Being Water: Collaborating with an LLM in an Interactive Digital Narrative (IDN) as Speculative Aesthetics

Rafaela Nunes[1,2][0009−0006−8588−6286], Tehri Marttila[1,2][0000−0003−2868−7206], Andrés Isaza-Giraldo[2][0009−0003−9054−1491], Paulo Bala[2][0000−0003−1743−0261], and Pedro F. Campos[2,3][0000−0001−7706−5038]

[1] Faculty of Fine Arts of the University of Lisbon, Portugal. rafaela@nenhures.net
[2] ITI/LARSyS, Lisbon, Portugal
[3] Wow!Systems, Funchal, Portugal

Abstract. *Being Water* is an IDN experience that speculates about the ways of being of water in the world. It combines 360 degree video, artifact-beings, voice-over and atmospheric sound. In a first iteration, users navigate through the environment to listen to authorial text. This paper's focus is on second iteration in which we experiment with integrating generative AI (genAI) to various degrees (e.g., augmenting interaction, making it replayable, sustaining a narrative) by using authorial text and scene context. We discuss preliminary findings in using LLMs as a collaborator to make meaningful additions to digital artworks.

Keywords: digital art · interactive digital narrative · large language models · speculative aesthetics.

1 Introduction

With the release of models like OpenAI's GPT [1], Midjourney [23], and Stable Diffusion [20] among others, we have witnessed a popularization of genAI technologies as they become more user-friendly, easy to access and highly adaptive to practice and research. Simply put, genAI generates outputs from learned features in data input, whether that is text, images, music or videos. GenAI models are trained on large datasets, learning patterns and structures. In recent years, user interaction with these models has been mostly done through text prompting and image prompting, which constitutes a practical way to explore the vast space of possible outputs, called latent space.

Despite the ongoing popularity surge of genAI within the general public, this technology has faced considerable controversy within creative practices. Mainly, the reported existence of intellectual property in the datasets used for the training of these models has sparked backlash from creators, which has often culminated in the filing of lawsuits against companies that developed these AI models [3, 21, 26]. While no final ruling has cleared out the intellectual property of said systems, the United States Copyright Office has revoked copyright

claims over images generated using MidJourney [23] for the visual novel *Zayra of the Dawn*[4] which constitutes a clear legal precedent. Moreover, AI bias is an emergent ethical concern, and while efforts to address it are largely focused on representativeness and fairness in datasets and algorithms, human and systemic biases seem to be the underlying factor, and AI should therefore be approached as a socio-technical system [22].

Despite these reactions, various artists have been thinking about and through these issues by working with different AI models. Work of artists like Mario Klingemann and Stephanie Dinkins represent artistic practices that engage with genAI in innovative ways. Motivated by this exploration of creative practice, in this work we aim to understand how artists are expanding their practices through the collaboration with genAI and engaging in speculative thinking about what could be meaningful ways to connect with others through interactive digital narratives.

In this paper, we present our work-in-progress artwork, *Being Water*, a digital space dedicated to seeing, feeling and thinking through the perspective of water in the spirit of speculative aesthetics. This interactive and immersive experience is a collaboration between four humans and genAI. Our aim is also to think how this artwork might transform by speculatively imagining the branching of possibilities afforded by this kind of collaboration. We present our artwork in its current shape, and proceed to present an iteration of it using a large language model (LLM) to deepen the interactive element of the artwork, allowing mutual imaginings between human and machine.

We worked with the AI through a perspective of collaboration, and not simply as a passive tool, acknowledging the way the different media we work with (text, code, film, audio and painting, as well as genAI) interact with us, with each other and with others that experience it. This way, we consider our collaboration with these different media and between ourselves as an agential type of creative practice. Nunes describes such an agential turn in art-making as "*a set of artistic practices which take as their theoretical and practical point of departure the whole of human and nonhuman relations and their natural-cultural contexts*" [17, p.7]. The work presented in this paper is of interest to other researchers and artists examining their creative practices as they incorporate genAI as a collaborator.

2 Related Work

Janet Murray has proposed concepts such as dramatic agency, meaningful replay, multisequential or multiform storyworlds as suggestions towards creating a vocabulary to describe how meaningful experience in interactive digital narrative (IDN) is created [16]. Murray observes that dramatic agency does not necessarily mean that the user can change the events of the story, but rather that the user feels that their interaction choices have coherent and meaningful consequences in the experience [16]. In intersecting Murray's work and genAI, LLMs could be

[4] https://www.copyright.gov/docs/zarya-of-the-dawn.pdf

used for attuning user interaction with consequences in the experience through programming the system instead of authoring for multiple possible interaction-consequence pairs. For example, by inviting the LLM to attune their textual response to the image they see on screen, it can create dramatic agency through coherence between interaction and experience.

Another important aesthethic value of IDNs, according to Murray, is whether or not the experience motivates replay, and whether these variations are trivial or dramatically meaningful [16]. Implementing genAI can contribute towards creating an immense quantity of material for IDNs with low effort (compared with the effort required for creating manually). Klingemann's work *Uncanny Mirror* [19] invites visitors to view themselves in a "mirror", in fact a screen, which shows them an image of themselves, yet created in real time by a GAN algorithm that was trained on images of the previous exhibition visitors. The image one sees is never the same, it changes constantly, and this invites to replay as a continued exploration of the GAN's vast latent space, whose dataset might also be changing.

Other aesthethic values of IDNs according to Murray include multisequential narratives as narratives in which multiple traversals are possible and multiform story systems in which a set of parameters can produce multiple parallel instantiations[16]. Koenitz calls the author of such a multiform story system the "*creator, system builder, and narrative architect*" [7] instead of referring to a storyteller, which would allude to the notion of a literary author. In *Not the only one* [4], artist Stephanie Dinkins trains a genAI on three generations of women's oral history of her Black American family. This results in a voice interactive experience in which interactors are able to converse with the system to learn about the Black American experience. Instead of following her initial intention to preserve her family's story, by implementing genAI, Dinkins creates a multiform narrative system in which interactors are free to explore the Black American experience according to their particular interests.

Nunes develops a practice-based research that stems from a transmedia creative practice which intersects painting, programming and genAI [17]. Nunes transforms her paintings into virtual entities, creating "artifact-beings" from an iterative process that imbues her paintings with human and nonhuman characteristics and concepts through the generative AI's interpretations of them. This way, the artist engages in imaginatively projecting artifact-being hybrids, questioning binary boundaries like beings and non-beings, artificial and natural. Nunes's work [17], albeit not implicitly focused on narrative as such, is another example of the creation of such a multiform story system in which parameters are set and both user input as well as the input of genAI influence the final "output" and experience.

3 Methdodology

Inspired by the work of other artists and the increasing availability of genAI models, we wanted to explore the process of adapting artworks to include genAI

as a collaborator. We do this through a practice-based research approach. Firstly, four collaborators (all authors in this paper) created a digital artifact sustaining a narrative that was meant to be experienced by an audience. Secondly, we expand our collaboration to include genAI models, looking for ways to not only deepen interaction within the digital artifact, but also maintain and reinforce the author's desired narrative for the artwork. In the following subsections, we describe some experimental iterations of the digital artwork.

3.1 *Being Water*

Our purpose with *Being Water* is to speculate about how water might perceive the world as a way of attempting to become sensitive to the perspectives of others. We follow Hayles' proposal of speculative aesthetics [6], in which one imaginatively projects into the worldview of other objects and beings based on evidence about their ways of being in the world as an attempt to escape anthropocentrism, never assuming that these speculations in any way represent the actual perception of the world of the entity, but as fictional entities that we called *waterbeings*.

Our initial iteration was based on a collaborative flow in which an authorial speculative text was written based on 360° videos of water sources. This text was used in Stable Difusion [20] to transform existing paintings from one of the authors to create images for the *waterbeings*' texture maps. In Blender [2], we procedurally generated organic shapes from these maps, the *waterbeings*. Using the Unity 2023.3 game engine [25], we brought together several types of media to create an interactive narrative experience (see fig. 1). The procedurally generated shapes and the abstract images were thus used to create *waterbeings* which move around the space and can be interacted with based on the user's distance to play voice recordings of the authorial text, alongside the ambient sound of the video and spatial music distributed around in droplets.

While this work used genAI, this was not done during runtime; conversely, the output was selectively curated. Although the moving authorial texts and changing videos guaranteed a new poetic experience for each player, the experience doesn't cease to be a finite game lacking the sense of ephemeral replayability that an ever-transforming object (like a mirror or video feedback [8]) might have.

3.2 LLM integration in *Being Water*

After completing the first iteration of *Being Water*, we reiterated on the work by experimenting with expanding the imaginative capacity of the speculative text through integrating LLMs. In this further development, we added an LLM layer to further deepen the interactive dimension of the work, allowing the AI to react to and generate content based on the text and environment that we created, as well as respond to the user's input. [5] This way, we aim to explore the AI's interpretation of poetic data.

[5] A video example of the different outputs is available online in https://vimeo.com/1017470826/6cd5d860df

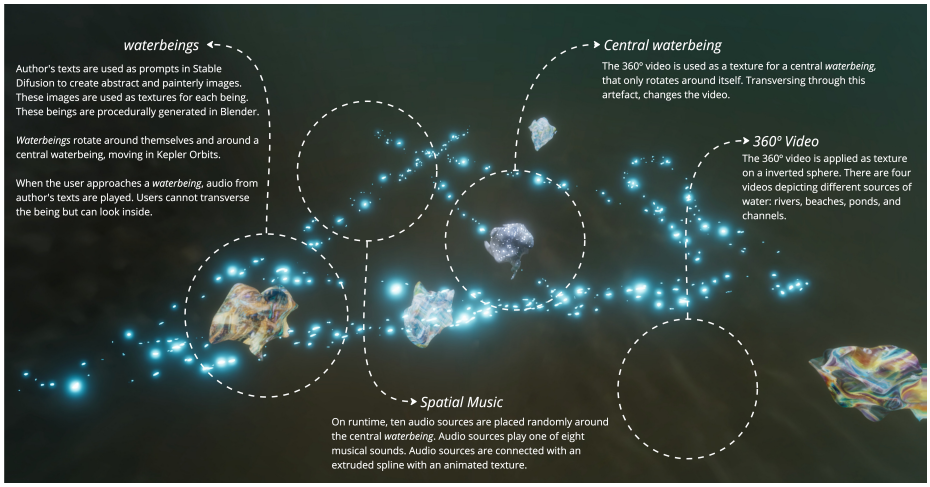

Fig. 1. First iteration of *Being Water*.

The purpose of this integration was to expand on the speculative imaginative aspect of the work by extending the human imagination to the collective, database-based universe of possibilities afforded by LLMs. Another purpose was to increase the replayability of the experience by making use of the variability that LLMs can provide for the content of the experience.

Building on the previous iteration, we used Ollama (version 0.3.10) [18], a framework designed to facilitate the deployment of LLMs on local environments, to communicate between Unity and a generative model. Given its versatility, we used a multimodal model LLaVA1.6 7b [9, 11, 10], to allow for text and image prompts. To replace voice recordings, we used the output of the LLM to create digital voice through Jets Text-to-Speech Model validated for Unity Sentis (Version 1.4.0-pre.2*) [24].

The authors had prompt writing sessions, which were streamlined iteratively after much trial and error. As instructions for the LLMs output, we used a base prompt (see fig. 2). Selecting this prompt required multiple iterations in order to guide the output of the LLM. The inclusion of the word "poem" or "poetic" in the prompt seemed to always result in the use of rhyme. The word artistic or art tended to add stereotypical notions to the output such as references to painting or creativity. At worst, the LLM would directly quote the key concepts in the prompt, revealing the inner workings of the system (e.g., "*An artistic reality is a mirror's smile.*").

We experimented with three different levels of integration of the LLM within the artwork, asking the LLM to generate text: (1) based on the authorial text only, (2) based on the user's current view and finally, (3) based on a combination of the authorial text and the image input.

Figure 2 shows some outputs from this experimentation. Asking the LLM to expand on the authorial text as the sole input resulted in the regurgitation

of recognizable portions of the text input. The LLM reordered sentences, mixed and matched and reformulated parts of the authorial text, but never managed to escape the confines of the imaginative capacity of the original text. In prior work, Marttila [14] experimented with an LLM-inspired performative approach to remixing their own text to draw new meanings from it and the resulting text output was far bolder than that of the LLM in these experiments. However, the LLM's imaginative capacity might in this case be limited by the small size of the text input, since the prompt confines the output to the authorial text.

Asking the LLM to base its output on the user view resulted in an overly abstract text, as evidenced in fig. 1, "*liquid thoughts weave*" or "*where color and form intertwine*". We presume this might be due to the abstract nature of the views which the experience presents, leading the LLM to interpret mostly colors and shapes, rather than objects.

By far the most meaningful speculative output was achieved by asking the LLM to base its output on both the authorial text and the user view. This is evidenced in fig. 2 – e.g., "*a frozen abstraction, the remnants of a body once fluid, now trapped in ice and stone*". While this output is clearly anchored in the authorial text, the LLM is expanding beyond the imaginative confines of the text by adding layers it "sees" in the experience such as "*a frozen abstraction*", or, as in another output, "*watery forms*". In this conjunction of embracing or interpreting the authorial text through the abstract visions it sees on screen, the LLM succeeds in expanding on human imagination by offering perspectives that were not present in the authorial speculative text.

4 Discussion

Collaboration with LLMs and generative art within an artistic context is still very recent, and yet, as we have seen, different artists have been taking advantage of this technology's capabilities, as well as exploiting its weaknesses to great effect. In this context, the importance arises of thinking about ways to meaningfully and thoughtfully collaborate with genAI within creative practices. In the case of the second iteration of *Being Water*, our purpose was first and foremost to invite genAI to expand on the imaginative and speculative components of the work.

As an author or artist, one expresses their perspective through their own unique view of the world. Collaborating with an LLM is quite different than working with another human, in the sense that the LLM does not have a singular, individual perspective, but it is a vast collection of different views, opinions and ideas — all represented through a single iteration at a time, "evened out" in some kind of virtual middle ground among this enormous multiplicity. Working with an LLM as we have described brought us the realization that we were effectively collaborating with a simulation of thought, a simulated non-perspective with such a type of collective.

In a sense, we are engaging in a mode of unprecedented second-order observation. Luhmann describes two modes of observation in society — first-order

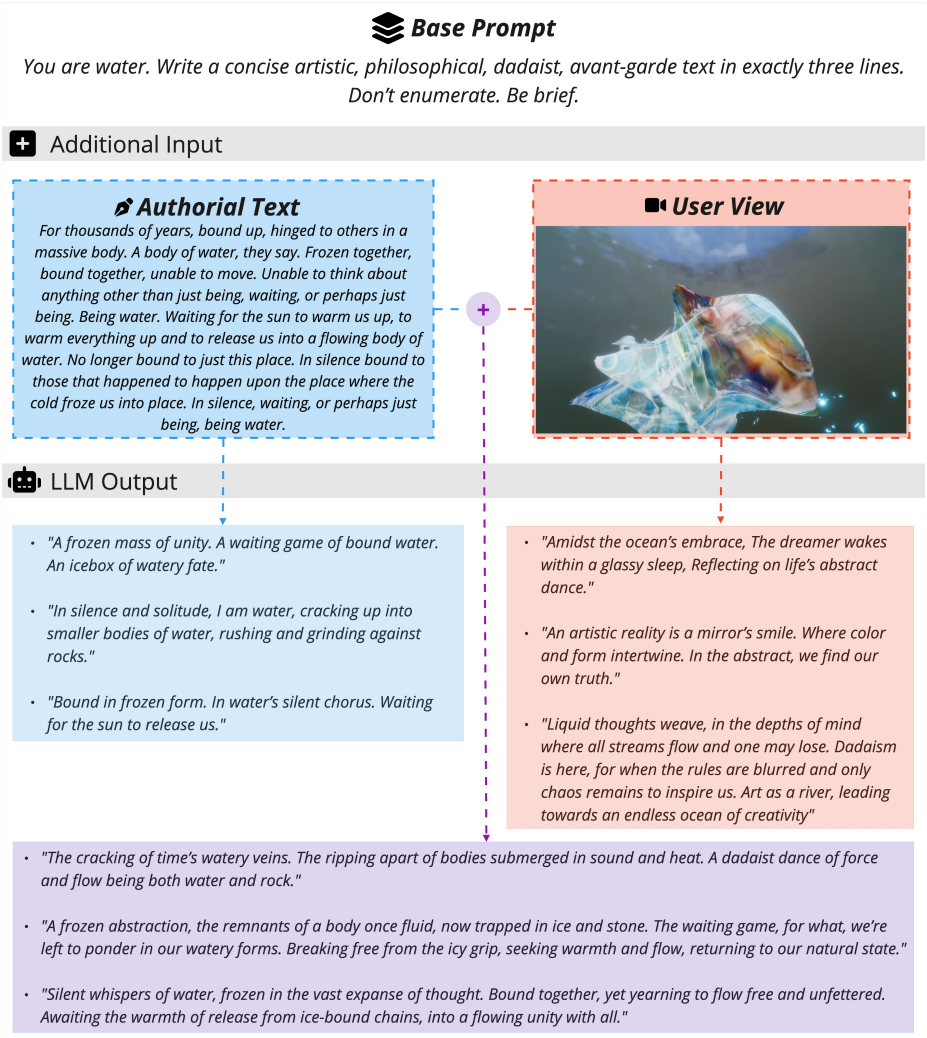

Fig. 2. Examples of outputs based on three scenarios of LLM integration

observation relates to what is directly experienced, and second-order observation to what is experienced as being observed by someone else — an observation of an observation [13]. Second-order observation has, according to Luhmann, "*become the advanced mode of perceiving the world in modern society*" [12, p.100], and this has been used to analyze the way we interact with others through and because of social media [15]. According to this theory, we can affirm that experiencing the AI's collective output through an LLM allows us to observe how we collectively observe the world. By interacting with AI like we do, we are engaging with AI-generated representations of human data.

In our LLM-integrated artwork, we are allowing the AI to create observations of what it is "seeing". It is undeniably transforming our experience based not on a single perspective, but on probability patterns in data. This attests to a uniquely different way of weaving interactive narratives with AI, and how we must take into account how it reflects a *representation* of multiple perspectives — limited as it reflects this "evening out" of perspectives into a single iteration. Such a view is a challenging way to think about the embodiment of knowledge and the gaze. Haraway describes attempts at "*views from nowhere*" as "*truly fantastic, distorted, and irrational*" [5, p.587].

Future work includes refining the prompt in order to better utilize the imaginative capacities of LLMs through expanding the level of agency afforded to it instead of limiting the LLM's output to the confines of the authorial text. We also suggest performing a fine-tuning training of the LLM, using a previously curated corpus of text, either of human generated or AI generated texts. This would generate results closer to those desired by the artists, guiding the model to a clearer form. Moreover, it could be possible to include conversational interaction within the user and the LLM, creating newly tailored text in each interaction.

5 Conclusion

This paper describes our process of experimenting with the integration of LLMs into an authorial text in the artwork *Being Water*. We discussed some of the affordances of IDN's and showed how existing work integrates LLMs in a meaningful way that expands the narrative experience. Supported by a first iteration of the artwork, we describe preliminary results of experimenting with LLM integration into the work. Future work includes further fine tuning our experiments with LLMs and analysing their use on the impact of the narrative experience.

Acknowledgments. This research was funded by the Portuguese Recovery and Resilience Program (PRR), IAPMEI/ANI/FCT under Agenda no.26, C645022399-00000057 (eGamesLab). The authors would also like to acknowledge the Portuguese Foundation for Science and Technology, for projects 10.54499/LA/P/0083/2020; 10.54499/UIDP/50009/2020 & 10.54499/UIDB/50009/2020.

References

1. AI, M.: ChatGPT (2022), https://chat.openai.com/
2. Community, B.O.: Blender (2018), http://www.blender.org, 3D modelling and rendering package
3. David, E.: George R.R. Martin and other authors sue OpenAI for copyright infringement. The Verge (Sep 2023), https://www.theverge.com/2023/9/20/23882140/george-r-r-martin-lawsuit-openai-copyright-infringement
4. Dinkins, S.: Not the Only One (2018), https://www.stephaniedinkins.com/ntoo.html
5. Haraway, D.: Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. Feminist Studies **14**(3), 575 (1988). https://doi.org/10.2307/3178066, https://www.jstor.org/stable/3178066?origin=crossref
6. Hayles, N.K.: Speculative Aesthetics and Object-Oriented Inquiry (OOI). In: Askin, R., Ennis, P.J., Hägler, A., Schweighauser, P. (eds.) Speculations V Aestehtics in the 21st Century. punctum books (May 2014). https://doi.org/10.2307/jj.2353877
7. Koenitz, H.: Understanding Interactive Digital Narrative: Immersive Expressions for a Complex Time. Routledge, London, 1 edn. (Mar 2023). https://doi.org/10.4324/9781003106425
8. Krauss, R.: Video: The aesthetics of narcissism. October **1**, 51–64 (1976), http://www.jstor.org/stable/778507
9. Liu, H., Li, C., Li, Y., Lee, Y.J.: Improved baselines with visual instruction tuning (2023), https://arxiv.org/abs/2310.03744
10. Liu, H., Li, C., Li, Y., Li, B., Zhang, Y., Shen, S., Lee, Y.J.: Llava-next: Improved reasoning, ocr, and world knowledge (January 2024), https://llava-vl.github.io/blog/2024-01-30-llava-next/
11. Liu, H., Li, C., Wu, Q., Lee, Y.J.: Visual instruction tuning. In: NeurIPS (2023)
12. Luhmann, N., Baecker, D., Gilgen, P.: Introduction to systems theory. Polity, Cambridge, UK ; Malden, MA (2013), oCLC: ocn819367834
13. Luhmann, N., Barrett, R.: Theory of society. Cultural memory in the present, Stanford University Press, Stanford, California (2012)
14. Marttila, T.: Speaking in loops that change - writing for the wandering listener in time, diffracted (2024), preprint on webpage at https://airdrive.eventsair.com/eventsairseasiaprod/production-experteventspublic/325a4a2820d4482ca779e149faf61c92
15. Moeller, H.G.: On second-order observation and genuine pretending: Coming to terms with society. Thesis Eleven **143**(1), 28–43 (Dec 2017). https://doi.org/10.1177/0725513617740968, http://journals.sagepub.com/doi/10.1177/0725513617740968
16. Murray, J.H.: Research into Interactive Digital Narrative: A Kaleidoscopic View. In: Rouse, R., Koenitz, H., Haahr, M. (eds.) Interactive Storytelling, vol. 11318, pp. 3–17. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-030-04028-4_1
17. Nunes, R.: AI, a Tool or an Author? A Posthuman Feminist Perspective on the Agency of Gen-AI in Creative Practices. Augmented Human Research **9**(1), 8 (Sep 2024). https://doi.org/10.1007/s41133-024-00074-8, https://doi.org/10.1007/s41133-024-00074-8
18. Ollama: Ollama. https://github.com/ollama/ollama (2024)

19. Onkaos: Uncanny Mirror by Mario Klingemann (May 2019), https://vimeo.com/336559940
20. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models (2021)
21. Schrader, A.: In a Blow for Artists, a Federal Judge Has Sided With Three A.I. Companies in a Copyright Dispute. Artnet (Oct 2023), https://news.artnet.com/art-world/federal-judge-sides-with-ai-companies-in-artists-copyright-dispute-2387654
22. Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., Hall, P.: Towards a standard for identifying and managing bias in artificial intelligence. Tech. Rep. NIST SP 1270, National Institute of Standards and Technology (U.S.), Gaithersburg, MD (Mar 2022). https://doi.org/10.6028/NIST.SP.1270, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf
23. Team, M.: Midjourney (Jun 2022), https://midjourney.com/
24. Unity: unity/sentis-jets-text-to-speech · Hugging Face, https://huggingface.co/unity/sentis-jets-text-to-speech
25. Unity Technologies: Unity (2023), https://unity.com/, game development platform
26. Vincent, J.: Getty Images sues AI art generator Stable Diffusion in the US for copyright infringement. The Verge (Feb 2023), https://www.theverge.com/2023/2/6/23587393/ai-art-copyright-lawsuit-getty-images-stable-diffusion