Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/133/2025.20611

A Comprehensive Review of Transformer and Diffusion Models in Game Design: Applications, Challenges, and Future Directions

Chenhui Zhong[1,a,*]

[1]*School of Computer Science, Sun Yat-Sen University, 132 Waihuan Rd E, Guangzhou, China*
a. zhong.gu.gu@qq.com
**corresponding author*

Abstract: With the rapid advancement of artificial intelligence (AI) technologies, Transformer and Diffusion models have emerged as powerful tools across various fields, including game design. These models offer new possibilities for generating high-quality content and enhancing user experiences through personalized narratives. The integration of AI into game development promises to revolutionize how games are designed and experienced, potentially addressing limitations in traditional methods and opening avenues for more dynamic and engaging gameplay. However, despite the great potential of these advanced models, their use in game design is still an area that has not been fully explored, with limited systematic understanding of their capabilities, challenges, and future directions.This Scoping Review aims to provide a comprehensive overview of the current landscape of Transformer and Diffusion models in game design. By mapping out existing research, identifying key concepts, and highlighting gaps in the literature, this review seeks to establish a foundation for further exploration. It will analyze a broad range of studies to understand the diverse applications of these models, assess the methodologies employed, and explore the emerging trends in AI-enhanced game development. Through this process, the review will contribute to the broader discourse on AI's role in creative industries and inform future research efforts.

Keywords: Transformer models, Diffusion models, LLMs, Game design, Procedural Content Generation

1. Introduction

In recent years, with the rapid development of Artificial Intelligence (AI) technology, especially the major breakthroughs made in the fields of Deep Learning, Natural Language Processing and Computer Vision, the field of game design has ushered in unprecedented opportunities, and the gaming industry has developed rapidly. With the number of gamers and market size continuing to grow globally, game developers need more efficient and creative tools to meet the gaming industry's urgent need for innovative technology and high-quality content.

In this context, AI-driven game design has gradually become the focus of attention in both academia and the industry. In particular, the development of Transformer Models [1-3] and Diffusion

DOI: 10.54254/2755-2721/133/2025.20611

Models[4-7] provides new solutions to key issues such as Procedural Content Generation (PCG), personalized experience, and intelligent NPC behavior.

Previous work has summarized the working principles and applications of AI in game procedural content generation [8-11] as well as video games [12-14]. Based on this, the Transformer model has more possibilities [15-16] to improve text generation[17-18] by introducing the self-attention mechanism, giving the possibility to generate dynamic dialogues[19], as well as the ability to generate personalized quests [20-21], levels [22] and narratives [23], and to play an important role in automated game development [24].The Diffusion model, on the other hand, excels in image and video generation, not only making it possible to create realistic game environments [25] but also demonstrating its possibilities as a game engine [26] and level generator [27-28].

While Transformer and Diffusion models have yielded results in game content generation, there is a dearth of research on how to effectively combine them to enhance procedural content generation (PCG), a fusion that could potentially improve the quality and coherence of generated content. Existing studies have initially explored AI-generated game plots, but most of them rely on preset templates, failing to take full advantage of the potential of AI to create personalized experiences; in particular, the research on responding to player behaviors in real time and adjusting plots accordingly is still in its infancy. With the popularity of these models in game design, evaluating the quality of their generated content has become a key topic, but there is a lack of a unified and systematic evaluation framework to guide researchers and developers to objectively measure whether AI-generated game elements achieve the desired effect.

This study aims to fill the above research gaps by posing the following research questions: How can Transformer and Diffusion models be used to optimize procedural content generation in games? Can Transformer and Diffusion models help to implement a real-time responsive narrative engine for a truly personalized game experience? How to establish a comprehensive system of evaluation metrics to measure the quality of AI-generated game content? Together, these three research questions form a complete theoretical framework that aims to comprehensively explore how game design can be revolutionized through advanced AI technologies.

By answering these questions, this study seeks to construct a complete framework that will guide future researchers on how to effectively apply the Transformer and Diffusion models to game design. Specifically, this thesis will: provide a comprehensive review of the application of Transformer and Diffusion Models to game design, covering all aspects from the underlying principles to the most recent advances; analyze the major challenges that currently exist and discuss possible solutions; and explore the potential directions of applying these techniques in future game design, including but not limited to smarter NPC dialog systems, dynamic level design, and interactive experience improvement in AR/VR environments.

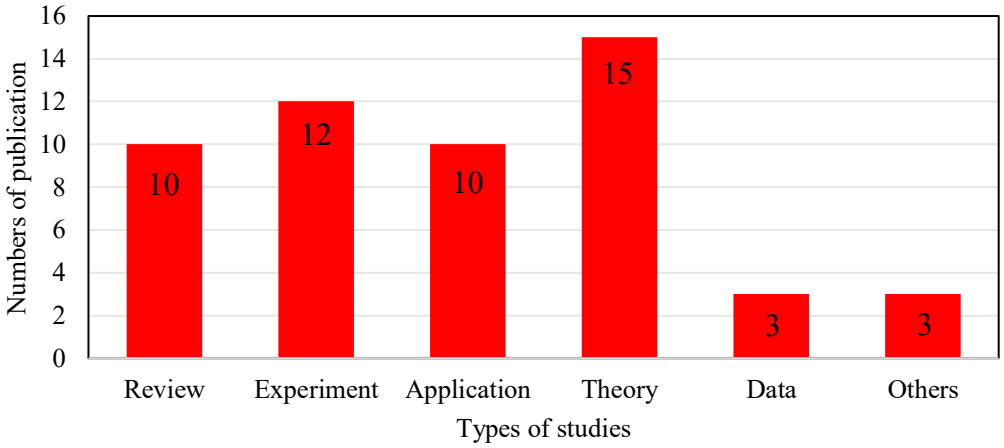
In conclusion, this review hopes to present readers with a clear technological panorama that will contribute to a broader discussion of the role of AI in the game industry, and provide a reference for future research efforts to stimulate more thinking and innovation about AI-enabled game design.

2. Literature Survey

Since Transformer models and Diffusion were only proposed in 2017 and 2018, respectively, and only started to be applied in games in recent years, the literature screening was mainly based on statistical indicators such as research methods and research fields. In Google Scholar, searching for literature in recent years with the keywords “Transformer models in game” and “Diffusion models in game” yielded 18,000 and 17,100 documents, respectively. After screening for relevance, 53 articles were selected for review. Table 1 shows the different types of studies involved in the literature.

DOI: 10.54254/2755-2721/133/2025.20611

Table 1: The different types of studies involved in the literature.



3. Applications of Transformer and Diffusion Models in Games

3.1. Text Generation and Dialog System

Akoury et al. [29] proposed a framework to evaluate the quality of LLM-generated NPC dialogues, showing that the Transformer model generates coherent and logical dialogues that enhance player immersion. Matyas and Csepregi's study showed that context-aware LLM-driven NPC dialogues significantly improved player engagement [30].

Promptable Game Models (PGM) implements text-guided game simulations using Masked Diffusion Models [28], allowing users to control the game progress through natural language and generating specific styles of dialogues to provide a personalized experience.

3.2. Level Design and Narrative Creation

Lee and Simo-Serra's work demonstrates the potential of Diffusion models as an efficient tool in Super Mario level generation [27], while Peebles and Xie explore a new type of Diffusion model based on the Transformer architecture, which performs well in ImageNet benchmarks [31].The SceneCraft framework utilizes large-scale language models to automate the generation of interactive narrative scenes, ensuring that storylines are both creative and consistent with authorial intent [32].

Ren et al. investigated the application of DDPMs and their variants in the Minecraft environment, demonstrating the ability of these models to generate high-quality images of game scenes [25].Che et al. introduced GameGen-X, the first Diffusion Transformer model dedicated to generating and interactively controlling open-world gameplay videos, which greatly improves the the diversity and interactivity of game content [33].

3.3. Intelligent NPC Behavior

The “Ghost in the Minecraft” framework proposed by Zhu et al. combines a large-scale language model with textual knowledge memory to create NPCs capable of handling complex tasks and adapting to uncertainty in open-world environments [34]. Hu et al. showed that a game agent based on a large-scale language model could realize highly anthropomorphic interactions through multiple functional components such as perceiving, thinking, and role-playing [35].

Valevski et al. explored the potential of Diffusion models as a real-time game engine, demonstrating that they can simulate the classic game DOOM at more than 20 frames per second on

DOI: 10.54254/2755-2721/133/2025.20611

a single TPU and dynamically adjust NPC actions to provide a smoother and more realistic interactive experience [26].

3.4. PCG Optimization

Menapace et al. showed how text-guided game simulations can be implemented through Masked Diffusion Models, incorporating Transformer to parse text descriptions as input to Diffusion models, guiding the latter to generate visual content that better matches expectations, improving the quality and flexibility of PCGs [28].

Buongiorno et al. proposed the PANGeA framework, which utilizes generative AI to generate narrative-consistent content and supports different sizes of LLMs for various devices [23]. Meanwhile, the Diffusion model can generate visuals according to the player's choices, ensuring that each player can have a unique gaming experience.

Nasir and Togelius showed how practical procedural content generation can be achieved through large-scale language models [36], while Che et al. emphasized the importance of combining multiple modalities in video generation, which increases the expressiveness of the generated content and provides a wide scope for future innovations [33].

4. Challenges and Future Directions

Despite the significant potential of Transformer and Diffusion models in game design, several challenges remain in optimizing procedural content generation (PCG), implementing real-time responsive narrative engines, and establishing comprehensive evaluation metrics for AI-generated game content [6-7].

4.1. Optimizing Procedural Content Generation (PCG)

Current research often focuses on single-model applications, lacking exploration into effective integration of Transformer and Diffusion models for multi-modal content generation [34]. Additionally, maintaining long-term coherence and consistency in generated content remains challenging [39]. Resource efficiency is another concern, particularly for deployment on mobile or resource-constrained devices.

To address these gaps, future research should explore efficient cross-modal fusion methods, such as using text to guide image generation or context-based adjustments. Techniques for fine-tuning pre-trained models can enhance long-term consistency while reducing repetition. Developing lightweight versions of these models will facilitate their application in resource-limited environments like mobile games or indie projects.

4.2. Implementing a Real-Time Responsive Narrative Engine

Most existing studies rely on preset templates, failing to leverage AI's full potential for creating personalized player experiences [40]. Real-time feedback mechanisms for dynamically adjusting plot developments based on player behavior are still in their infancy, especially in complex open-world games [41]. Moreover, current NPC behaviors are largely rule-based or statistically driven, lacking deep emotional understanding and dynamic responses to player emotions [42].

Advanced adaptive algorithms should be developed to provide highly customized narratives that reflect players' unique preferences and behaviors. Building robust real-time feedback loops will enable rapid analysis of player actions and corresponding plot adjustments, ensuring each player has a unique experience. Enhancing natural language processing capabilities, particularly in emotion

DOI: 10.54254/2755-2721/133/2025.20611

recognition and dialogue management, can lead to more emotionally intelligent NPCs capable of responding appropriately to player sentiments.

4.3. Establishing Comprehensive Evaluation Metrics

A unified, systematic evaluation framework for assessing AI-generated content quality is lacking, making it difficult to compare results across studies. Existing user experience measurement tools, such as miniPXI [43], may not fully capture the nuanced dimensions of player engagement. Automated evaluation methods like BERTScore [44] perform well in some areas but fall short when evaluating creative and artistic aspects.

A multidisciplinary platform should be established to develop standardized evaluation metrics involving experts from computer science, psychology, and human-computer interaction. Diversified user experience measurement tools, including eye-tracking and physiological signal monitoring, can offer richer data on player engagement. Improving automated evaluation methods by integrating human judgment through crowd-sourced assessments or hybrid scoring models could better capture the artistic and innovative qualities of AI-generated content.

5. Conclusion

This study explores how to optimize procedural content generation (PCG) in games, build real-time responsive narrative engines, and establish a comprehensive evaluation metrics system using Transformer and Diffusion models. Through a literature review, we found that although these models have made significant progress in text generation, level design, and intelligent NPC behaviors, challenges remain in cross-modal integration, long sequence processing, resource efficiency, and personalized experience. Existing automatic evaluation methods have limited performance in measuring creative and artistic content, and more comprehensive evaluation frameworks are needed.

It is worth noting that this study has some limitations. Due to the literature review approach, we were unable to analyze empirical data, which limits the depth of the conclusions. There are relatively few studies in certain areas, leading us to make inferences based on limited information. Therefore, future research should focus more on empirical validation to fill these gaps.

Future research should focus on developing efficient cross-modal fusion methods, such as combining text and image generation to enhance PCG quality; improving content coherence over long time spans by fine-tuning pre-trained models; developing lightweight versions of the models for resource-constrained environments; enhancing natural language processing techniques, especially sentiment analysis and dialog management, to create intelligent NPCs that are more aware of players' emotions; setting up a multidisciplinary cooperation platform to develop a standardized evaluation index system to assess the quality of AI-generated content.

References

[1] *Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. Neural Information Processing Systems.*

[2] *Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language Models are Few-Shot Learners. ArXiv, abs/2005.14165.*

[3] *Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners.*

[4] *Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. ArXiv, abs/2006.11239.*

[5] *Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2021). High-Resolution Image Synthesis with Latent Diffusion Models. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10674-10685.*

DOI: 10.54254/2755-2721/133/2025.20611

[6] Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E.L., Ghasemipour, S.K., Ayan, B.K., Mahdavi, S.S., Lopes, R.G., Salimans, T., Ho, J., Fleet, D.J., & Norouzi, M. (2022). Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. ArXiv, abs/2205.11487.

[7] Blattmann, A., Dockhorn, T., Kulal, S., Mendelevitch, D., Kilian, M., & Lorenz, D. (2023). Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets. ArXiv, abs/2311.15127.

[8] Yannakakis, G.N., & Togelius, J. (2011). Experience-Driven Procedural Content Generation. IEEE Transactions on Affective Computing, 2, 147-161.

[9] Hendrikx, M.J., Meijer, S., Velden, J.V., & Iosup, A. (2013). Procedural content generation for games: A survey. ACM Trans. Multim. Comput. Commun. Appl., 9, 1:1-1:22.

[10] Shaker, N., Togelius, J., & Nelson, M.J. (2016). Procedural Content Generation in Games. Computational Synthesis and Creative Systems.

[11] Liu, J., Snodgrass, S., Khalifa, A., Risi, S., Yannakakis, G.N., & Togelius, J. (2020). Deep learning for procedural content generation. Neural Computing and Applications, 33, 19 - 37.

[12] Levytskyi, V., Tsiutsiura, M., Yerukaiev, A., Rusan, N., & Li, T. (2023). The Working Principle of Artificial Intelligence in Video Games. 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST), 246-250.

[13] Yi Chan, E.M., Seow, C.K., Wee Tan, E.S., Wang, M., Yau, P.C., & Cao, Q. (2024). SketchBoard: Sketch-Guided Storyboard Generation for Game Characters in the Game Industry. 2024 IEEE 22nd International Conference on Industrial Informatics (INDIN), 1-8.

[14] Salge, C., Short, E., Preuss, M., Samothrakis, S., & Spronck, P. Applications of Artificial Intelligence in Live Action Role-Playing Games (LARP).

[15] Gallotta, R., Todd, G., Zammit, M., Earle, S., Liapis, A., Togelius, J., & Yannakakis, G.N. (2024). Large Language Models and Games: A Survey and Roadmap. ArXiv, abs/2402.18659.

[16] Sweetser, P. (2024). Large Language Models and Video Games: A Preliminary Scoping Review. ACM Conversational User Interfaces 2024.

[17] Xu, Y., Chen, L., Fang, M., Wang, Y., & Zhang, C. (2020). Deep Reinforcement Learning with Transformers for Text Adventure Games. 2020 IEEE Conference on Games (CoG), 65-72.

[18] Sudhakaran, S., Gonz'alez-Duque, M., Glanois, C., Freiberger, M.A., Najarro, E., & Risi, S. (2023). MarioGPT: Open-Ended Text2Level Generation through Large Language Models. ArXiv, abs/2302.05981.

[19] Nananukul, N., & Wongkamjan, W. (2024). What if Red Can Talk? Dynamic Dialogue Generation Using Large Language Models. ArXiv, abs/2407.20382.

[20] Ashby, T., Webb, B.K., Knapp, G., Searle, J., & Fulda, N. (2023). Personalized Quest and Dialogue Generation in Role-Playing Games: A Knowledge Graph- and Language Model-based Approach. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.

[21] Värtinen, S., Hämäläinen, P., & Guckelsberger, C. (2024). Generating Role-Playing Game Quests With GPT Language Models. IEEE Transactions on Games, 16, 127-139.

[22] Todd, G., Earle, S., Nasir, M.U., Green, M.C., & Togelius, J. (2023). Level Generation Through Large Language Models. Proceedings of the 18th International Conference on the Foundations of Digital Games.

[23] Leandro, J., Rao, S., Xu, M., Xu, W., Jojic, N., Brockett, C.J., & Dolan, W.B. (2023). GENEVA: GENErating and Visualizing branching narratives using LLMs. 2024 IEEE Conference on Games (CoG), 1-5.

[24] Chen, D., Wang, H., Huo, Y., Li, Y., & Zhang, H. (2023). GameGPT: Multi-agent Collaborative Framework for Game Development. ArXiv, abs/2310.08067.

[25] Ren, L. (2024). Application of Denoising Diffusion Probabilistic Models in the Minecraft Environment. CAIBDA.

[26] Valevski, D., Leviathan, Y., Arar, M., & Fruchter, S. (2024). Diffusion Models Are Real-Time Game Engines. ArXiv, abs/2408.14837.

[27] Lee, H.J., & Simo-Serra, E. (2023). Using Unconditional Diffusion Models in Level Generation for Super Mario Bros. 2023 18th International Conference on Machine Vision and Applications (MVA), 1-5.

[28] Menapace, W., Siarohin, A., Lathuilière, S., Achlioptas, P., Golyanik, V., Ricci, E., & Tulyakov, S. (2023). Promptable Game Models: Text-guided Game Simulation via Masked Diffusion Models. ACM Transactions on Graphics, 43, 1 - 16.

[29] Akoury, N., Yang, Q., & Iyyer, M. (2023). A Framework for Exploring Player Perceptions of LLM-Generated Dialogue in Commercial Video Games. Conference on Empirical Methods in Natural Language Processing.

[30] Matyas, L., & Csepregi The Effect of Context-aware LLM-based NPC Conversations on Player Engagement in Role-playing Video Games.

[31] Peebles, W.S., & Xie, S. (2022). Scalable Diffusion Models with Transformers. 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 4172-4182.

DOI: 10.54254/2755-2721/133/2025.20611

[32] Kumaran, V., Rowe, J., Mott, B.W., & Lester, J.C. (2023). SceneCraft: Automating Interactive Narrative Scene Generation in Digital Games with Large Language Models. Artificial Intelligence and Interactive Digital Entertainment Conference.

[33] Che, H., He, X., Liu, Q., Jin, C., & Chen, H. (2024). GameGen-X: Interactive Open-world Game Video Generation. ArXiv, abs/2411.00769.

[34] Zhu, X., Chen, Y., Tian, H., Tao, C., Su, W., Yang, C., Huang, G., Li, B., Lu, L., Wang, X., Qiao, Y., Zhang, Z., & Dai, J. (2023). Ghost in the Minecraft: Generally Capable Agents for Open-World Environments via Large Language Models with Text-based Knowledge and Memory. ArXiv, abs/2305.17144.

[35] Hu, S., Huang, T., Ilhan, F., Tekin, S.F., Liu, G., Kompella, R.R., & Liu, L. (2024). A Survey on Large Language Model-Based Game Agents. ArXiv, abs/2404.02039.

[36] Nasir, M.U., & Togelius, J. (2023). Practical PCG Through Large Language Models. 2023 IEEE Conference on Games (CoG), 1-4.

[37] Hu, C., Zhao, Y., Wang, Z., Du, H., & Liu, J. (2023). Games for Artificial Intelligence Research: A Review and Perspectives. IEEE Transactions on Artificial Intelligence, 5, 5949-5968.

[38] Salge, C., Short, E., Preuss, M., Samothrakis, S., & Spronck, P. Applications of Artificial Intelligence in Live Action Role-Playing Games (LARP).

[39] Hämäläinen, P., Tavast, M., & Kunnari, A. (2023). Evaluating Large Language Models in Generating Synthetic HCI Research Data: a Case Study. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.

[40] Sweetser, P. (2024). Large Language Models and Video Games: A Preliminary Scoping Review. ACM Conversational User Interfaces 2024.

[41] Buongiorno, S., Klinkert, L.J., Zhuang, Z., Chawla, T., & Clark, C. (2024). PANGeA: Procedural Artificial Narrative Using Generative AI for Turn-Based, Role-Playing Video Games. Artificial Intelligence and Interactive Digital Entertainment Conference.

[42] Hendrikx, M.J., Meijer, S., Velden, J.V., & Iosup, A. (2013). Procedural content generation for games: A survey. ACM Trans. Multim. Comput. Commun. Appl., 9, 1:1-1:22.

[43] Abeele, V.V., Spiel, K., Nacke, L.E., Johnson, D.M., & Gerling, K.M. (2020). Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. Int. J. Hum. Comput. Stud., 135.

[44] Zhang, T., Kishore, V., Wu, F., Weinberger, K.Q., & Artzi, Y. (2019). BERTScore: Evaluating Text Generation with BERT. ArXiv, abs/1904.09675.

[45] Yang, D., Kleinman, E., & Harteveld, C. (2024). GPT for Games: An Updated Scoping Review (2020-2024). ArXiv, abs/2411.00308.

[46] Pérez-Liébana, D., Liu, J., Khalifa, A., Gaina, R.D., Togelius, J., & Lucas, S.M. (2018). General Video Game AI: A Multitrack Framework for Evaluating Agents, Games, and Content Generation Algorithms. IEEE Transactions on Games, 11, 195-214.

[47] Levytskyi, V., Tsiutsiura, M., Yerukaiev, A., Rusan, N., & Li, T. (2023). The Working Principle of Artificial Intelligence in Video Games. 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST), 246-250.

[48] Guerrero-Romero, C., Lucas, S.M., & Pérez-Liébana, D. (2018). Using a Team of General AI Algorithms to Assist Game Design and Testing. 2018 IEEE Conference on Computational Intelligence and Games (CIG), 1-8.

[49] Liebana, D.P., Samothrakis, S., Togelius, J., Schaul, T., & Lucas, S.M. (2016). General Video Game AI: Competition, Challenges and Opportunities. AAAI Conference on Artificial Intelligence.

[50] Lanzi, P.L., & Loiacono, D. (2023). ChatGPT and Other Large Language Models as Evolutionary Engines for Online Interactive Collaborative Game Design. Proceedings of the Genetic and Evolutionary Computation Conference.

[51] Volum, R., Rao, S., Xu, M., DesGarennes, G., Brockett, C., Durme, B.V., Deng, O., Malhotra, A., & Dolan, B. (2022). Craft an Iron Sword: Dynamically Generating Interactive Game Characters by Prompting Large Language Models Tuned on Code. Proceedings of the 3rd Wordplay: When Language Meets Games Workshop (Wordplay 2022).

[52] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M.A., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518, 529-533.

[53] Hafner, D., Lillicrap, T.P., Ba, J., & Norouzi, M. (2019). Dream to Control: Learning Behaviors by Latent Imagination. ArXiv, abs/1912.01603.