Masters of Negotiated Studies: Negotiated Project 3 (40 Credit)

Large Language Model Implementation into Games

Christopher Boyce

Student Number: 19016871
Email: b0196871j@student.staffs.ac.uk
Supervisor: Shaun Reeves

Glossary

Large Language Model (LLM): A deep learning system that processes and generates human-like text based on vast datasets.

GPT (Generative Pre-trained Transformer): An AI model that uses transformer architecture to generate text from pre-training on large datasets and fine-tuning.

Natural Language Processing: A computational techniques to understand, interpret, and generate human language.

Fine-Tuning: Is the process of refining a pre-trained AI model on a specific dataset to improve task-specific performance.

Real-Time Processing: The immediate handling and analysis of data as it is generated or received.

Key Words

Large Language Models, OpenAI, AI Integration, Games Development, Narrative Generation

Contents

Section 1: Introduction

Large language models are a form of artificial intelligence used to understand and generate text outputs. They utilize large data sets to train a machine-learning algorithm, allowing for the interpretation and relaying of data. In 2023, an extreme spike in users occurred with the release of ChatGPT, which garnered over one hundred million users in less than a year and reportedly had two million developers building integrations with OpenAI's API (Duarte, 2024). 2023 has been recognized by several sources as the first year of the "AI Revolution." This theory suggests that artificial intelligence will have an impact as significant as the Industrial and Digital Information Revolutions (Makridakis, 2017).

AI has entered many sectors and has had a significant impact, with the games industry being no exception. The development of video games using large language models is currently in its infancy, but researchers have begun using LLMs for tasks such as text-based action games (Yao, et al., 2020) and procedural content generation (PCG) for level creation (Sudhakaran, et al., 2023). This has become more accessible recently due to locally run LLMs becoming more available, along with AI-focused hardware technologies like Nvidia CUDA Acceleration, which significantly improves the efficiency of running complex AI models, making real-time applications in gaming more feasible.

However, there are still many limitations to the current use of AI models, such as the real-time responsiveness of the models. This is caused by processing time, whether locally or via servers through an API. Therefore, research into mitigation techniques to address the challenges associated with real-time responsiveness is crucial for the future of AI in games.

The success of integrating AI into the gaming industry could profoundly impact the gaming experience and the development process. It can enhance storytelling and realism within virtual environments, leading to more dynamic content and unique experiences for players. AI integration can also accelerate the development process by reducing the time required to create NPCs and develop storylines. In the long term, it can enable new and innovative design possibilities that haven't been explored yet, much like Virtual Reality and Augmented Reality have done in their respective domains.

Aims

This research paper will investigate and develop a solution for implementing large language models into modern game engines and explore player interaction with these models through NPCs. The focus will be on creating dialogue systems and AI-generated quests to facilitate this interaction.

Objectives

- Identify potential use cases for large language models in games and analyse the methods used in research.
- Develop suitable solutions for Unreal Engine and implement them into a vertical prototype.
- Analyse the features of large language models in the vertical prototype.
- Reflect on issues and devise potential fixes or mitigation techniques.

Project Planning

This project will begin by reviewing the current integrations and solutions that have been developed. Using this knowledge of existing research in the field, the project will create an

Unreal Engine project that uses OpenAI's API and locally run large language models (LLMs). The focus will be on dialogue systems and AI-generated quests to create interactive experiences for players. The outcomes will be analysed during a testing phase, with identified issues broken down and discussed.

Section 2 : Literature Review

Quest Generation

A study on quest generation was conducted using a modified version of GPT-2 (Värtinen, et al., 2024).Researchers used data from six role-playing games to train GPT-2 on the general structure of quests from these games. Their findings indicated that the quest description was unacceptable for the intended quest 80% of the time, a significant percentage of instances where the output didn't fit the quest. They suggested that this could be improved with GPT-3, and with GPT-2 now deprecated, GPT-3.5 Turbo and GPT-4 remain the current viable options, which should not pose a problem.

An important section of the study discusses the "Quest Ingredients" (Ammanabrolu, et al., 2020) needed to generate each quest, providing a comprehensive list that can guide the development of this artifact. The study also explored the impact of top-p values and temperature on the quality and accuracy of narrative output, discovering a trade-off depending on the values. They recommended values that balance these factors, which will be used in the project's development but may need adjustment depending on the outcomes.

Another study that uses Llama-2 studies several ways to create quests and compares them to prewritten solutions (Mishra, 2023) .As it's shown when giving the LLM integrated world the performance of the quest generation is significantly improved. The quests generated without world data produced worse descriptions and would generate conflicting information. The researcher advised including relevant information and details to create quests that stay true to the overall game narrative.

Narrative Generations

A researcher using LLMs made a charades game. They concluded the biggest limitation of using LLMs was that you can't control the interactions precisely (Frans, 2020), which was more apparent when using pre-trained models making the LLM display behaviors the designer didn’t intend. This is something that will have to be evaluated and something that the prototype will have to consider.

When looking into dialogue generation one of the key issues presented is the generation of concepts that don’t exist within the game world (Kalbiyev, 2022). It is noted the closer the game is to the real world the less impactful this will be on the generated text. It is important to specify to the LLM the world around them as they can only relay information they have been given. An investigation into the coherence and grammar returned data that showed that it didn’t have any major issues but was comparable to a poorly written human response which wouldn’t be suitable for a narrative game (Kalbiyev, 2022). This again was using a modified GPT-2 called “FallouGPT” which used Fallout 4 (2015) dialogue as training data. The researcher had an issue making conclusive reasons for why the program produced “Lackluster performance” but recommended using datasets with large data pools and more comprehensive training data which will be investigated as options for this project.

A research paper into the dangers of using LLMs has a section on narrative in 2D and 3D games (Murray, et al., 2023). There are many interesting points first being the engagement of the users being heavily reliant on the willingness to give creative responses to the AI. This has been shown to also similarly cause issues with narrative coherence as the player can meld the NPCs to believe or act in unintended ways. The user can also “Jailbreak” the AI as many LLMs have security risks or prompts that can break the intended uses. A commonly known one is called

“DAN” Do-Anything-Now which allows GPT to bypass the intended safeguards. Inworld AI a service that leverages GPT-4 has claimed to have added restricted models therefore NPCs and AI characters cannot be broken, but their service is a pay-per-use. Safety is a key feature that is going to have to be studied if this project is going to be widely distributed and basic safeguards will have to be checked to make testing this project safe for the participants in the survey. The conclusion of this paper suggests that LLMs aren’t proven as dangerous but the testing of novel use cases is still not known and at this current time we can only look at what has been widely done.

Other LLM Applications in Games

In research conducted by Microsoft, a Large Language Model was used to generate levels in VR (Roberts, et al., 2022). This used a prompt to level generation using ChatGPT3.5 and the OpenAI codex. This research explored the capabilities of generating not only visual assets but also generated code. They found serval challenges throughout the research the first being latency, many of the problems they encountered was downloading and loading the 3D models into Unity. The interesting part is the compiled and interpreted languages. This is an issue with the engine not being able to add external DLL assemblies at runtime therefore the researchers used “Roslyn” a compiler that allows the code to compile dynamically at runtime. Looking at this will potentially become a critical part when making the quest system in the project. One of the technical changes that has happened since this paper was released was the depreciation of the OpenAI codex which now is recommended to migrate to GPT-4 (OpenAI, 2024).

A similar study using GPT-3 was used to generate levels in a game called “Sokoban”. This is a 2D game that had the map made from a tile set (Todd, et al., 2023). The first difference with this is they don’t use code interpretation they instead have it linearly generate the map that uses characters to represent the different tiles. They used three measurements to rate the levels that the LLM created playability, novelty, and diversity. In the conclusion they were impressed with the range of tasks that the LLM could produce and once using training data they could improve the results. This is something that this project will have to see if it is viable.

Llama2 Model

Llama2 as a locally run LLM has many benefits compared to using an API. The first is data protection as the data doesn’t get transferred to a server it allows the player to have full control of input data and doesn’t have a reliance on third-party services (Roumeliotis, et al., 2023). Using a local LLM also allows uninterrupted service as it doesn’t rely on the internet or servers' computing power. The third main benefit is the latency and responsiveness of the LLM as it does not have to wait for the upload and download of data and it can instantly be accessed. On the contrary, running locally means more power and computing is placed on the user which will make the game inaccessible to a large percentage of players. This can impact performance in the game and if not on a newer computing architecture like CUDA from Nvidia it might not be able to run. LLMs also have a major issue of unforeseen scenarios as data isn’t sent to a server if issues arise the developer cannot see the issues and analysis them compared to if they could read the log of all interactions.

Llama2-7b has implemented safety boundaries dubbed Llama Guard when the user inputs potentially dangerous content. This covers "Threats of violence, Hate Speech, Illegal activity, self-harm and suicide, Sexual exploitation and abuse, severe misinformation and cybersecurity threats" (Hakan Inan, 2023). When testing the guard, it performed better than OpenAI's API and Perspective API; the researchers wanted to improve this, so they used training data from the

ToxicChat dataset. This measurably improved the guard's performance, even with only 20% of the dataset. Llama Guard can also be modified with novel datasets that can increase data efficiency and performance. Llama only knows information inside its training data therefore it can lead to unsuccessful assessments when asked outside that knowledge base. It also has a limitation with language with the training data being English therefore the guardrails are weaker in other languages. There are also issues with the misuse of the guard rails because it can be bypassed or retrained with faulty data to allow for unethical content to be entered and generated.

Ethical Considerations

Carbon Output

A unique ethical consideration is the high energy use and resources required to create, maintain, and update Large Language Models. It is reported that each query to OpenAI's GPT-4 emits 4.32g of carbon dioxide, with every 16 queries being equivalent to the energy used to boil a kettle highlighting the substantial compute power required (Wong, 2023).

The initial training phase of these models is highly resource intensive. An instance of this was the training of GPT-3 which involved Nvidia A100s running for ninety to one hundred days. MIT Technology Review estimated this generated 500 metric tons of carbon dioxide. Meanwhile, Meta is reported to have emitted 539 metric tons to produce multiple versions of Llama2 ranging from 2 billion to 70 billion parameters (Wong, 2023).

However, this carbon output is only expected to increase with the addition of more training data, as evidenced by GPT-3.5 having 154 billion parameters and ChatGPT-4 having 1.76 trillion parameters, with no clear data on the carbon impact of this cost.

Solutions to this issue have been put forward including using more efficient and smaller LLMs for tasks that don't require extensive data, improvements in hardware could support more compute power for less energy, thereby reducing the overall costs. Additionally new optimizations in algorithms that reduce the compute cost for each generation (Strubell, et al., 2019).

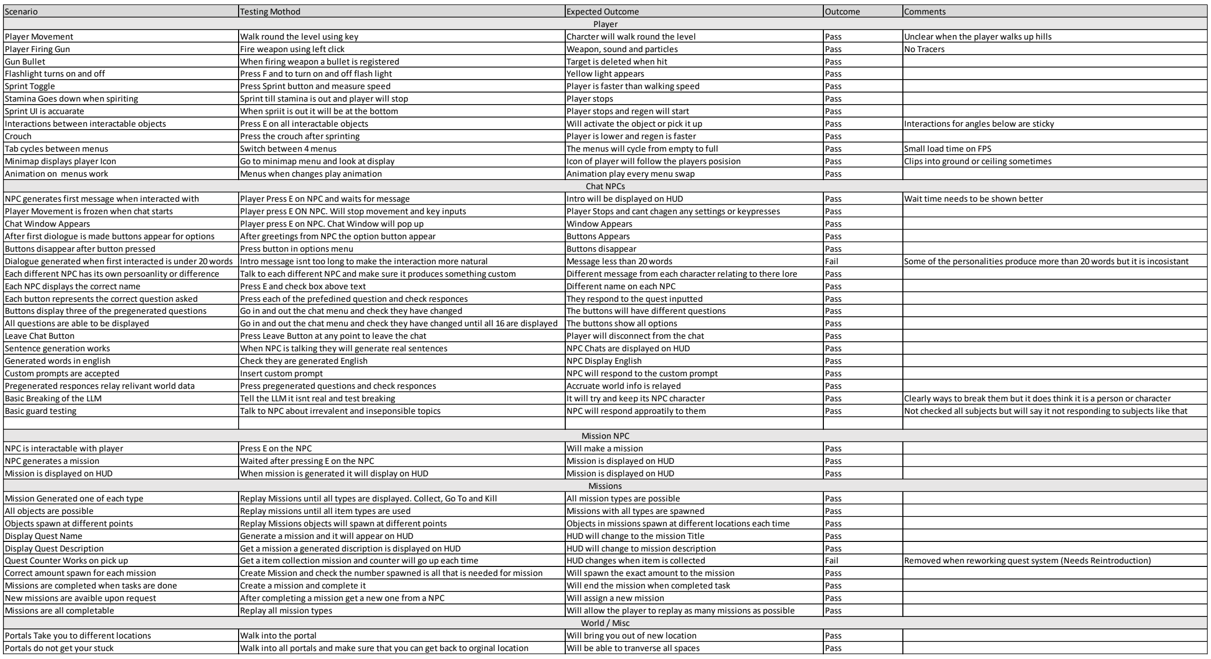
Bias and Representation

Large Language Models have been predominantly trained on Western bias and the English language which means translations or genres outside of Western scopes may have increased mistakes in outputting narrative (Wenlong Sun, 2020). This also will translate to dialects or accents that the data isn't trained on therefore it uses a very standard dialect. This translates to potentially little diversity and representation in games which in a modern game is rare as developers want to encourage representation. This is exacerbated when using a smaller LLM as they have fewer data points on each culture, and some aren't trained at all on other languages. The solution to this is to train the data on many varieties of cultures dialects and languages therefore the model will understand and be representative of all people.

Section 3 : Testing

This project will focus on black-box testing to ensure that all intended functionality works and will gather user feedback to evaluate the success of implementing a large language model into the game. The black-box testing will examine the core functionality of the game, reporting any bugs or errors for potential fixes. The black-box testing will be conducted on the version of the build intended for user testing, ensuring consistency in the data. Any features added after this point will not be included in this phase. The results from user testing will be discussed and analysed at a later stage in the documentation.

Blackbox Testing



Scenario	Testing Mothod	Expected Outcome	Outcome	Comments
		Player		
Player Movement	Walk round the level using key	Charcter will walk round the level	Pass	Unclear when the player walks up hills
Player Firing Gun	Fire weapon using left click	Weapon, sound and particles	Pass	No Tracers
Gun Bullet	When firing weapon a bullet is registered	Target is deleted when hit	Pass	
Flashlight turns on and off	Press F and to turn on and off flash light	Yellow light appears	Pass	
Sprint Toggle	Press Sprint button and measure speed	Player is faster than walking speed	Pass	
Stamina Goes down when spiriting	Sprint till stamina is out and player will stop	Player stops	Pass	
Sprint UI is accuarate	When spriit is out it will be at the bottom	Player stops and regen will start	Pass	
Interactions between interactable objects	Press E on all interactable objects	Will activate the object or pick it up	Pass	Interactions for angles below are sticky
Crouch	Press the crouch after sprinting	Player is lower and regen is faster	Pass	
Tab cycles between menus	Switch between 4 menus	The menus will cycle from empty to full	Pass	Small load time on FPS
Minimap displays player Icon	Go to minimap menu and look at display	Icon of player will follow the players posision	Pass	Clips into ground or ceiling sometimes
Animation on menus work	Menus when changes play animation	Animation play every menu swap	Pass	
		Chat NPCs		
NPC generates first message when interacted with	Player Press E on NPC and waits for message	Intro will be displayed on HUD	Pass	Wait time needs to be shown better
Player Movement is frozen when chat starts	Player press E ON NPC. Will stop movement and key inputs	Player Stops and cant chagen any settings or keypresses	Pass	
Chat Window Appears	Player press E on NPC. Chat Window will pop up	Window Appears	Pass	
After first diologue is made buttons appear for options	After greetings from NPC the option button appear	Buttons Appears	Pass	
Buttons disappear after button pressed	Press button in options menu	Buttons disappear	Pass	
Dialogue generated when first interacted is under 20 words	Intro message isnt too long to make the interaction more natural	Message less than 20 words	Fail	Some of the personalities produce more than 20 words but it is incosistant
Each different NPC has its own persoanlity or difference	Talk to each different NPC and make sure it produces something custom	Different message from each character relating to there lore	Pass	
Each NPC displays the correct name	Press E and check box above text	Different name on each NPC	Pass	
Each button represents the correct question asked	Press each of the prefedined question and check responces	They respond to the quest inputted	Pass	
Buttons display three of the pregenerated questions	Go in and out the chat menu and check they have changed	The buttons will have different questions	Pass	
All questions are able to be displayed	Go in and out the chat menu and check they have changed until all 16 are displayed	The buttons show all options	Pass	
Leave Chat Button	Press Leave Button at any point to leave the chat	Player will disconnect from the chat	Pass	
Sentence generation works	When NPC is talking they will generate real sentences	NPC Chats are displayed on HUD	Pass	
Generated words in english	Check they are generated English	NPC Display English	Pass	
Custom prompts are accepted	Insert custom prompt	NPC will respond to the custom prompt	Pass	
Pregenerated responces relay relivant world data	Press pregenerated questions and check responces	Accruate world info is relayed	Pass	
Basic Breaking of the LLM	Tell the LLM it isnt real and test breaking	It will try and keep its NPC character	Pass	Clearly ways to break them but it does think it is a person or character
Basic guard testing	Talk to NPC about irrevalent and inseponsible topics	NPC will respond approatily to them	Pass	Not checked all subjects but will say it not responding to subjects like that
		Mission NPC		
NPC is interactable with player	Press E on the NPC	Will make a mission	Pass	
NPC generates a mission	Waited after pressing E on the NPC	Mission is displayed on HUD	Pass	
Mission is displayed on HUD	When mission is generated it will display on HUD	Mission is displayed on HUD	Pass	
		Missions		
Mission Generated one of each type	Replay Missions until all types are displayed. Collect, Go To and Kill	All mission types are possible	Pass	
All objects are possible	Replay missions until all item types are used	Missions with all types are spawned	Pass	
Objects spawn at different points	Replay Missions objects will spawn at different points	Objects in missions spawn at different locations each time	Pass	
Display Quest Name	Generate a mission and it will appear on HUD	HUD will change to the mission Title	Pass	
Display Quest Description	Get a mission a generated discription is displayed on HUD	HUD will change to mission description	Pass	
Quest Counter Works on pick up	Get a item collection mission and counter will go up each time	HUD changes when item is collected	Fail	Removed when reworking quest system (Needs Reintroduction)
Correct amount spawn for each mission	Create Mission and check the number spawned is all that is needed for mission	Will spawn the exact amount to the mission	Pass	
Missions are completed when tasks are done	Create a mission and complete it	Will end the mission when completed task	Pass	
New missions are avaible upon request	After completing a mission get a new one from a NPC	Will assign a new mission	Pass	
Missions are all completable	Replay all mission types	Will allow the player to replay as many missions as possible	Pass	
		World / Misc		
Portals Take you to different locations	Walk into the portal	Will bring you out of new location	Pass	
Portals do not get your stuck	Walk into all portals and make sure that you can get back to orginal location	Will be able to tranverse all spaces	Pass	

User Testing Results

The survey conducted involved 7 participants and contained 19 questions about the game's prototype. The questions were divided into three categories: narrative, missions, and general questions about the game. Many questions were open-ended to gather opinions and thoughts on the technology, while some were designed to produce numerical data. Additionally, data on the users' PC specifications were collected to understand how the game performed on different systems and how that affected the Large Language Model.

Regarding bugs and sentence structure, Question 4 highlighted several issues, including extra punctuation and unnecessary lines added by the LLM. This bug was encountered during development, and some participants noticed it regularly. Another issue involved NPCs exceeding the text box limits, causing players to miss parts of the dialogue. This appeared to be due to the LLM not following instructions set. There were also instances where the LLM would not start or respond, which could have resulted from various factors requiring code logs to debug.

When asked about the personality types, many participants noticed distinct differences among the characters, with people able to identify four or five personality types among the ten AIs they interacted with. This observation was supported by high ratings in Question 5, which asked how pronounced the characters' personalities were.

Question 6, which asked how believable the characters were as NPCs, leaned towards the "Somewhat" category, with many participants encountering dialogue that an NPC wouldn't typically use. However, one participant voted "Yes," indicating that they could see these characters as NPCs in a video, with no one saying "No," which is a positive outcome.

When asked about changes to align the NPCs with traditional ones, participants mostly suggested reducing extraneous text generated by the LLM. This feedback highlighted a clear difference between prewritten dialogue and LLM-generated text, which can sometimes be unverified or unchecked. One person also mentioned that the characters needed to be more accurate and provide more details, suggesting the LLM hadn't been sufficiently "gamified," a valid point that requires further exploration.

Regarding response speed, participants generally rated it in the middle range, with most finding it adequate but slower than a typical modern video game. This data was then compared to PC specifications, generally on the higher end, indicating that higher-end PCs performed better and influenced response speed.

When asked about changes in responses when starting a quest and interacting with an NPC, all participants noted no significant differences. This inconsistency might be due to backend factors that require further analysis and better implementation.

Participants were also asked if custom prompts could break the character's personality, revealing that the AI could be jailbroken or deviate from its intended behaviour. This was evident in extreme cases where the AI responded with potentially dangerous or harmful scenarios, indicating a need for further investigation. Llama, the LLM used, has set bounds, but these were frequently broken.

Regarding immersion, participants noted that NPC dialogue was vague and not helpful to the plot or quest, suggesting the LLM needs more data or a better structure to understand its role. The mission system, however, showed positive signs of responsiveness, with many participants not noticing that the missions were generated by the LLM, indicating its potential for expansion.

The survey's comments section mentioned a few areas needing improvement, such as the lack of a sensitivity slider and texture issues. However, the core concept received positive feedback, with some participants calling it "cute." Other comments highlighted the demanding nature of locally run LLMs, leading to FPS drops and lag, which were also noted.

Section 4 : Analysis

User testing provided valuable feedback on the systems implemented, with many issues and bugs identified during the development process. However, these would need to be addressed in greater depth for a full release of this technology. Many participants noted the high hardware requirements. During this testing, a variety of hardware was used, but it became clear that, at this stage, the technology or its implementation is not viable for a current Triple-A game, as the impact is too significant. The high minimum requirements would also exclude a large percentage of the market. This problem was addressed by using server-based processing, but the delay between user input and response might not meet modern gaming standards, suggesting an area for further investigation.

Since the time each user spent with the artifact wasn't tracked, it's challenging to determine the exact amount of interaction, or the specific inputs provided. However, it's clear that a more personalized approach could be achieved with in-person user testing or with additional guidance. The data collected was generally fair and yielded a mostly positive response, with most participants recognizing that the technology is still in its infancy. Problems with the LLM might be due to the type of model used, with participants reporting sentence structure issues and erroneous punctuation. These could be resolved with larger and newer models, as the system can be updated with newer models using the GUFF format.

The generation of missions, though basic, worked well, with users not reporting any glaring issues. This supports the potential use of these systems in games like RPGs, where players might want a variety of missions to complete. However, due to the lack of control over the content, it's unlikely that this technology will be used in story-driven games or to control the game's plot. To improve this, more types of missions and their content could be expanded and tested to collect more data for a larger system.

Section 5: Conclusion

These projects' objectives were to research, find, and produce a vertical slice that had LLM integration and use it for narrative development, as well as reflect on issues found by player testing and black-box testing. The implementation of this project was more complex than originally thought, with limited research into integration with game engines when starting. Therefore, a large amount of collaboration between people researching this was required. This project helped facilitate collaboration with other developers in the same field, and this open-source approach aided the development of this technology.

The key findings and takeaway from this project are the significant potential and opportunity for this technology, with key industry players now working on it as well. However, some bugs need to be addressed, such as character-breaking and extraneous punctuation in some scenarios. Additionally, hardware limitations also restrict LLM capabilities, as smaller models need to be used, which can cause more bugs. In the future, if LLMs reduce in size or personal computing becomes more powerful, this technology could be used in a shipped title.

Using external API LLMs is more viable today but has several limitations, including the cost of maintenance and increased response times. Using LLMs for a publishable game is possible today, but it depends on whether a game studio wants to invest the time in developing these features. Future investigations may be required to determine the best prompts to obtain the best responses.

A key ethical problem with current LLMs is that they can break safety nets and produce explicit content. This issue emerged during the investigation, indicating that a higher level of testing and model creation will be required if this technology is to enter the market.

The benefits of this technology entering games are significant, with the potential for fully customizable stories and experiences and a potential reduction in development time. The AI revolution is happening, and it's time for the gaming industry to adopt and contribute.

Bibliography

Ammanabrolu, P. et al., 2020. *Toward Automated Quest Generation in Text-Adventure Games,* Atlanta: School of Interactive Computing Georgia Institute of Technology.

Duarte, F., 2024. *Number of ChatGPT Users (Apr 2024).* [Online]
Available at: https://explodingtopics.com/blog/chatgpt-users
[Accessed 22 04 2024].

Frans, K., 2020. *AI Charades: Language Models as Interactive Game Enviroments.* s.l.:Cross Labs.

Hakan Inan, K. U. J. C. R. R. K. I. Y. M. M. T. Q. H. B. F. D. T. M. K., 2023. *GenAI : Llama Guard: LLM-based Input-Output.* [Online]
Available at: https://ai.meta.com/research/publications/llama-guard-llm-based-input-output-safeguard-for-human-ai-conversations/
[Accessed 22 04 2024].

Kalbiyev, A., 2022. *Affective Dialogue Generation for Video Games.* [Online]
Available at: https://essay.utwente.nl/89325/1/Kalbiyev_MA_EEMCS.pdf
[Accessed 22 04 2024].

Makridakis, S., 2017. The forthcoming Artifical Intelligence (AI) reveoluation : Its impact on society and firms. *Futures,* Volume 90, pp. 46 - 60 .

Mishra, M. K., 2023. *Generating Video Game,* s.l.: University of Twente.

Murray, J. T., Murray, J. & Salter, A., 2023. Playing with AI Chat: Positioning “Dangerous” Language Model Futures through Interactive Fiction. *Proceedings of the 41st ACM International Conference on Design of Communication,* 1(1), pp. 82 - 88.

OpenAI, 2024. *Deprecations.* [Online]
Available at: https://platform.openai.com/docs/deprecations
[Accessed 22 04 2024].

Roberts, J., Banburski-Fahey, A. & Lanier, J., 2022. *Steps towards prompt-based creation of virtual worlds,* s.l.: Microsoft.

Roumeliotis, K., Tselikas, N. & Nasiopoulos, D., 2023. *Llama 2: Early Adopters' Utilization of Meta's New Open-Source Pretrained Model.* [Online]
Available at: https://doi.org/10.20944/preprints202307.2142.v2
[Accessed 22 04 2024].

Strubell, E., Ganesh, A. & McCallum, A., 2019. *Energy and Policy Considerations for Deep Learning in NLP,* s.l.: College of Information and Computer Sciences.

Sudhakaran, S. et al., 2023. *MarioGPT: Open-Ended Text2Level Generation,* Copenhagen: University of Copenhagen.

Todd, G. et al., 2023. *Level Generation Through Large Language Models.* [Online]
Available at: https://arxiv.org/pdf/2302.05817.pdf
[Accessed 22 04 2024].

Värtinen, S., Hämäläinen, P. & Guckelsberger, C., 2024. Generating Role-Playing Game Quests. *IEEE TRANSACTIONS ONGAMES,* 16(1), p. 127.

Wenlong Sun, O. N. ,. S., 2020. Evolution and impact of bias in human and machine learning algorithm interaction. 15(8).

Wong, V., 2023. *Gen AI's Environmental Ledger: A Closer Look at the Carbon Footprint of ChatGPT.* [Online]
Available at: https://piktochart.com/blog/carbon-footprint-of-chatgpt/#:~:text=An%20early%20estimate%20suggested%20that,30%2C000%20units%2C%20if%20not%20more
[Accessed 22 04 2024].

Yao, S., Rao, R., Hausknecht, M. & Narasimhan, K., 2020. *Language Models for Action Generation in Text-based Games,* s.l.: Princeton University.

Appendices

Appendix A : Questionnaire

https://docs.google.com/forms/d/e/1FAIpQLSd8u29ANpctwOpJJrhpP_cykv8OiH0cXrR6BECN1An7NCxtDg/viewform?usp=sf_link