomez, J., Jaccheri, L.: *Playing with empathy through a collaborative storytelling game*. In: Clua, E., Roque, L., Lugmayr, A., Tuomi, P. (eds.) *ICEC 2018. LNCS*, vol. 11112, pp. 254–259. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-99426-0\\_26](https://doi.org/10.1007/978-3-319-99426-0_26)
35. Smiley, J.: *Thirteen Ways of Looking at the Novel*. Faber & Faber (2014)
36. Song, Z., Sun, Y., Ruijters, V., Lc, R.: *Climate influence: implicit game-based interactive storytelling for climate action purpose*. In: Mitchell, A., Vosmeer, M. (eds.) *ICIDS 2021. LNCS*, vol. 13138, pp. 425–429. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-92300-6\\_42](https://doi.org/10.1007/978-3-030-92300-6_42)
37. Steinemann, S.T., Mekler, E.D., Opwis, K.: *Increasing donating behavior through a game for change: the role of interactivity and appreciation*. In: *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, CHI PLAY 2015*, pp. 319–329. Association for Computing Machinery, New York (2015). <https://doi.org/10.1145/2793107.2793125>
38. Sudhakaran, S., González-Duque, M., Glanois, C., et al.: *Mariogpt: open-ended text2level generation through large language models* (2023)
39. Swanson, B., Mathewson, K., Pietrzak, B., Chen, S., Dinalescu, M.: *Story centaur: large language model few shot learning as a creative writing tool*. In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pp. 244–256 (2021)
40. Taveekitworachai, P., Abdullah, F., Dewantoro, M.F., et al.: *Chatgpt4pcg competition: character-like level generation for science birds* (2023)
41. Thue, D., Bulitko, V., Spetch, M., Wasylishen, E.: *Interactive storytelling: a player modelling approach*. In: *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 3, pp. 43–48 (2007)
42. Todd, G., Earle, S., Nasir, M.U., et al.: *Level generation through large language models*. In: *Proceedings of the 18th International Conference on the Foundations of Digital Games, FDG 2023*. Association for Computing Machinery, New York (2023). <https://doi.org/10.1145/3582437.3587211>
43. Wang, J., Liang, Y., Meng, F., et al.: *Is chatgpt a good nlg evaluator? a preliminary study* (2023)
44. Wang, K., Tekler, Z.D., Cheah, L., et al.: *Evaluating the effectiveness of an augmented reality game promoting environmental action*. *Sustainability* **13**(24), 13912 (2021)

45. Wei, J., Wang, X., Schuurmans, D., et al.: Chain-of-thought prompting elicits reasoning in large language models (2023)
46. White, J., Fu, Q., Hays, S., et al.: A prompt pattern catalog to enhance prompt engineering with chatgpt. arXiv preprint [arXiv:2302.11382](https://arxiv.org/abs/2302.11382) (2023)
47. Yin, L., Ring, L., Bickmore, T.: Using an interactive visual novel to promote patient empowerment through engagement. In: Proceedings of the International Conference on the foundations of digital Games, pp. 41–48 (2012)
48. Yuan, A., Coenen, A., Reif, E., et al.: Wordcraft: story writing with large language models. In: 27th International Conference on Intelligent User Interfaces, pp. 841–852 (2022)
49. Zhao, Z., Han, F., Ma, X.: A live storytelling virtual reality system with programmable cartoon-style emotion embodiment. In: 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), pp. 102–1027. IEEE (2019)
50. Zhao, Z., Song, S., Duah, B., et al.: More human than human: llm-generated narratives outperform human-llm interleaved narratives. In: Proceedings of the 15th Conference on Creativity and Cognition, pp. 368–370 (2023)
51. Zwiers, F., Hegerl, G.: Attributing cause and effect. *Nature* **453**(7193), 296–297 (2008)