NarrativePlay: An Automated System for Crafting Visual Worlds in Novels for Role-Playing

Runcong Zhao[1*], Wenjia Zhang[1, 2*], Jiazheng Li[1*], Lixing Zhu[1], Yanran Li[3], Yulan He[1, 2, 4], Lin Gui[1]

[1]King's College London,
[2]University of Warwick,
[3]Independent Researcher,
[4]The Alan Turing Institute
{runcong.zhao, wenjia.1.zhang, jiazheng.li, lixing.zhu}@kcl.ac.uk
yanranli.summer@gmail.com, {yulan.he, lin.1.gui}@kcl.ac.uk

Abstract

In this demo, we present NarrativePlay – an innovative system enabling users to role-play a fictional character and interact with dynamically generated narrative environments. Unlike existing predefined sandbox approaches, NarrativePlay centres around the main storyline events extracted from the narrative, allowing users to experience the story from the perspective of a character they chose. To design versatile AI agents for diverse scenarios, we employ a framework built on a Large Language Models (LLMs) to extract detailed character traits from text. We also incorporate automatically generated visual displays of narrative settings, character portraits, and character speech, greatly enhancing the overall user experience.

Introduction

Most recent works utilise Large Language Models (LLMs) to produce human-like responses by using memories and thoughts stored in the databases (Ouyang et al. 2022; Park et al. 2023; AutoGPT 2023). It presents an exciting opportunity for creating an immersive and interactive environment akin to those featured in the television series "Westworld". However, current LLM-based methods for interactive agents typically focus on specific capabilities in set scenarios, often relying on manual settings for characters and environments (Park et al. 2023; Gao and Emami 2023; Xu et al. 2023), which require significant manual efforts and lack generalisability. Yet, we lack a universal framework for designing adaptable AI agents for varied scenarios.

In general, narratives contain extensive character-centric details such as "Who," "Objective," and "Relationships" (including Family, Friend, Competitor, etc.). They also provide details on Appearance (including Age, Gender, Hair colour/style, Eye colour, etc.), and Experiences (including Characters involved, Location, Description, Conversation), all of which can be utilised to craft vivid characters.

Most existing researches mainly focus on agent behaviours in handcrafted sandboxes (Riedl and Bulitko 2012; Côté et al. 2018; Hausknecht et al. 2020), which are labour-intensive and lack broad applicability. However, directly generating environments from descriptions presents challenges: (1) *Environment Extraction*. Settings are often ambiguously defined unless plot-critical. Thus, we suggest a method emphasising major storyline events from a chosen character's view, simplifying the identification of narrative settings. For instance, if the event is "Grandpa Joe discussing Willy Wonka's Chocolate Factory" in the "Grandparents' room", our goal is to visualise this scene. (2) *Environment Generation*. By utilising diffusion models (Koh et al. 2021; Rombach et al. 2022) and image generators, we address absent environment details (Alayrac et al. 2022). While generating knowledge for specific narrative settings is challenging, models trained on certain image styles, like fairy tales or animations, excel in this task.

We developed NarrativePlay (Zhao et al. 2023), a novel web-based platform capable of transforming narrative inputs into immersive interactive experiences. Our system synchronises text with visual displays of story settings, character portraits and speech, leveraging advanced multi-modal LLMs to enhance user experience. The demo[1] is available online.

Architecture of NarrativePlay

Main Storyline Extraction

We utilise the ChatGPT model `gpt-3.5-turbo` to extract structured information from chunked text. For an input narrative, our initial step is to solicit a list of the **Characters** involved using prompt. Subsequently, for each newly introduced character, we extract their core traits, appearance, and quotes. For each **Event**, we use prompts to extract descriptions, involved characters, locations, and conversations. This approach links events to their respective characters and locations, allowing concurrent event descriptions, through a character can't be in multiple events simultaneously. Conversations are extracted to capture any subevents and facilitate smoother transition to new conversations between users (i.e., users' chosen narrative characters) and agents (i.e., other characters in a narrative). To clarify the vaguely described

[*]These authors contributed equally.

[1]http://narrative-play.eastus.cloudapp.azure.com/

Figure 1: Our system's interactive process begins when a user provides a narrative to the system. They then choose a character as their narrative identity, through whom they can engage with the story. Users can have conversations with other characters, thereby experiencing the story in a more immersive way.

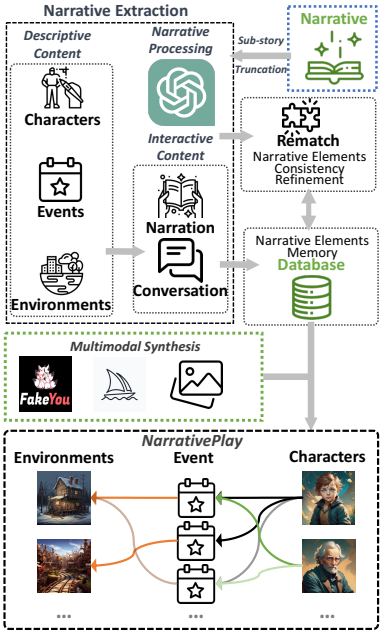

Figure 2: Demonstration of our *NarrativePlay* through a pipeline view.

Environments in narratives, we propose a location focused strategy of main storyline events and creating environment visuals rooted in the event descriptions. Generating images from event environment descriptions partly alleviates the issues of location co-referencing. Dynamic location changes, such as the onset of snowfall in winter, can be more easily represented in the generated images. Furthermore, our framework simplifies the process by directing the visibility among agents through their participation in shared events.

Narrative Image and Speech Synthesis

In this demo, we have incorporated an API request-based image generation service offered by Hotpot AI[2] into our framework for generating character portraits, which offers a more stable generation style. Additionally, we employ Midjourney[3] for event image generation as it provides more varieties and detailed pictures. Moreover, our framework also includes transforming narrative text into compelling speech, enriching the experience with an auditory dimension. We primarily employ Text-to-Speech models from the FakeYou platform[4] for this crucial task.

Main Storyline Progression

We progress the storyline with three stages (Figure 1):

1. **Narrative Input** The user begins by selecting or uploading their chosen narrative.
2. **Character Selection** Following above, NarrativePlay extracts the main storyline and subsequently presents information about the background and characters. Users are then asked to select from the listed major characters to begin their adventure. The agent's memory is initialised at this stage, laying the groundwork for future interactions.
3. **Story Progression** Once a character is selected, we present events related to the chosen character to the user. The scene image is displayed as the background picture, and the event description appears at the top of the page. Each event displays the involved characters on the left, with the user-selected character on the right. If there are conversations extracted for this event, they will be played first with voice renditions. Then, the user can click on other characters to chat.

Conclusions and Future Work

NarrativePlay, a novel platform, transforms narratives into interactive experiences, addressing challenges of storyline extraction, authentic character creation, and versatile environment design automatically. By focusing on the main events and leveraging advanced LLMs, it aligns text, image, and speech, marking a step forward in immersive interactive narratives. With a potential for wider applications like game generation, NarrativePlay paves the way for future advancements in narrative understanding.

[2]https://hotpot.ai/.

[3]https://www.midjourney. com/.

[4]https://fakeyou.com/.

Acknowledgements

This work was supported in part by the UK Engineering and Physical Sciences Research Council (grant no. EP/T017112/2, EP/V048597/1, EP/X019063/1). YH is supported by a Turing AI Fellowship funded by the UK Research and Innovation (grant no. EP/V020579/2).

References

Alayrac, J.-B.; Donahue, J.; Luc, P.; Miech, A.; Barr, I.; Hasson, Y.; Lenc, K.; Mensch, A.; Millican, K.; Reynolds, M.; Ring, R.; Rutherford, E.; Cabi, S.; Han, T.; Gong, Z.; Samangooei, S.; Monteiro, M.; Menick, J.; Borgeaud, S.; Brock, A.; Nematzadeh, A.; Sharifzadeh, S.; Binkowski, M.; Barreira, R.; Vinyals, O.; Zisserman, A.; and Simonyan, K. 2022. Flamingo: a Visual Language Model for Few-Shot Learning. In *Processing of the 36th Conference on Neural Information Processing Systems*. New Orleans, Louisiana, USA.

AutoGPT. 2023. Auto-GPT: An Autonomous GPT-4 Experiment. https://github.com/Significant-Gravitas/Auto-GPT. Accessed: 2023-08-01.

Côté, M.-A.; Kádár, A.; Yuan, X.; Kybartas, B.; Barnes, T.; Fine, E.; Moore, J.; Tao, R. Y.; Hausknecht, M.; Asri, L. E.; Adada, M.; Tay, W.; and Trischler, A. 2018. TextWorld: A Learning Environment for Text-based Games. arXiv:1806.11532.

Gao, Q. C.; and Emami, A. 2023. The Turing Quest: Can Transformers Make Good NPCs? In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics - Student Research Workshop*, 93–103. Toronto, Canada.

Hausknecht, M.; Ammanabrolu, P.; Côté, M.-A.; and Yuan, X. 2020. Interactive Fiction Games: A Colossal Adventure. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 7903–7910.

Koh, J. Y.; Baldridge, J.; Lee, H.; and Yang, Y. 2021. Text-to-Image Generation Grounded by Fine-Grained User Attention. In *2021 IEEE Winter Conference on Applications of Computer Vision*, 237–246. virtual.

Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C. L.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; Schulman, J.; Hilton, J.; Kelton, F.; Miller, L. E.; Simens, M.; Askell, A.; Welinder, P.; Christiano, P. F.; Leike, J.; and Lowe, R. J. 2022. Training language models to follow instructions with human feedback. In *Proceedings of the 36th Conference on Neural Information Processing Systems*. New Orleans, Louisiana, USA.

Park, J. S.; O'Brien, J. C.; Cai, C. J.; Morris, M. R.; Liang, P.; and Bernstein, M. S. 2023. Generative Agents: Interactive Simulacra of Human Behavior. arXiv:2304.03442.

Riedl, M. O.; and Bulitko, V. 2012. Interactive Narrative: An Intelligent Systems Approach. *AI Magazine*, 34(1): 67.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10674–10685. Los Alamitos, CA, USA: IEEE Computer Society.

Xu, Y.; Wang, S.; Li, P.; Luo, F.; Wang, X.; Liu, W.; and Liu, Y. 2023. Exploring Large Language Models for Communication Games: An Empirical Study on Werewolf. arXiv:2309.04658.

Zhao, R.; Zhang, W.; Li, J.; Zhu, L.; Li, Y.; He, Y.; and Gui, L. 2023. NarrativePlay: Interactive Narrative Understanding. arXiv:2310.01459.