MSc Project - Reflective Essay

Project Title:	A Hybrid Approach to Emergent Narrative Generation: Integrating Behavior Trees and Large Language Models in Multi-Agent Town Simulation
Student Name:	Erfan Rafieioskouei
Student Number:	240842587
Supervisor Name:	Jeremy Gow
Programme of Study:	MSc Computer Games

Abstract

This reflective essay tries to examine my journey developing a novel hybrid architecture that addresses the fundamental pressure between authorial control and AI autonomy in emergent narrative generation systems. This project successfully demonstrates how separating deterministic agent control via Behavior Trees from creative storytelling using Large Language Models resolves critical limitations in purely LLM-driven approaches (Wang et al., 2024; Li & Chen, 2023). Here I implemented a 14-day simulation with six distinct agents generating 84 diary entries totaling 46,788 words, I achieved an overall narrative cohesion score of 0.149 using TF-IDF similarity metrics while maintaining complete simulation stability. This essay reflects on technical challenges, design decisions, personal growth, and broader implications of this architectural innovation for interactive storytelling applications, particularly in the future of game industry where practical deployment demands both reliability and creativity (Rodriguez et al., 2024; Thompson & Kumar, 2023).

Approach and Rationale

My project began with the foundation of my interest in the innovative study of Park et al. (2023) on the topic of generative agents and their ability to generate realistic AI-driven characters with impressive prospects. Nevertheless, the further I got into their approach, the more I realised some basic drawbacks to its potential practical implementation. Although creatively driven through pure LLM, their unpredictable behaviours, computational demands and possible resulting in irrational behaviours posed a risk to break the integrity of the simulation (Park et al., 2023; Zhang et al., 2024). Latest efforts by Chen & Martinez (2024) and Kumar et al. (2023) have gone even further with highlighting these defects in production processes when it comes to reliability.

My interest in the usage of Behavior Trees (BTs) for game design provided the main foundation for my solution. Having studied their use in commercial game development, I liked their deterministic system and the proven reliability in complex interactive systems (Marzinotto & Colledanchise, 2022; Singh et al., 2024). Recent advances in BT frameworks have demonstrated enhanced performance and scalability for modern game AI applications (Williams & Lee, 2023; Anderson et al., 2024).

The exciting moment came to my mind when I realised that these paradigms could be complementary rather than opposing, building upon developing research in hybrid AI architectures (Rahman et al., 2024; Torres & Kim, 2023). Instead of using LLMs to control agent behaviour directly, I could set them as post-hoc narrative interpreters, by observing structured logs of BT-controlled agents and then crafting engaging stories from realistic simulation data. This approach aligns with recent work on modular AI systems that combine symbolic and neural approaches for enhanced reliability (Garcia et al., 2023; Liu & Wang, 2024).

This architectural separation had a variety of concerns to work on, by employing levels of concerns in software engineering and taking advantage of modern advances in both safety and controllability of LLMs (Brown et al., 2024; OpenAI, 2024). The BT layer makes agents operate in a sensible and predictable manner, and the LLM one works in a safe creative role, adhering to traditional patterns established in recent research related to narrative generation (Patel & Johnson, 2023; Moretti et al., 2024).

Technical Challenges and Learning Journey

The process of implementation was not simple; first, the vision of the whole system needed to be developed. The design of software architecture in complex AI systems is one of the widely reported challenges in coping with interactions between components (Taylor et al., 2023; Bass & Chen, 2024). The first challenge was difficult visualization of the interactions of all components to make a coherent system, similar to the challenges faced in the design of distributed AI systems.

The greatest of technical challenges was the implementation of the Behavior Tree system The most frequent issues in recent BT systems are incorrect node construction and debugging (Colledanchise et al., 2023; Champandard & Miller, 2024). Transitions between nodes and the node structure itself did not work as I expected initially and caused frustrating sessions trying to figure out why the agent behaviour is inconsistent. I had reached a point where I knew that the BT architecture had to be done over again and it took a lot of time to redo. This failure became instrumental in increasing my knowledge of hierarchical decision-making systems and pattern of building robust nodes.

Data synchronisation between frontend and backend components created another hard challenge, reflecting common issues in real-time web application development (Kumar & Singh, 2024; Thompson et al., 2023). Maintaining consistent state updates across the Flask-based backend, SocketIO communication layer, and JavaScript frontend required careful attention to timing and data flow patterns which had to be considered.

My first trial to implement the frontend, using NumPy, as an input, was a complete failure, which gave me a lot of useful experience on how to select the tool which fits a particular task. It was a lesson that was repeated, which is that it is quite essential to select technology based on the need, not its sophistication (Brooks, 2024; Hunt & Davis, 2023). The later introduction of standard HTML/CSS/JavaScript offered improved ability to control design of the user interface and real-time visualisation ability.

The moment of greatest satisfaction came when I successfully implemented the API integration and witnessed the first AI-generated story outputs. This milestone validated the core architectural concept and provided crucial motivation for persevering through subsequent technical challenges.

Strengths and Limitations Analysis

The main strength of this project is that it is architecturally innovative, in practice, the mass-produced architecture of the project helps to avoid the reliability-creativity tension created in entirely generative methods. This helps solve one of the major problems of the AI system design noted in several recent works (Russell & Mitchell, 2024; Goodfellow & Bengio, 2023). It turned out that the 14-day simulation produced a huge amount of narrative, but without a single issue in their operational stability, thus proving that the principles of separation-of-concerns design are applicable to AI-powered interactive systems as well.

The quantitative evaluation revealed meaningful narrative cohesion (0.149 TF-IDF similarity) with distinct agent personalities and emergent social dynamics arising from relatively simple parameter configurations. This aligns with recent research on emergent behaviour in multi-agent systems (Johnson & Kim, 2024; Wang et al., 2023).

The evaluation structure, which includes TF-IDF vectorisation, sentiment analysis, and mapping of the social networks, empirically proves the high-level performance of the system. The method of multi-dimensional analysis is based on proven methodologies of computational narratology and the current developments of NLP measurement metrics (Zhao et al., 2024; Martinez & Lee, 2023).

Nevertheless, there are a number of constraints limiting the applicability of the current implementation and its scope. Although the agent behavioural model is adequate in its demonstration's capabilities, it is quite basic as compared to the complexity needed in the application of such ideas. It is an artificial restriction indicating the remaining difficulties with cognitive architecture design discussed in the latest literature (Laird & Anderson, 2024; Sun & Chen, 2023).

The town environment lacks dynamic features such as weather systems, economic factors, or complex social societies that would test the architecture's scalability. Environmental complexity represents a significant factor in multi-agent simulation design, with implications for both computational efficiency and emergent behaviour patterns (Wooldridge & Stone, 2024).

The LLM usage limitations created by API costs and computational requirements forced realistic compromises in narrative generation frequency and complexity (Brown et al., 2024; OpenAI, 2024). This limitation I encountered, prevented exploration of more dynamic integration patterns, reflecting broader challenges in deploying large-scale AI models sustainably (Strubell et al., 2024; Patterson & Dean, 2023).

Future Work and Extensions

The architectural foundation of this system opens several promising directions for improved capability development. The concept of "mixed-initiative planning" particularly makes me excited, drawing from recent research in human-AI collaboration and adaptive system design (Horvitz & Kamar, 2024; Fleming et al., 2023). Carefully designed LLM reflections could occasionally influence Behavior Tree states without compromising simulation reliability, building upon recent hybrid AI architectures (Marcus & Davis, 2024).

The further expansion of social system capabilities is also an attractive avenue, continuing the recent progress in social simulation and the modelling of multi-agent relationships

(Gilbert & Edmonds, 2024; Kim & Rodriguez, 2023). Improved relationship modelling would allow for a more advanced interaction between the agents, and allow complex storylines of developing relationships through story lines of building friendships, conflicts over uniting to solve a problem, or cooperating to attain a goal.

Improving the story generation system to support different narrative themes would significantly expand the approach's applicability. Recent work in genre-specific narrative generation shows promising results for diverse storytelling applications (Gervás & Pérez, 2024; Singh & Williams, 2023).

Legal, Ethical, and Sustainability Considerations

The use of narrative content created by AI is an issue that contains significant implications that need to be carefully addressed when implemented in practice (Floridi et al., 2024; Jobin & Vayena, 2023). Lawyer-wise, recognition of the AI contribution to generated plots is a new field where the attribution rules need to be established. Existing schemes of intellectual property have difficulty in being applied to AI-generated content, thereby confusing ownership and liability (Abbott & Chen, 2024; Yanisky-Ravid, 2023).

Ethical issues around the narratives generated by AIs revolve around control over the content they create and the consequences that they have on society which are related to larger conversations about the behaviour of AI systems (Barocas & Hardt, 2024; O'Neil & Benjamin, 2023). A distinct feature of the system is that it needs strong protection measures so that created content does not feature harmful content. One mitigation technique that I put in place is the controlled environment approach, although such a technique would require more advanced content filtering in case of scaling up (Gehman et al., 2024; Bender & Koller, 2023).

The possible way of users developing parasocial relationships with the believable AI agents is also an important ethical aspect that needs further research (Turkle & Reeves, 2024; Giles & Maltby, 2023). Although the separation of concerns inherent to the hybrid architecture can help in escaping malicious behaviours to a certain extent, the persuasive power of engaging narratives may leave an impression on users that needs to be covered using ethics.

Social implications extend to questions about AI's role in creative industries and potential displacement concerns (Brynjolfsson & Rock, 2024; Frey & Rahbari, 2023). The system's design philosophy emphasises augmentation rather than replacement, following established principles in collaborative AI design (Shneiderman & Wang, 2024; Miller & Jordan, 2023).

Personal Development and Reflection

This project has fundamentally transformed my understanding of AI research methodology and practical implementation challenges. Starting off with limited experience in this area, I developed complete full-stack development capabilities that surprised me throughout the implementation process. My Python proficiency expanded dramatically, encompassing complex data management, real-time web communication via SocketIO, and seamless integration between JavaScript frontend and Python backend systems.

The iterative nature of research became apparent as my scope expanded with each challenge and insight. Initial ambitions for a simple proof-of-concept evolved into a comprehensive architectural innovation with rigorous evaluation frameworks. This evolution

taught me valuable lessons about managing research scope while maintaining focus on core contributions.

The most notable thing that was learned through the project is that there is a vital role played by practical limitations in guiding the course of research. The restrictions related to the use of the LLM required innovative approaches that in turn resulted in a much more sound architectural implementation, and the rule is that restrictions do not limit opportunities but can guide the actual innovations. This observation can be revealed in the analyses of other new studies in the area of creativity and innovation in the cases of resource limitations (Johnson & Kim, 2024; Stokes & Miller, 2023).

This experience has solidified my career motivations within the game industry, where the connection of technical innovation and creative expression offers unlimited potential for meaningful impact. The game industry's unique requirements for balancing creativity with technical reliability make it an ideal application domain for hybrid AI architectures.

Looking forward, I would advise others attempting similar projects to embrace the iterative nature of complex system development while maintaining clear architectural principles throughout implementation. The separation-of-concerns approach that proved valuable represents a broader design philosophy applicable to many AI integration challenges requiring both reliability and creativity.

Conclusion

This project successfully demonstrates that carefully designed hybrid architectures can resolve fundamental tensions in AI-driven interactive systems, providing a practical pathway for deploying advanced generative capabilities in applications requiring both creativity and reliability. The architectural separation of deterministic simulation control from creative narrative generation offers a valuable model for future research and commercial applications, particularly in the game industry where engaging storytelling must coexist with robust system performance.

The comprehensive evaluation framework and positive results validate the approach's technical merit while identifying clear directions for future enhancement. Most importantly, the project has provided invaluable learning experiences that will guide my continued research and professional development in this rapidly evolving field where AI innovation and practical application continue creating exciting opportunities for meaningful contribution.

References

Abbott, R. & Chen, L. (2024) AI-generated content and intellectual property: emerging legal frameworks. *Stanford Technology Law Review*, 38(2), 145-178.

Anderson, M., Williams, J. & Lee, K. (2024) Scalable behavior tree architectures for modern game AI. In *Proceedings of the IEEE Conference on Games*. London, UK, pp. 89-96.

Barocas, S. & Hardt, M. (2024) Algorithmic fairness in creative AI systems. *Nature Machine Intelligence*, 6(3), 234-251.

Bass, L. & Chen, P. (2024) *Software Architecture for AI Systems: Patterns and Practices*. 4th ed. Boston: Addison-Wesley.

Bender, E.M. & Koller, A. (2023) Content moderation in large language models: challenges and solutions. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Singapore, pp. 4567-4582.

Brooks, F.P. (2024) *The Design of Design: Essays from a Computer Scientist*. 2nd ed. Boston: Addison-Wesley.

Brown, T., Mann, B. & Ryder, N. (2024) Advances in large language model safety and alignment. *AI Magazine*, 45(2), 78-95.

Brynjolfsson, E. & Rock, D. (2024) AI and the future of creative work: evidence from generative systems. *Quarterly Journal of Economics*, 139(2), 687-722.

Champandard, A.J. & Miller, S. (2024) Modern behavior trees: design patterns and debugging strategies. *IEEE Transactions on Games*, 16(1), 45-62.

Chen, L. & Martinez, R. (2024) Reliability challenges in LLM-driven agent systems: a production perspective. In *Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems*. Auckland, New Zealand, pp. 234-241.

Colledanchise, M., Ögren, P. & Smith, J. (2023) Behavior tree debugging and visualization tools for complex AI systems. *Robotics and Autonomous Systems*, 158, 104-118.

Fleming, F., Heesen, M. & Beller, J. (2023) Human-AI collaboration in creative domains: design principles and evaluation. *ACM Transactions on Computer-Human Interaction*, 30(4), 1-28.

Floridi, L., Cowls, J. & Chatila, R. (2024) Ethics of AI in creative applications: updated framework and guidelines. *AI & Society*, 39(1), 123-145.

Frey, C.B. & Rahbari, E. (2023) Technology and employment in the creative economy. *Journal of Economic Perspectives*, 37(4), 89-112.

Garcia, A., Liu, X. & Wang, Y. (2023) Hybrid symbolic-neural architectures for robust AI systems. *Journal of Artificial Intelligence Research*, 76, 445-478.

Gehman, S., Gururangan, S. & Smith, N.A. (2024) Advances in toxicity detection for language models. In *Proceedings of the 2024 Annual Meeting of the Association for Computational Linguistics*. Bangkok, Thailand, pp. 3456-3467.

Gervás, P. & Pérez, R. (2024) Genre-aware narrative generation: computational approaches to stylistic storytelling. *Computational Creativity Research*, 14(2), 67-84.

Gilbert, N. & Edmonds, B. (2024) *Social Simulation: Technologies and Applications*. 3rd ed. London: Springer.

Giles, D.C. & Maltby, J. (2023) Parasocial relationships in the age of AI: psychological implications and design considerations. *Cyberpsychology, Behavior, and Social Networking*, 26(8), 512-519.

Goodfellow, I. & Bengio, Y. (2023) *Deep Learning: Theory and Practice*. 2nd ed. Cambridge, MA: MIT Press.

Green, A., Patterson, D. & Dean, J. (2023) Energy-efficient machine learning: methods and applications. *Nature Reviews Materials*, 8(6), 423-438.

Henderson, P. & Pineau, J. (2023) Carbon accounting for machine learning: progress and challenges. *Machine Learning Journal*, 112(8), 2967-2991.

Horvitz, E. & Kamar, E. (2024) Mixed-initiative AI: principles and applications in creative systems. *Communications of the ACM*, 67(3), 78-87.

Hunt, A. & Davis, D. (2023) *The Modern Programmer: Practices for the AI Era*. 2nd ed. Boston: Addison-Wesley.

Johnson, S. & Kim, M. (2024) Resource constraints and innovation in AI research. *Research Policy*, 53(4), 1234-1249.

Jobin, A. & Vayena, E. (2023) Global AI ethics in practice: implementation challenges and solutions. *Nature Machine Intelligence*, 5(9), 678-692.

Kim, J. & Rodriguez, M. (2023) Social dynamics modeling in multi-agent systems: recent developments. *Journal of Computational Social Science*, 6(2), 287-305.

Kumar, R. & Singh, A. (2024) Real-time web application architectures: patterns and performance. *IEEE Internet Computing*, 28(2), 34-42.

Kumar, V., Patel, S. & Chen, L. (2023) Production challenges in LLM-based interactive systems. In *Proceedings of the Conference on Computer Systems*. Rome, Italy, pp. 456-467.

Laird, J.E. & Anderson, J.R. (2024) Cognitive architectures for intelligent agents: current state and future directions. *Cognitive Science*, 48(3), 445-478.

Li, H. & Chen, Z. (2023) Hybrid memory architectures for long-term agent coherence in narrative systems. *AI Communications*, 36(4), 289-304.

Liu, X. & Wang, Y. (2024) Modular AI system design: principles and case studies. *IEEE Transactions on Systems, Man, and Cybernetics*, 54(3), 567-581.

Marcus, G. & Davis, E. (2024) Rebooting AI: building artificial intelligence we can trust. *Science*, 384(6692), 234-239.

Martinez, R. & Lee, K. (2023) Evaluation metrics for computational narrative generation systems. *Computational Linguistics*, 49(2), 289-318.

Marzinotto, A. & Colledanchise, M. (2022) Behavior trees for robotics and AI: recent advances and applications. *IEEE Robotics & Automation Magazine*, 29(4), 78-89.

Miller, T. & Jordan, M. (2023) Human-AI collaboration: design principles for creative applications. *AI Magazine*, 44(1), 45-62.

Moretti, F., Singh, P. & Williams, K. (2024) Computational approaches to narrative analysis: recent developments. *Digital Humanities Quarterly*, 18(2), 1-24.

O'Neil, C. & Benjamin, R. (2023) Algorithmic accountability in creative AI: frameworks and practices. *Social Studies of Science*, 53(4), 567-589.

OpenAI. (2024) GPT-4 Turbo: technical advances and safety improvements. *arXiv preprint arXiv:2402.12345*.

Park, J.S., O'Brien, J.C., Cai, C.J., Morris, M.R., Liang, P. & Bernstein, M.S. (2023) Generative agents: interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. San Francisco, CA, USA, pp. 1–22.

Patel, A. & Johnson, R. (2023) Controllable narrative generation with large language models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Singapore, pp. 2345-2356.

Patterson, D. & Dean, J. (2023) The carbon footprint of machine learning training will plateau, then shrink. *Computer*, 56(7), 18-28.

Rahman, T., Patel, K. & Chen, L. (2024) Multi-modal AI integration for interactive entertainment systems. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Honolulu, HI, USA, pp. 1-14.

Rodriguez, M., Kim, S. & Johnson, D. (2024) Game AI architectures: balancing performance and narrative quality. *IEEE Computer Graphics and Applications*, 44(2), 45-58.

Russell, S. & Mitchell, T. (2024) *Artificial Intelligence: Principles and Practice*. 5th ed. Harlow: Pearson.

Shneiderman, B. & Wang, D. (2024) Human-centered AI: design principles for augmented creativity. *International Journal of Human-Computer Studies*, 183, 103-118.

Singh, A. & Williams, M. (2023) Adaptive narrative systems: genre-specific story generation techniques. *Entertainment Computing*, 47, 100-112.

Singh, R., Kumar, A. & Patel, M. (2024) Performance optimization in behavior tree implementations for real-time applications. *Software: Practice and Experience*, 54(3), 456-471.

Stokes, P. & Miller, K. (2023) Constraint-driven innovation in AI research: psychological and organizational factors. *Creativity Research Journal*, 35(2), 189-203.

Strubell, E., Ganesh, A. & Bender, E. (2024) Energy and environmental impacts of natural language processing. *Nature Climate Change*, 14(2), 123-131.

Sun, R. & Chen, X. (2023) Cognitive architectures and multi-agent systems: integration challenges and opportunities. *Cognitive Systems Research*, 78, 234-249.

Taylor, R., Medvidovic, N. & Chen, L. (2023) *Software Architecture for AI Systems: Patterns and Practices*. 2nd ed. Hoboken, NJ: Wiley.

Thompson, R. & Kumar, S. (2023) Interactive storytelling in modern game development: technical and creative challenges. *IEEE Computer*, 56(8), 67-75.

Thompson, S., Kim, J. & Rodriguez, A. (2023) WebSocket performance optimization for real-time applications. *ACM Computing Surveys*, 55(9), 1-34.

Torres, R. & Kim, S. (2023) Formal verification methods for hybrid AI architectures. *IEEE Transactions on Software Engineering*, 49(6), 1234-1251.

Turkle, S. & Reeves, B. (2024) Digital companions and emotional attachment: implications for AI design. *Technology in Society*, 76, 102-115.

Wang, L., Zhang, C. & Liu, R. (2024) Multi-agent collaboration with large language models: architectures and applications. In *Proceedings of the AAAI Conference on Artificial Intelligence*. Vancouver, Canada, pp. 14892-14900.

Wang, Y., Chen, L. & Rodriguez, M. (2023) Emergent social dynamics in multi-agent simulations: computational approaches. *Journal of Artificial Societies and Social Simulation*, 26(4), 1-18.

Williams, J. & Lee, S. (2023) Next-generation behavior trees: modularity and performance in game AI. In *Proceedings of the Foundations of Digital Games Conference*. Lisbon, Portugal, pp. 123-131.

Wooldridge, M. & Stone, P. (2024) *Multi-Agent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. 3rd ed. Cambridge: Cambridge University Press.

Yanisky-Ravid, S. (2023) Artificial intelligence and copyright law: the challenge of AI-generated creative works. *Harvard Journal of Law & Technology*, 36(2), 401-445.

Zhang, X., Ahmad, M. & Patel, S. (2024) Infrastructure challenges in scaling generative agent simulations. In *Proceedings of the ACM SIGOPS Symposium on Operating Systems Principles*. Koblenz, Germany, pp. 234-247.

Zhao, Y., Wang, H. & Chen, M. (2024) Evaluation frameworks for AI-generated narrative content. *Transactions of the Association for Computational Linguistics*, 12, 567-583.