

# **A Comparative Analysis of Game Asset Creation Using Conventional and AI Methods**

**Viktoriya Markutsa**

**A Thesis  
in  
The Department  
of  
Computer Science and Software Engineering**

**Presented in Partial Fulfillment of the Requirements  
for the Degree of  
Master of Computer Science at  
Concordia University  
Montréal, Québec, Canada**

**July 2024**

**© Viktoriya Markutsa, 2024**

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Viktoriya Markutsa**

Entitled: **A Comparative Analysis of Game Asset Creation Using Conventional  
and AI Methods**

and submitted in partial fulfillment of the requirements for the degree of

**Master of Computer Science (Computer Science)**

complies with the regulations of this University and meets the accepted standards with respect to  
originality and quality.

Signed by the Final Examining Committee:

\_\_\_\_\_ Chair  
*Dr. Leila Kosseim*

\_\_\_\_\_ Examiner  
*Dr. Leila Kosseim*

\_\_\_\_\_ Examiner  
*Dr. Dhrubajyoti Goswami*

\_\_\_\_\_ Supervisor  
*Dr. Charalambos Poullis*

Approved by

\_\_\_\_\_  
Joey Paquet, Chair  
Department of Computer Science and Software Engineering

\_\_\_\_\_ 2024

\_\_\_\_\_  
Mourad Debbabi, Dean  
Faculty of Engineering and Computer Science

# Abstract

## A Comparative Analysis of Game Asset Creation Using Conventional and AI Methods

Viktoriya Markutsa

The rapidly evolving Artificial Intelligence (AI) field is having a significant impact on many industries, as well as individuals. ChatGPT and image generating applications are undoubtedly among the most well-known and widely utilized AI tools. It is interesting to explore how these AI tools may be used in video games, specifically for asset creation, and how they compare to more traditional methods.

This thesis investigates the creation of video game assets using traditional techniques and AI tools such as ChatGPT and Midjourney. A group of 34 participants, each with varying levels of experience in different techniques and asset production, were given the task of creating basic game components using both traditional and AI tools. The goal was to rank both methods according to the overall rating, the satisfaction of the end result, and the ease of use to determine which one is preferred. The sample scene was a simple 2D platformer game, similar to Mario, with which most people are familiar and most likely played at some point. The participants were involved in both the creation process of game assets, as well as their evaluation, which brings a new perspective on the matter. It was discovered that, on average, AI tools are rated higher and are simpler to use than traditional approaches. Participants were highly satisfied with the results. However, the content created in such a way may not be as creative or as tailored for the specific needs as content made by people. Using these technologies makes it difficult to maintain a consistent style or create exactly what the person envisions, as compared to producing them manually, when the artist has complete control over every step and detail. Both methodologies have value, and depending on the project's goals and available resources, one may be preferred over the other. A hybrid method, which combines AI efficiency with artist creativity, may be the best option.

# Acknowledgments

I would like to express my sincere gratitude to my supervisor, Dr. Charalambos Poullis, for his compassion, understanding, and invaluable advice that he provided throughout this journey. He was always available to answer any inquiries and provide help, and he shown understanding and patience at a time when I was dealing with personal issues. I could not have done it without his support.

I would also like to thank my family and friends for their encouragement and unwavering belief in me.

# Contents

<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>1</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Abstract . . . . .	1
1.2 Motivation . . . . .	1
1.3 Experiment . . . . .	2
1.4 Contributions . . . . .	3
1.5 Thesis Organization . . . . .	3
<b>2 Preliminaries</b>	<b>5</b>
2.1 Abstract . . . . .	5
2.2 GAN and early versions . . . . .	5
2.2.1 ImageNet and CNNs . . . . .	5
2.2.2 GANs . . . . .	6
2.3 Stable Diffusion models . . . . .	9
2.4 ChatGPT and text-to-text models . . . . .	11
2.5 State of the Art Technologies . . . . .	13
2.5.1 AI Storytellers and Generative texts . . . . .	13
2.5.2 SORA . . . . .	13
2.5.3 GigaByte and Nvidia . . . . .	14

<b>3</b>	<b>Comparative Analysis of Game Asset Creation Using Conventional and AI Methods</b>	<b>15</b>
3.1	Abstract . . . . .	16
3.2	Introduction . . . . .	16
3.3	Related Work . . . . .	17
3.3.1	ChatGPT . . . . .	18
3.3.2	AI-Driven Asset and Content Creation . . . . .	19
3.3.3	AI in Game Development . . . . .	20
3.3.4	3D And Animation . . . . .	22
3.4	Application . . . . .	24
3.5	Methodology . . . . .	25
3.5.1	Experimental Setting . . . . .	25
3.5.2	Pilot Test . . . . .	26
3.5.3	Participant Demographics . . . . .	26
3.5.4	Analysis . . . . .	26
3.6	Evaluation . . . . .	26
3.6.1	Previous Knowledge . . . . .	26
3.6.2	Written Assets . . . . .	27
3.6.3	Visual Assets . . . . .	29
3.6.4	Comparison . . . . .	30
3.7	Conclusion and Future Work . . . . .	36
<b>4</b>	<b>Conclusion and Future Work</b>	<b>40</b>
	<b>Appendix A Questionnaire</b>	<b>43</b>
A.1	Questionnaire on Google forms . . . . .	43
A.2	Ethics Review Board Approval . . . . .	53
A.3	Instructions PowerPoint Group A . . . . .	54
	<b>Bibliography</b>	<b>69</b>

# List of Figures

Figure 2.1	Basic CNN Architecture (Phung & Rhee, 2019)	7
Figure 2.2	AlexNet Architecture (Krizhevsky, Sutskever, & Hinton, 2012)	7
Figure 2.3	GAN Architecture (Bengesi et al., 2023)	8
Figure 2.4	Stable Diffusion Architecture (Chollet, Wood, & Gupta, 2023)	10
Figure 2.5	ChatGPT’s Architecture (Bengesi et al., 2023)	12
Figure 3.1	Base scene with basic placeholders for assets, as presented to participants to update with the images they create themselves (b). Resulting scene of participant #28 using conventional methods such as finding available assets on the Internet and creating from scratch, (a) and using Midjourney as AI tool (c)	15
Figure 3.2	Experimental setting	24
Figure 3.3	Participants’ reported skill levels in game and asset creation	27
Figure 3.4	Skill level vs satisfaction level of written assets created by the participants manually (a) and using ChatGPT (b)	29
Figure 3.5	Ease of creation of written assets (1-very difficult, 10-very easy), by respective skill level of participants	29
Figure 3.6	Skill level vs satisfaction level of visual assets created by the participants manually (a) and using Midjourney (b)	31
Figure 3.7	Ease of creation of visual assets (1-very difficult, 10-very easy), by respective skill level of participants	31
Figure 3.8	Comparison of rating (a), satisfaction level (b), and ease of creation (c) of written assets created using different methods as reported by participants	32

Figure 3.9	Comparison of rating (a), satisfaction level (b), and ease of creation (c) of visual assets created using different methods as reported by participants . . . . .	32
Figure 3.10	Resulting scene with conventional methods (a) and AI tools (b) of participant #12 . . . . .	35
Figure 3.11	Resulting scene with conventional methods (a) and AI tools (b) of participant #25 . . . . .	36
Figure 3.12	Sample dialogue created with AI by participant #6 . . . . .	37
Figure 3.13	Sample dialogue created with conventional methods by participant #21 . . . . .	37
Figure 3.14	Sample dialogue created with AI by participant #25 . . . . .	38
Figure 3.15	Sample dialogue created with conventional methods by participant #25 . . . . .	39
Figure 3.16	Resulting scene with conventional methods (a) and AI (b) of participant #32 . . . . .	39
Figure A.1	Questionnaire part 1 . . . . .	44
Figure A.2	Questionnaire part 2 . . . . .	45
Figure A.3	Questionnaire part 3 . . . . .	46
Figure A.4	Questionnaire part 4 . . . . .	47
Figure A.5	Questionnaire part 5 . . . . .	48
Figure A.6	Questionnaire part 6 . . . . .	49
Figure A.7	Questionnaire part 7 . . . . .	50
Figure A.8	Questionnaire part 8 . . . . .	51
Figure A.9	Questionnaire part 9 . . . . .	52
Figure A.10	Certification of Ethical Acceptability . . . . .	53



# Chapter 1

## Introduction

### 1.1 Abstract

The recent AI breakthroughs have resulted in a lot of innovative and interesting tools that may be utilized for a wide range of purposes. Video games have also evolved, and any type of player may find something to their liking, whether it is a simple mobile game to pass the time, an old-school platformer, or an immersive Role Playing Game (RPG), the options are countless. However, creating one is a difficult undertaking that requires a great deal of thinking and effort, in addition to the necessity of specific skill sets. The development of visual assets and story-telling are crucial aspects of what makes games unique and engaging, thus it is worth investigating how these AI technologies can assist game developers and artists, as well as how they compare to traditional techniques.

### 1.2 Motivation

Since computers became widely available, video games have grown in popularity as a recreational and even competitive activity. The assets utilized in-game are among the characteristics that make them appealing and unique: Captivating story telling, fleshed out dialogues and characters, compelling visuals of the locations, items and characters. Moreover, given that the majority of the population has access to smartphones, mobile gaming has grown in popularity, and even

non-gamers have most certainly engaged with some of them. In fact, according to Entertainment Software Association (ESA), in 2021, over 70% of Americans have at some point played video games (Cerezo-Pizarro, Revuelta-Domínguez, Guerra-Antequera, & Melo-Sánchez, 2023). They are now part of our culture and people of all backgrounds and generations can find something for their taste. They are a source of entertainment, however they may also help develop communication skills, problem-solving abilities, and critical thinking, among others (Duggan, 2015).

Traditionally, asset production necessitates a specific set of abilities and a team of people to create an appealing and high-quality game. Of course it is possible to create one as a one man team, but a single individual does not always have the same high level of expertise in all areas of development, story-telling, visual asset creation, voice acting, audio engineering, and so on. There is always a tradeoff between the quality and the available resources, in terms of both time and money. In recent years, there has been an emergence of AI-based tools such as ChatGPT for text-to-text generation, as well as image-generating programs such as Dall-E and Midjourney, which are available to the public for free or at a reasonable cost. These tools serve a number of purposes, both personal and professional. ChatGPT is a text-to-text model that may be used as a personal assistant for creating code, writing emails, postings, captions, and essays, among other things. Text-to-image models such as Midjourney create highly detailed images based on the user's prompt in mere minutes.

It is interesting to explore these tools specifically for video game asset creation, to determine how efficient they are compared to conventional approaches and identify any limitations. They could potentially be very time and cost effective, which is appealing to video game producers.

### **1.3 Experiment**

Previous relevant works about AI use in games were important to investigate. The capabilities of AI to generate text or images have been explored by other researchers and offered valuable information, however none used people from diverse background to experiment with it within the context of a project.

For this experiment, ChatGPT-3.5 and Midjourney were selected, and 34 people were asked to

create assets for a sample game scene using the aforementioned AI tools as well as conventional methods. The scene consisted of three dialogues with a progressive story, unless the participant had another vision that could be easily incorporated. Dialogues were developed both manually and with ChatGPT. For the visual assets, they were asked to create sprites for the main character, the three NPCs (or items/places) and some background and tiles if time permitted.

Although AI tools were generally easier and faster to use, end-user satisfaction varied. ChatGPT had mostly positive feedback, whilst Midjourney received mixed reviews. It produced high-quality images, but some of the most common concerns were about the irregularity of the style, the AI's inability to create exactly what the user desired, as well as the soullessness and lack of creativity and emotional attachment of the created assets. Time constraints, the requirement for participants to come up with ideas on the spot, and restrictions on the setting and scripts accessible were some of the project's limitations.

## **1.4 Contributions**

The findings of the research are valuable as they not only assess the objective advantages and disadvantages of using conventional methods or AI tools, but also subjective ones, which are just as important because, in the end, both the creators and the users are humans. What sets this research apart is the fact that participants are both involved in the creative process of the assets, as well as the evaluation.

## **1.5 Thesis Organization**

Chapter 1 summarizes the present state of video games and AI, emphasizing its relevance for game asset generation. A comprehensive literature review on the evolution of generative models including Convolutional Neural Networks, Generative Adversarial Networks, Stable diffusion and state-of-the-art technologies that are relevant in video game asset creation are presented in Chapter 2. It is critical to understand the history of modern technologies, how they came to be, and how they continue to evolve at such a rapid pace. Chapter 3 will deep dive into the experiment itself, where the participants were asked to create video game assets for a sample 2D platformer scene

using conventional methods as well as AI tools, specifically ChatGPT-3.5 and Midjourney. The goal was to assess the current AI capabilities for generating video game assets as compared to more traditional methods, as well as record the participants' experiences using various parameters such as ease of use, satisfaction level, and any other subjective remarks they may have. Finally, Chapter 4 summarizes all the previous information, as well as gives an overview of other relevant topics to explore to further this research.

With the research goal set, the first step is to explore the origins of AI, to understand what led to its development and how these previous technologies expanded. It will also help understand how and why they weren't as used in video games at the time and what makes them a popular topic presently.

# Chapter 2

## Preliminaries

### 2.1 Abstract

To understand the current state of AI, it is necessary to recall the path that led to its development: from the earliest generative models and Convolutional Neural Networks (CNNs) to Generative Adversarial Networks (GANs) and, finally, stable diffusion models. Most are used to generate images based on the user input or other images. In terms of text-to-text generation, ChatGPT is the most well-known Natural Language Processing (NLP) technology, with a wide range of applications due to its human-like responses and ability to interpret prompts. Most recent advancements include video-generating applications such as SORA ([OpenAI, 2024](#)). As with any new technologies, it is crucial to assess their influence on society and individuals, as well as all of the ways they will transform our lives, to ensure they cause no harm and adhere to a stringent ethical code.

### 2.2 GAN and early versions

#### 2.2.1 ImageNet and CNNs

The first generative models in computer vision can be traced back to 1950s ([Raut & Singh, 2024](#)), and the development of CNNs to the 1980s ([Bengesi et al., 2023](#)), but the launch of the ImageNet competition in 2010 marked real progress. It provided the largest database of over 15 million high-resolution, labeled images from 22 000 distinct categories ([Krizhevsky et al., 2012](#)),

allowing different models to be trained and CNNs became of particular interest for image classification (Bengesi et al., 2023). AlexNet marked a breakthrough in 2012 after winning the ImageNet competition and setting a new standard for CNN algorithms (Krizhevsky et al., 2012). Since then, CNN architecture has evolved into different models such as ResNet (He, Zhang, Ren, & Sun, 2016), DenseNet (Huang, Liu, van der Maaten, & Weinberger, 2017), MobileNet (Howard et al., 2017), and EfficientNet (Tan & Le, 2019) (Bengesi et al., 2023). However, all of the above models are mainly descriptive, with the main purpose of illustrating existing data patterns and predicting outcomes based on the information available (Bengesi et al., 2023).

A basic CNN consists of two main components: a feature extraction layer, which is made up of pooling layers that separate and identify the features of an image, and a fully connected layer that predicts the image's class based on feedback from previous layers. The first layer of feature extraction is a convolution layer and is used to extract data from the input using the dot product with a  $M \times M$  filter. The second layer is a pooling layer which main function is to reduce the size of the convoluted feature map obtained from the previous step, in order to decrease the computational costs. The fully connected layer consists of weights and biases and is fed the flattened input image from the previous layers. OA dropout layer is frequently used to reduce the risk of overfitting. Each layer transforms the image into a more abstract representation, with the first layers identifying edges, colors, and textures, and the deeper layers more complex structures. To approximate the relationship, an activation function like as ReLU, Softmax, tanH, and Sigmoid is required (Gurucharan, 2023). A simple CNN is illustrated in 2.1. Different models feature varying numbers of layers, hyperparameters and activation functions, among others. For contrast, AlexNet CNN architecture is depicted in 2.2. It consists of eight learnt layers (five convolutional and three fully connected), with ReLU activation function and Softmax on the last fully connected layer.

## 2.2.2 GANs

In 2014, GANs were introduced by Goodfellow et al (Goodfellow et al., 2020), revolutionizing the state of AI. Unlike CNNs, GANs are able to create new data, whether it is text, images, or audio, based on training datasets that resemble the original data. For classification problems, supervised learning is used, which requires a large amount of human-labeled data, with example inputs and

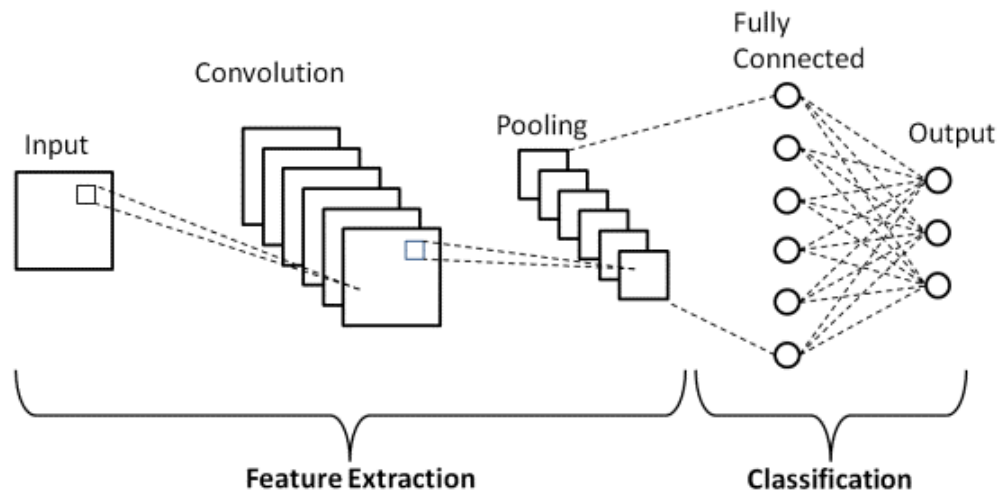


Figure 2.1: Basic CNN Architecture (Phung & Rhee, 2019)

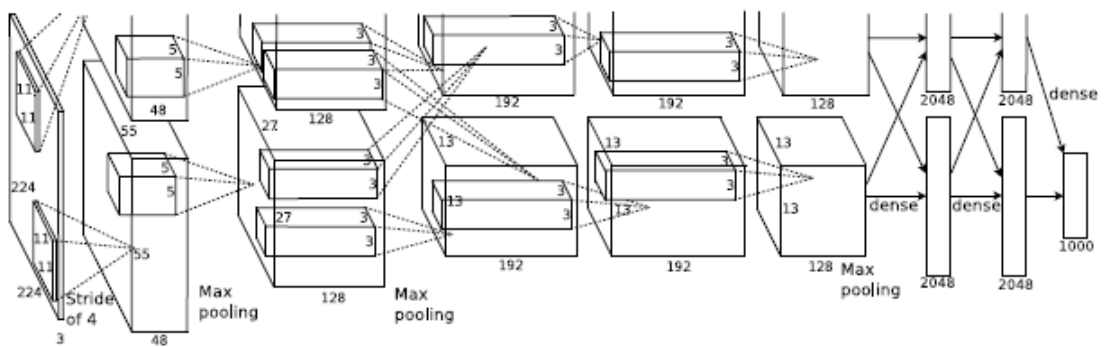


Figure 2.2: AlexNet Architecture (Krizhevsky et al., 2012)

outputs. To outperform humans, the model would require millions of such data points, which would be very resource consuming. Generative models, such as GANs, are trained via unsupervised learning, which requires less human supervision and fewer data (Goodfellow et al., 2020). The data is unlabeled, so the model must analyze it and identify patterns. Given a set of examples  $x$  and an unknown distribution  $p_{data}(x)$ , the goal of the model is to create an algorithm  $p_{model}$  that matches it as closely as feasible. The most common approach is the Maximum Likelihood Estimation, and consists of minimizing the Kullback-Leibler divergence between  $p_{data}$  and  $p_{model}$  (Goodfellow et al., 2020).

A GAN consists of two neural networks: the generator and the discriminator. The generator, as the name suggests, generates new data that closely mimics real data. The discriminator tries to determine whether that new data is real or synthetic (Bengesi et al., 2023). The generator does not have access to any real images and receives feedback only from the discriminator, who has access to both real and synthetic data. Noise is used by the generator to create new data, which is then sent into the discriminator, which uses a sigmoid activation function and binary cross-entropy loss to determine whether or not the data is fake. Backpropagation then takes place to optimize the training process until the fake data is almost indistinguishible from the real one by the discriminator (Creswell et al., 2018). This process is illustrated in Figure 2.3.

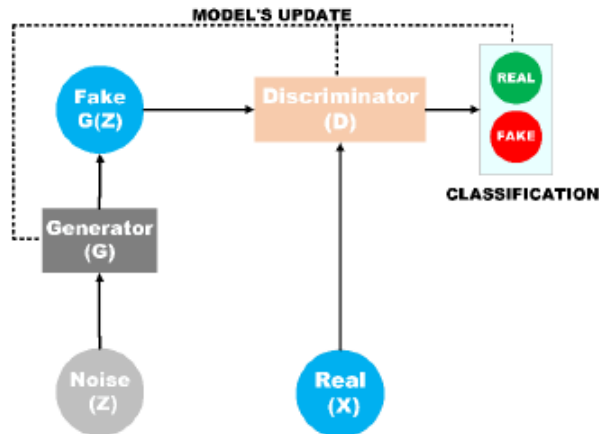


Figure 2.3: GAN Architecture (Bengesi et al., 2023)

However, they also suffered from a few drawbacks, such as the risk of mode collapsing, when the generator can only produce one type of output (Bhagyashree, Kushwaha, & Nandi, 2020). This



can happen either because of the discriminator overfitting or catastrophic forgetting, where learning a new task destroys any knowledge of the previous one (Bengesi et al., 2023). Another issue that may occur is non-convergence and instability. GANs are particularly sensitive to any change in hyperparameters, making it more challenging to ensure training consistency as it has a significant impact on the results (Bengesi et al., 2023). If the discriminator happens to learn too quickly in distinguishing between real and synthetic data, the generator may have difficulties keeping up and will not improve based on the feedback provided, slowing or stopping the learning process. This phenomenon is known as gradient vanishing.

To address these limitations, different variations of GANs were developed, including Conditional Generative Adversarial Network (cGAN), Deep Convolutional GAN (DCGAN) and Wasserstein GAN (WGAN) (Bengesi et al., 2023). WGAN solves the issue of gradient vanishing; cGAN adds a layer with extra information about the class of the image; and DCGAN adds a deep learning model for the discriminator and the generator, which enhances the model's stability (Radford, Metz, & Chintala, 2016).

## 2.3 Stable Diffusion models

The next major step after GANs was the proposal of Stable Diffusion models. They address some of the previously discussed disadvantages GANs. Stable Diffusion models consist of three parts: a text encoder, a diffusion model, and a decoder, represented in Figure 2.4. The diffusion model itself is a series of denoising autoencoders designed to learn a data distribution by gradually denoising a normally distributed variable (Rombach, Blattmann, Lorenz, Esser, & Ommer, 2022). This is a reverse Markov chain: denoising where every step depends only on the previous step. By repeatedly denoising a pure noise image, it is capable of generating remarkable new synthetic images. The more it loops over the denoising (default being 50), the crispier the output image will be (Chollet et al., 2023). The Latent Diffusion Model was proposed shortly after, in which a diffusion model is trained in a latent space, resulting in better scaling and the need for only one network pass to obtain a highly-detailed image (Rombach et al., 2022). Another advantage is that the autoencoding stage only needs to be trained once, allowing the model to be reused for many

image-to-image and text-to-image applications. The text encoder transforms the user prompt into a latent vector, that is fed into the diffusion model alongside the randomly generated noise patch. The diffusion model denoises the patch, and the decoder converts the produced image into a high-resolution picture.

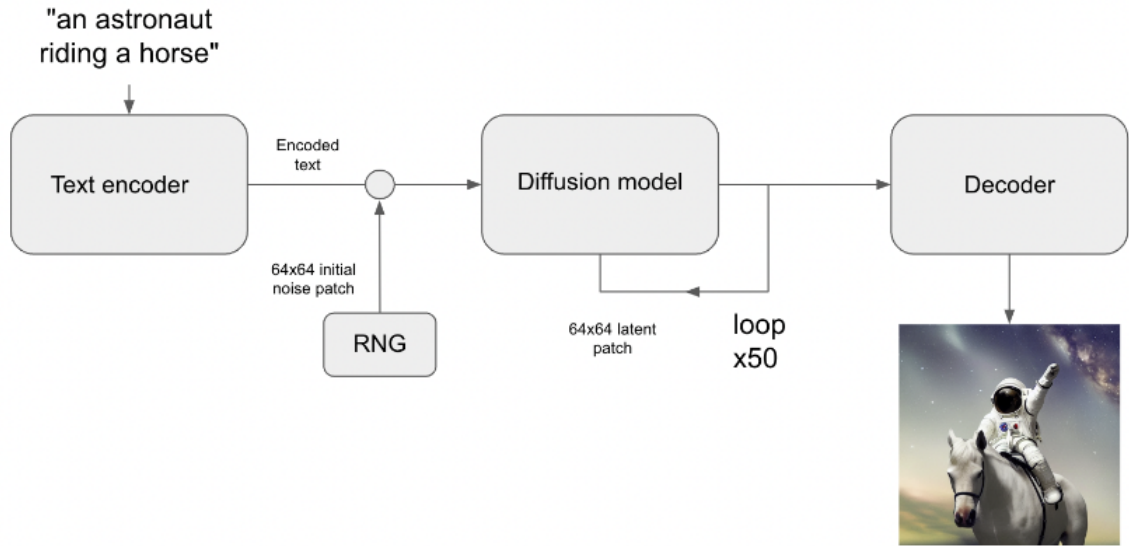


Figure 2.4: Stable Diffusion Architecture (Chollet et al., 2023)

However, the model is not without its drawbacks. Generating high-quality samples requires a lot of processing power and memory. They also need larger datasets for training (Vahdat & Kreis, 2023). As they are trained on data created by humans, they can also potentially perpetrate bias in the images they generate and infringe on copyright, so users must be mindful of how they use it. The generated images amplify the cultural and societal prejudices, depending on the type of data used for training (Luccioni et al., 2023). For example, requesting Dall-E 2 to generate managers or CEOs yields images of predominantly white men wearing suits and ties, which lacks diversity. As it is a fairly new technology, no solution was found, but some attempts have been made to correct the bias. Some outputs are reviewed by people, and then some weights can be reassigned within the model to counteract them. However, there are no established benchmarks for what is ethical for generative AI yet. It is also possible to enhance the models to have an "ideal" representation of the society, however the "ideal" may vary by person, based on the cultural and societal factors

(Luccioni et al., 2023). Some training data also contains artists' work, and the generated images can imitate their style extremely accurately, resulting in copyright violation. To safeguard the creators, a few frameworks were proposed. CopyScope is implemented at the model level and quantifies infringement in it (Zhou, Gao, Wang, & Wei, 2023). It can be quite valuable in assisting users in creating legally compliant artwork while protecting original content creators. DiffusionShield is a system that creates uniform patterns for watermarks and flags any generated images based on copyrighted material (Cui et al., 2023). Unfortunately, it will not solve all issues since most small internet companies are unregulated, and unethical ones frequently sell as much merchandise as possible before disappearing due to a large number of complaints and reappearing under a different name. They might employ AI-generated images that infringe copyrights, sell as many items as they can and disappear. Schemes like these are already common and will continue to exist.

## 2.4 ChatGPT and text-to-text models

ChatGPT is a Generative Pretrained Transformer (GPT), a Large Language Model (LLM) that is capable of performing diverse NLP tasks (“Understanding the ChatGPT Architecture: A Comprehensive Guide”, 2023). These tasks include generating human-like text, translating, answering questions, and generating code, among others. GPT training consists of two phases: language modeling, a training phase during which the model attempts to predict the next word in a sentence, and fine tuning, during which it is trained on a smaller dataset for a specific purpose (Dsouza, 2024). ChatGPT can serve as a personal assistant, as well as help with work related tasks. It was first introduced by OpenAI in 2018 and the model consisted of two stages, the first being the language understanding component and the second being response generation (Bengesi et al., 2023) (Dsouza, 2024). The model analyzes the semantics and syntax of the user's text input to determine the most appropriate response based on the message's fluency and relevance (Dsouza, 2024). This process is represented in Figure 2.5. The original ChatGPT was trained on BooksCorpus dataset, which contains over 7000 distinct unpublished books in different genres (Bengesi et al., 2023). Every subsequent iteration of ChatGPT improved its ability to grasp and output useful results. The training dataset now comprises millions of webpages, books, and other media; it can do speech recognition

and handle increasingly difficult tasks. (Bengesi et al., 2023). ChatGPT-4, released in March 2023, can take not only text but images as input, and is capable of performing complex tasks comparable to those performed by humans, such as performing well on a bar or medical test. The model was trained using publically available data and fine-tuned with Reinforcement Learning from Human Feedback (RLHF) (Bengesi et al., 2023).

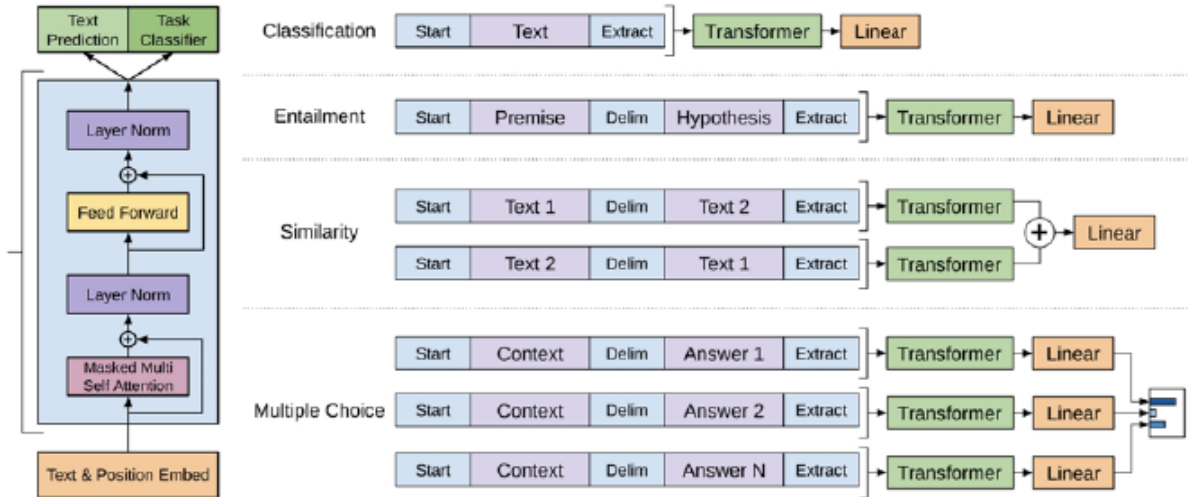


Figure 2.5: ChatGPT’s Architecture (Bengesi et al., 2023)

Although each iteration significantly enhances the model’s features, there are still some issues that need to be considered. Since the data that is fed to the model is created by humans, it is inherently biased, and the responses created by ChatGPT are equally skewed (Dsouza, 2024). It does not have access to the most recent data, and it may provide incorrect responses; therefore, if the information sought is critical, it should be double-checked. There are also security concerns as the users can expose personal information or use it with malicious intentions. It also lacks emotional intelligence and empathy.

## 2.5 State of the Art Technologies

### 2.5.1 AI Storytellers and Generative texts

Several games are experimenting with integrating LLMs as part of storytelling or Non Playable Characters' (NPCs) dialogues and behaviours (Will & David, 2024). Generative agents are agents that simulate believable human behaviour, such as the NPCs in games like Sims. Using LLM-powered AI characters, it is possible to get more realistic behaviour from them, such as lifelike conversations, and novel behaviours between the NPCs emerge rather than being preprogrammed (Park et al., 2023). The characters form new relationships, partake in common activities and develop their personalities. There are still some ethical concerns to consider, similar to the ones already mentioned of possible bias perpetration and misuse, there is also the possibility of the user's emotional attachment to characters, which can be mitigated by clearly stating that it is only an agent and ensuring that the models have proper values and will not be inappropriate for a given context, such as reciprocating love.

### 2.5.2 SORA

Another interesting model is the text-to-video model SORA, developed by OpenAI (OpenAI, 2024). It is able to build complex scenes based on the user's input, which can be an interesting path to explore for video games as it could potentially make the creation of cutscenes (non-interactive video sequence in a game) faster and cheaper. SORA produces detailed movies in various formats, lengths, and quality (OpenAI, 2024). The videos are incredibly realistic, even given silly and magical scenarios, in which case the interactions will flow naturally and realistically within the context of the prompt. It accepts text, images, and videos as input, making it ideal for photo and video editing, gif creation, and other tasks (OpenAI, 2024). However, this also raises the question of whether it was trained on copyrighted data, as well as how it would impact human creators (Turner Lee & White, 2024).

### 2.5.3 GigaByte and Nvidia

With all of the rapid improvements in AI, hardware is a significant consideration. All of these generative models require a significant amount of resources, which not all modern computers and servers can handle. In January 2024, GigaByte Technology announced their new hardware innovations that will allow users to achieve the extensive computation needed by modern generative AI models, while fulfilling sustainable goals through its thermal design (GIGABYTE, 2024). It is adaptable and scalable, with reliable performance in severe environments. Their cutting-edge servers are adaptive to fast shifting trends, and they provide both gaming and creator computers that take this into consideration, while also incorporating latest Nvidia AI. (GIGABYTE, 2024).

During the Consumer Technology Association (CES) of 2024, Nvidia presented a few exciting projects involving generative AI, such as Nvidia ACE, which brings digital avatars to life and may be used to revolutionize interactions between players and NPCs (Caulfield, 2024). It uses speech recognition from the player's voice input, after which an LLM generates the character's response, which is then passed to an animation model to render a realistic lip sync. They also introduced new GPU series capable of rendering the most graphically intensive games while utilizing a fraction of the power and remaining affordable. Creators will also be able to generate stable diffusion videos 1.5 times faster and images 1.7 times faster (Caulfield, 2024). Some video games are very beloved within the community, and mods are being relentlessly created for them. With this in mind, Nvidia launched a beta of RTX Remix, which can be used by modders to remaster older games with full ray tracing, using AI texture tools to turn low resolution textures into 4k ones (Caulfield, 2024).

The examples above are only a handful of the new technologies that use generative AI, and many more are on the way. From the current AI tools available, Midjourney and ChatGPT-3.5 were selected for this research to see how well they can be applied in video game asset creation, as compared to conventional methods.

## Chapter 3

# Comparative Analysis of Game Asset Creation Using Conventional and AI Methods

The following is a verbatim copy of the manuscript currently under review at IEEE Transactions on Games, titled "Comparative Analysis of Game Asset Creation Using Conventional and AI Methods, authored by Viktoriya Markutsa and Charalambos Poullis.

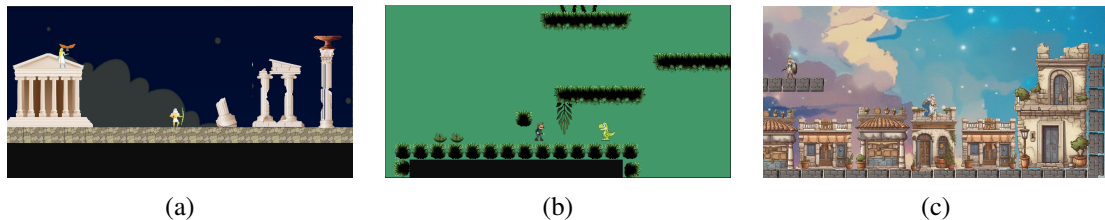


Figure 3.1: Base scene with basic placeholders for assets, as presented to participants to update with the images they create themselves (b). Resulting scene of participant #28 using conventional methods such as finding available assets on the Internet and creating from scratch, (a) and using Midjourney as AI tool (c)

### **3.1 Abstract**

This paper investigates the creation of video game assets using traditional techniques and AI tools such as ChatGPT and Midjourney. A group of 34 participants, each with varying levels of experience in different techniques and asset production, were given the task of creating basic game components using both traditional and AI tools. It was discovered that, on average, AI tools are rated higher, based on overall rating and satisfaction of the participants, and are simpler to use than traditional approaches. Participants were highly satisfied with the results. However, the content created in such a way may not be as creative or as tailored for the specific needs as content made by people. Using these technologies makes it difficult to maintain a consistent style or create exactly what the person envisions, as compared to producing them manually, when the artist has complete control over every step and detail. Both methodologies have value, and depending on the project's goals and available resources, one may be preferred over the other. A hybrid method, which combines AI efficiency with artist creativity, may be the best option.

### **3.2 Introduction**

The rapid advancement of generative AI has sparked widespread interest across various communities, catering to both personal and professional needs. With these advancements, media creators, both professionals and amateurs, now have access to powerful AI tools that facilitate the swift generation of content. This is predominantly enabled by text-to-image models and Large Language Models (LLMs), which have transformed creative workflows in numerous industries. One such industry where the potential of these tools is particularly noteworthy is game development. Here, the use of Diffusion-based applications for image generation and ChatGPT for narrative creation is especially intriguing.

Diffusion-based models such as those utilized by Midjourney and Dall-E allow for the generation of incredibly detailed images in mere minutes. These tools not only make artistic creation more accessible but also provide the capability for users to train models and fine-tune them for specific applications. This aspect opens up exciting possibilities for customized content creation that aligns



with the unique needs of different projects (Cahyadi, Rafi, Shan, Lucky, & Moniaga, 2023). However, while these innovations offer substantial benefits, they also come with their set of challenges, notably the risk of perpetuating biases and stereotypes related to gender and race. This paper aims to delve into these critical issues, assessing the implications of such biases in technology.

Similarly, ChatGPT has been shown to be a versatile tool that can produce human-like text, aiding in various tasks from brainstorming ideas to generating code and essays. Its application in writing video game narratives, dialogues, and quests showcases its potential to revolutionize content creation in gaming (Hassija, Chakrabarti, Singh, Chamola, & Sikdar, 2023). Despite the promising capabilities of these technologies, it is imperative to critically evaluate both their advantages and the ethical dilemmas they pose. Extensive research has been conducted on the potential of these technologies to generate media content and the drawbacks associated with their use, which will be explored more in detail in Section 3.3.

This study focuses specifically on the application of Midjourney and ChatGPT-3.5 in generating video game assets, comparing these modern tools with traditional methods. For the scope of this study, traditional approaches involve individuals manually creating dialogues and narratives and either drawing their own assets or sourcing them for free use on the Internet. The applications employed in this research and the experimental setup are discussed in greater detail in Sections 3.4 and 3.5, respectively. Section 3.6 presents a thorough analysis and discussion of the experimental results, highlighting key findings and their implications for the field. Finally, Section 3.7 outlines potential avenues for future research, posing questions that could further explore the expansive potential of generative AI in creative industries.

### **3.3 Related Work**

In recent years, the appearance and development of AI has greatly impacted many aspects of our lives. It had a major influence on a variety of applications, particularly video game development, whether for level design or asset generation.

### 3.3.1 ChatGPT

In the last few years, LLMs have transformed the Natural Language Processing (NLP) by showcasing an unmatched ability to interpret and generate natural and human-like language (Gaur & Saunshi, 2022). ChatGPT, the most well-known and widely used LLM, is useful in a variety of domains, including social chatbots and entertainment, language translation and comprehension, text generation and summarization, personal assistance and virtual companionship, and customer service. (Hassija et al., 2023). It is also an excellent tool for invention and productivity, as it may aid with ideas, proofreading, and writing papers or code, among other purposes (Hassija et al., 2023). It can also assist teachers and serve as a learning tool for a variety of subjects (Yu, Liu, & Guo, 2023). However, it comes with some drawbacks regarding ethical issues, data privacy and security, effectiveness and reliability (Yu et al., 2023). Content generated by ChatGPT should be validated and used with care, taking into account all the challenges it presents.

Unsurprisingly, all these characteristics make it a potentially amazing tool for generating dialogue and quests based on intricate narratives and character decisions. In “Generating Role-Playing Game Quests With GPT Language Models”, (Vartinen, Hamalainen, & Guckelsberger, 2022) fine-tuned ChatGPT-2 to create Quest-GPT-2, a specialized tool for creating RPG quests. The results have been mainly positive, with most problems being addressed by utilizing later versions of ChatGPT. For optimal results, this tool can be used as an assistant to co-create quests with the human author (Vartinen et al., 2022). Moreover, ChatGPT-3 addressed most issues brought up by participants throughout the experiment, such as lack of continuity between quests, and their illogical nature. However, with ChatGPT-4 and GPTs, it is likely that the quality of the quests and the cohesiveness will be improved even more (Vartinen et al., 2022). This shows how promising the assistance of ChatGPT in games is, and how it has the potential to bring more diversity, help with narrative composition, and generate more complex dialogues and quests that would reflect the setting of the game.

### 3.3.2 AI-Driven Asset and Content Creation

The use of Text-to-Image AI tools have been increasingly popular in the past years and they are now widely used in a variety applications. The art produced by these applications can be employed in a wide range of industries such as E-Commerce, fashion, medical imaging, architecture and criminal procedure (Chauhan, Chauhan, Nainwal, Arora, & Bhatt, 2023). The authors in (Singh, 2023) explore the existing capabilities of AI tools to generate images and videos from text and their application in game development and other media. It highlights that the output produced is high-quality, and the photorealism is only improving (Singh, 2023). However, some limitations include the lack of coherence in the videos and the need for a lot of resources otherwise (Singh, 2023). Diffusion models are used by the majority of popular text-to-image applications, including Midjourney and DALL-E-2 (Dayma & Cuenca, 2022). They currently spark more interest among the researchers and public communities than their first counterparts that used Generative Adversarial Networks (GANs) (Cahyadi et al., 2023). They also provide the possibility for individuals to train their own models, which can turn out to be an interesting feature but also poses the issue of it being unregulated and can be misused to perpetuate bias and abuse (Cahyadi et al., 2023). Current AI generative tools present a heavy bias, especially towards gender and race roles in society (Chauhan et al., 2024), therefore users should be mindful of the content they create. Key factors to consider are clear communication and expectations, trust and transparency in collaboration process, appropriate training and education for human collaborators, and ensuring AI tools align with human values and goals (Kunal, Rana, & Bansal, 2023).

Stable diffusion models can be created and finetuned to tackle various issues, which is demonstrated by (Nguyen, Sone Aung, Le, Park, & Hong, 2023), where the scarcity of relevant images was overcome by creating a custom stable diffusion model to synthesize the medical images. (Vayadande et al., 2023) present a web API that leverages DALL-E-2 but fine tune it using their own model, which further showcases the versatility and adaptation of diffusion models for a variety of specific topics. (Göring, Ramachandra Rao, Merten, & Raake, 2023a) found a correlation between appeal and realism when comparing realistic generated images, with Own and Midjourney having the best

appeal ratings and Own and DALL-E-2 having the highest realism ratings. Further research indicated that Midjourney and DALL-E-2 had the most appealing realistic images when compared to real pictures (Göring, Ramachandra Rao, Merten, & Raake, 2023b). Furthermore, due to their high resemblance to real photographs, AI generated ones are used in journalism, which raises more concerns since it is not necessarily validated content and contributes to the spread of false news and misinformation (Paik et al., 2023). To combat this side-effect, it is crucial to be able to distinguish between real and fake images, hence some models have been created to recognize AI-generated images (Hossain, Uz Zaman, & Islam, 2023). However, they are growing more sophisticated and harder to distinguish, which in addition can cause the generated images to look too similar to the original artwork, thereby breaching copyrights (Zhang, Wang, Tohidypour, & Nasiopoulos, 2023). Trained models to detect synthetic images are also getting increasingly advanced (one of the models reaching 99.46 accuracy in a case study) (Zhang et al., 2023), while humans are still having trouble recognizing a synthesized image (46.97% accuracy) (Zhang et al., 2023). Papa et al. conducted a similar experiment and constructed a model to assess if a generated human face is real or synthetic. The model had an accuracy rate of 99%, compared to 60% for humans (Papa, Faiella, Corvitto, Maiano, & Amerini, 2023). Some bias issues as well as data quantity for some subjects may be partially addressed: Zambrano and Senouci explored adding synthetic data to train image classification models based partly on synthetic datasets (Zambrano & Senouci, 2023). The results were promising, especially on mixed datasets, and could open doors for more further studies (Zambrano & Senouci, 2023).

### **3.3.3 AI in Game Development**

AI is commonly utilized in video games for a variety of purposes, whether it is during the design and creation process, or for gameplay. (Goodman, Wallat, Perez-Liebana, & Lucas, 2023) investigated a use case in which an AI play tester was used with human play testers during the design phase of a tabletop game. Although the AI could not capture all design flaws as they were due to comprehension or intuitive use, the AI was still a helpful tool to use in addition to the manual process, thus reducing costs and potentially catching some uncommon bugs that human play testers may have missed (Goodman et al., 2023) (Powley, Colton, Gaudl, Saunders, & Nelson,

2016). The authors in (Ostuni & Galante, 2021) aimed to create an AI that would be able to beat the hardest difficulty of Touhou<sup>1</sup> (a popular bullet hell game), by looking at pixels only, something no AI has been able to do so far. The models were very accurate but not enough to succeed. However, this research is still valuable since it illustrates how AI playtests are evolving, offering game designers more insights and providing players with potential opponents or showcase how to flawlessly beat a level. Powley et al. developed an AI application that generates and playtests physics-based game levels for casual phone games (Powley et al., 2016). Although it still needs refinement, it might potentially open up the doors of game development to a wider population (Powley et al., 2016). Players often have interest in creating extra content for their favorite games, with noteworthy examples being Counter-Strike<sup>2</sup> and DOTA<sup>3</sup> created from Half-Life<sup>4</sup> and Warcraft III<sup>5</sup> respectively (Duan, Huang, Zhao, Huang, & Cai, 2022), and all the modding communities for varied titles (Duan et al., 2022). Making game asset and levels development more accessible to the general public through the use of generative AI would enable users to contribute more content for their favorite games, resulting in a richer and more diversified universe.

The military, healthcare, and education are just a few of the industries interested in the developments in AI-driven image production and classification. Due to their easily expandable utility, these can also have a direct impact on how the AI tools are used when creating video games. The outpainting technique was used to construct panoramic views from existing military images to train a model to detect military objects in panoramic views (Lee, Moon, & Nam, 2023). This can be further expanded to detect objects in various distorted views such as fisheye or ultra-wide (Lee et al., 2023). It can be an interesting feature for some military or strategy-based video games, among others. Midjourney has been found to be able to generate scientifically accurate representations of tornadoes, which can potentially be extrapolated to other natural phenomena, fusing science and art (Chen, 2023). This interesting research can potentially be used in both teaching and recreational games, enabling the creation of more immersive environments and accurate representations of the dynamics of natural phenomena with a custom artistic flair.

---

<sup>1</sup>Touhou Wiki [https://en.wikipedia.org/wiki/Touhou\\_Project](https://en.wikipedia.org/wiki/Touhou_Project)

<sup>2</sup>Counter-Strike Wiki <https://en.wikipedia.org/wiki/Counter-Strike>

<sup>3</sup>DOTA Wiki <https://en.wikipedia.org/wiki/Dota>

<sup>4</sup>Half-Life Wiki [https://en.wikipedia.org/wiki/Half-Life\\_\(video\\_game\)](https://en.wikipedia.org/wiki/Half-Life_(video_game))

<sup>5</sup>Warcraft III Wiki [https://en.wikipedia.org/wiki/Warcraft\\_III:\\_Reign\\_of\\_Chaos](https://en.wikipedia.org/wiki/Warcraft_III:_Reign_of_Chaos)

AI Stable Diffusion v1.5<sup>6</sup> can be a very intriguing tool to utilize in custom character development because of its capacity to build visually appealing images, especially human faces, and to add unique features based on user prompts (Papa et al., 2023). Furthermore, the Character-based Outfit Generator (COG) (Forouzandehmehr et al., 2023), a model that generates a complete outfit tailored to the age and gender based on a specific prompt of a known character, could be used in video games to create relevant styles for characters and add some diversity. The prompt quality for a stable diffusion application has a significant impact on the resulting's image relevance, especially for more complex scenes (Hariffadzillah, Fadil, Harun, Ayob, & Razak, 2023). Although there are still drawbacks such as the style not always being consistent and some character emotions not being rendered well, it can be an interesting tool for generating complex images to help in the narrative of a video game, similar to what has been done for a children's storybook by Tengku et al (Hariffadzillah et al., 2023). In (Hattori, Kado, Matsumura, & Yamasaki, 2024), the authors created three distinct models that enable furniture placement in a room without altering the layout by combining FP-LoCon (furniture placement LoCon) and LP-CN (layout preservation controlNet) with stable diffusion. These models may also be investigated for use in video games to facilitate asset placement. Phillips, Jiao, and Clubb explore the use of AI generated images in urban design (Phillips, Jiao, & Clubb, 2024), which can be expanded and used to create gaming levels that have a more realistic and urban feel. Stable diffusion can also effectively generate objects and characters that scarcely exist in the real world, as demonstrated by the authors in (Dubrovinskaya & Tuhtan, 2023), who successfully generated realistic images of fish species even with a dataset as small as 5. This is an interesting advancement that may help generate animals and characters for video games. AI tools can also act as assistants to developers. For instance, the C2Create, which is primarily targeted at novice developers, assists them in suggesting various artwork or graphics for a card-based video game, based on the current input (Lopes, Liapis, & Yannakakis, 2014).

### 3.3.4 3D And Animation

For many game assets, simply having a static image is insufficient. Animation is necessary, as are 3D modeling, texturing, and lighting in the case of 3D assets (Hristov & Kinaneva, 2021). Deep

---

<sup>6</sup>Stable Diffusion 1.5 demo <https://huggingface.co/spaces/runwayml/stable-diffusion-v1->

learning models have been used to generate game assets such as textures and faces, but only for low-poly graphics (3D models that are constructed using a relatively small number of polygons) (Tilson & Gelowitz, 2019). Sketch2Map is a tool that converts a sketch into a game asset containing the elevation map and terrain geometry (Wang & Kurabayashi, 2020). It provides a quick and easy method to turn a simple sketch into playable game terrain, regardless of whether it was created by hand or by AI. Since it currently has some restrictions and limitations, more intricate drawings might not be appropriate. It does, however, provide a peek of how these technologies might be used to create usable landscape from an input image. Karras et al. developed an intriguing picture-to-video model that takes into account different individual identities, textures, and patterns while animating a human character and their clothing based on an image input and a set of positions (Karras, Holynski, Wang, & Kemelmacher-Shlizerman, 2023). With the help of this intriguing tool, game characters may be readily animated while still producing realistic and eye-catching results, especially in cut scenes. Although fairly basic for now, this offers an interesting glimpse at how collaboration between people and AI can be done in order to foster both creativity and efficiency. Lai, Leymarie, and Latham go into additional detail about the current state of content production, the role AI plays in it, and the ways in which AI may help people create relevant and appealing content (Lai, Leymarie, & Latham, 2022). Depending on the requirements of the user or designer, computer agents can assume a variety of functions, offering a useful and personalized assistance.

Video games became so popular that over half the planet are now playing them, primarily mobile games (Buser, 2023). One of the biggest costs associated with making video games is content development, which generative AI may lower by producing more assets faster (Buser, 2023). Additionally, as real time gen-AI is explored, it could very soon offer asset creation based on the player's in-game actions and choices. It will also enable realistic and life-like in-game characters that interact naturally with the player (Buser, 2023). Furthermore, Google's DeepMind's new generative model is able to generate game frames by extracting and combining different player actions, based on the gameplay videos that were used to train it (Douglas Heaven, 2024). This will eventually become a game-making tool, however, the researchers aim to go further and train AI bots in a sandbox, where they have already been observed to be cooperating to solve challenges (Douglas Heaven, 2024). In the future, AI bots are expected to be trained based on training videos. Nonetheless, AI remains a

tool to be used by people to enhance their creativity and efficiency and not as a replacement, since it cannot replace human imagination and originality (Buser, 2023).

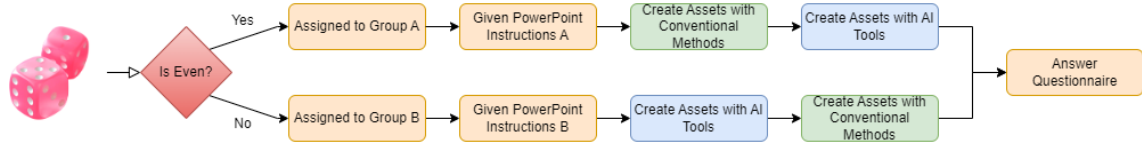


Figure 3.2: Experimental setting

### 3.4 Application

Using Unity, we developed a simple platformer game scene that includes some written and graphic elements. There are several objects in the scene, including platforms, decorative background objects, the main character and a few Non Playable Characters (NPCs). In addition, three discussion scenes—one at the start, one during interaction with the NPC, and one at the conclusion upon arriving at a particular location—are activated. Simple placeholder sprites (2D images used for characters and environment objects) are used for all visual elements, and the participant’s creations are intended to take their place. A screenshot of the main character placeholder in front of the NPC is displayed in Figure 3.1b. To keep things simple, no animations are used. The main character can move around with a simple script, jump on platforms and interact with other characters and specific objects. The participants were asked to imagine a setting for the game, and a plot that would necessitate three points of evolution that would be clarified through dialogue or narration. For the visual assets, participants were asked to replace as many as they wanted with sprites and images they created. They were also free to rearrange the objects around or create new ones. At the end, along with general questions, they were asked to evaluate how satisfactory the results were using each type of method and how easy or complicated it was to get the desired result. They were also given some more open-ended questions, asking which method they preferred and how would they go about creating assets if they had to in a real world setting. No prior knowledge of game development or asset creation is required for this study; however, having participants from different backgrounds and skill levels offered some valuable insights.



## 3.5 Methodology

### 3.5.1 Experimental Setting

The experiments received approval by the University Research Ethics Committee and were conducted over the period of 3 months and involved 34 participants. Prior to the study, the participants were given the details about the goal, as well as a consent form that they needed to sign. A dice roll was used to randomly allocate the participant to group A or group B at the start of the experiment (even for group A and odd for group B). To accommodate them as much as possible, participants were offered the option of conducting the study at the lab at Concordia University or in a café. A few did the study remotely as they were living too far and could not attend in person. In that situation, the Unity project was sent to them, and the study was conducted via screenshare over Discord. The participant received instructions via a PowerPoint presentation. Group A would start by creating the assets manually, whereas Group B would start with AI tools (Midjourney and ChatGPT). The instructions reflected that and included the general setting description and what was expected from the participant. Additionally, they explained what it means to generate the assets manually, how to do it, and provided a few examples for each stage. The researcher was also present to assist with the replacement of the gaming scene's assets, as the participants were not assessed on that aspect, although given the explanations in the presentation. Each participant had access to a Discord channel with MidJourney available for use, as well as internet access including ChatGPT. Throughout the experiment, the participants had the research representative available to answer any clarification questions. The experiment took between 1.5 and 4hrs for each participant, including brief unstructured breaks for eating lunch or snacking and using the restroom. Upon completion, each participant was required to respond to a questionnaire that asked about their demographics, their prior knowledge of diverse methods of asset creation, their satisfaction level with what they created, the ranking of each method, their ease of use, as well as some open-ended questions. This process is represented in Figure 3.2.

### **3.5.2 Pilot Test**

A pilot test of the application was conducted with two participants to provide feedback about the instructions, questions and overall process to ensure it is accessible to a range of knowledge levels and includes all pertinent questions.

### **3.5.3 Participant Demographics**

Using G\*Power tool, a sample size estimation of two tails distribution with an effect size of 0.5 showed that a minimum sample size of 34 participants is required for the study (Faul, Erdfelder, Lang, & Buchner, 2013). The statistical test used was the difference between two dependent means (matched pairs) with two tails since the same participants would do both the manual and AI steps and compare them as two equally valid methods. A total of 34 participants (64.7% male and 35.3% female) were recruited for the experiments. Although most had programming experience, their background was varied. The mean age was 34.51 years (SD = 7.03), with 85.3% being between the ages of 25 and 34.

### **3.5.4 Analysis**

The questionnaire that the participants were asked to answer after the study included some quantitative questions. Participants were asked to rate their skills in both manual and AI creation of written and visual assets. They were also asked to assess the ones they created by grading them out of 10, based on ranking, satisfaction level, and ease of creation. Python scripts with matplotlib and seaborn libraries were used to generate graphs based on the quantitative data gathered from the questionnaires.

## **3.6 Evaluation**

### **3.6.1 Previous Knowledge**

Participants were asked to rate their knowledge on related subjects out of 5 (1 being no knowledge at all, and 5 meaning they are comfortable in that field). Figure 3.3 illustrates the responses.

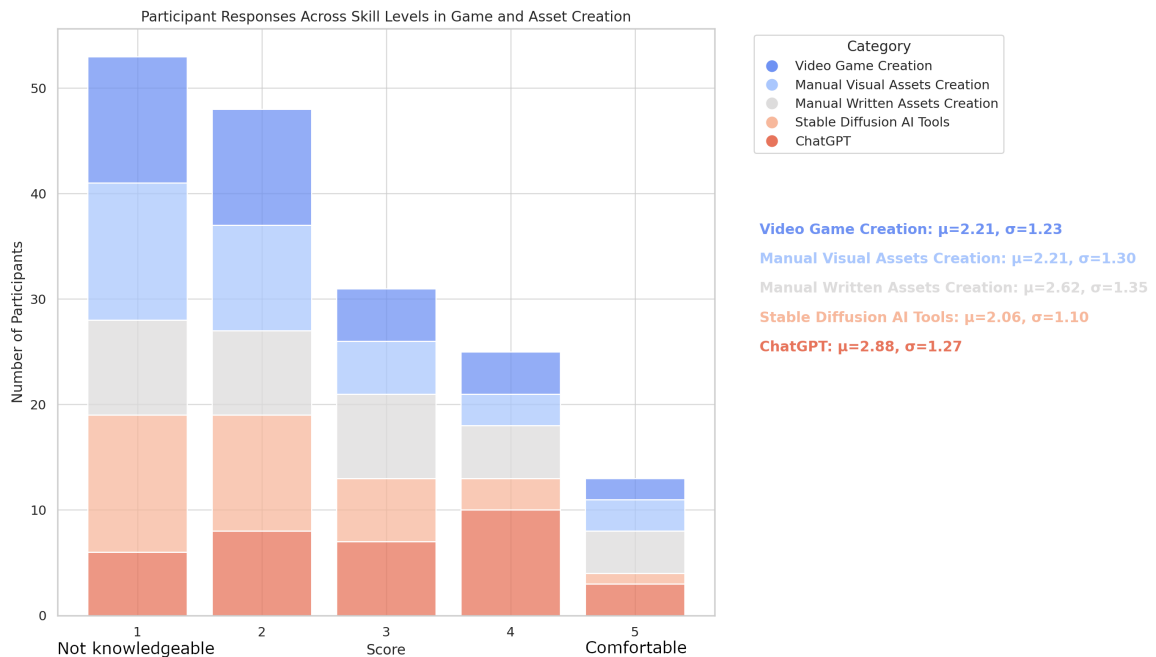


Figure 3.3: Participants' reported skill levels in game and asset creation

Participants had the most familiarity with ChatGPT ( $M = 2.88$ ,  $SD = 1.27$ ), followed by the manual written assets creation ( $M = 2.62$ ,  $SD = 1.35$ ). Stable diffusion AI tools have been rated the lowest ( $M = 2.06$ ,  $SD = 1.10$ ). Some were proficient in video game creation ( $M = 2.21$ ,  $SD = 1.23$ ), and manual visual asset creation ( $M = 2.21$ ,  $SD = 1.30$ ).

### 3.6.2 Written Assets

Participants were asked to assess their manual and AI-created textual assets on a scale of 10. They were also asked to rate the satisfaction and ease of both procedures on a scale of 10. For manually created assets, there is a correlation between the skill level of written assets and the rating of assets created, including how satisfied the participants were with it. Figure 3.4a clearly shows that the ranking increases as the skill level increases. However, for the highest skill level it suddenly declines, which could be attributed to the fact that they may have higher expectations of themselves, while being on a time limit, or simply not having time for creativity. As mentioned by participant #31, who ranked his skills as 5/5, his rating of the assets created as 1/10 and satisfaction as 1/10 "I'm a busy person, and I have little time for creativity.", mentioning that the main challenge is because

he is "very time constrained". As shown in Figure 3.4b, there is no correlation between the skill level and rating for assets created with AI, and they all rated relatively high. Manually developed written assets have a rating of ( $M = 4.76$ ,  $SD = 2.35$ ), and satisfaction of ( $M = 5.24$ ,  $SD = 2.28$ ), as shown in Figure 3.4a. In comparison, written assets generated using ChatGPT are rated higher ( $M = 7.94$ ,  $SD = 1.89$ ) and satisfaction is ( $M = 8.26$ ,  $SD = 1.96$ ) as demonstrated in Figure 3.4b.

Many participants described their hand written dialogue as "simple", "not creative enough", claiming that if they were given more time, they could have come up with something more elaborate and creative. Some struggled to express their thoughts adequately, such as participant #15 who stated, "I was not certain in how to phrase it in a more elegant manner (English being my second language)." Some participants were satisfied with the outcomes, saying it was funny and allowed for creativity (participant #21), see Figures 3.13a, 3.13b, and 3.13c. As challenges, a lot mentioned lack of imagination, time constraints, and the need to come up with a story on the moment. Group B participants also reported difficulty in "not straight up copying the AI generated one" (participant #7), although some viewed it as inspiration (participant #19). Overall, participants perceived written assets as fairly easy to write themselves ( $M = 6.79$ ,  $SD = 2.61$ ) as shown in Figure 3.5, where 10 implies very easy.

Most participants reported a positive experience with AI-generated textual assets, describing it as "creative and funny" (participant #16). A few found that it was a bit "too wholesome" (participant #5) and "wasn't exactly what they wanted". Nonetheless, an overwhelming majority believed that ChatGPT did a great job at delivering creative interactions. Some details are required in the prompt in order to generate exactly what the user intended, but they were usually effective. In one scenario, ChatGPT was unable to provide the proper context based on a very precise current world situation, but it is understandable as it does not have the current data. "I could never write something as good" (participant #32), "Fun, witty, and beautifully written" (participant #27), are just a few of the many positive comments the participants had to say about the AI, along with how easy and intuitive it is to use. Indeed, Figure 3.5 demonstrates this by displaying ( $M = 9.03$ ,  $SD = 1.60$ ) for ChatGPT use versus the manual ( $M = 6.79$ ,  $SD = 2.61$ ).

It is safe to conclude that AI is an useful tool for written assets because it is simple to use, creative and intuitive. However, the user must provide detailed prompts to have a more tailored

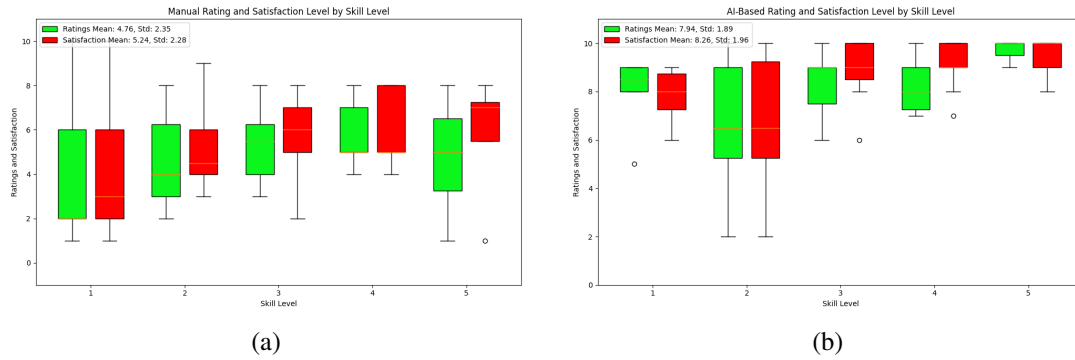


Figure 3.4: Skill level vs satisfaction level of written assets created by the participants manually (a) and using ChatGPT (b)

response. ChatGPT may have certain blind spots and biases towards how it writes stories, something to keep in mind.

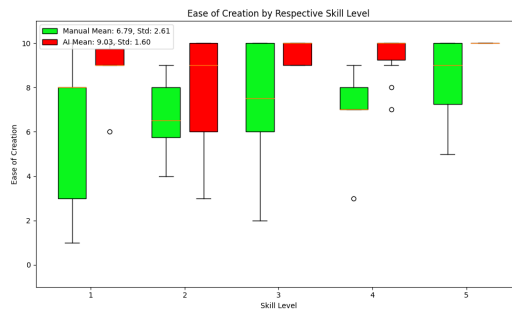


Figure 3.5: Ease of creation of written assets (1-very difficult, 10-very easy), by respective skill level of participants

### 3.6.3 Visual Assets

Figure 3.6a shows a correlation between participants' skills and their ranking of manually made graphic elements. The manual creation rating ( $M = 4.79$ ,  $SD = 2.61$ ) and satisfaction level ( $M = 4.94$ ,  $SD = 2.67$ ) demonstrate that, as expected, as the skill increases, so does the rating. Similarly to written assets with AI, visual assets with AI do not have a direct correlation with skill levels 3.6b and are rated higher than the manual counterpart at ( $M = 7.97$ ,  $SD = 1.87$ ) and satisfaction level of ( $M = 8.00$ ,  $SD = 1.48$ ).

The majority of poor ranks and challenges in manual visual asset generation stemmed from the participants' lack of drawing ability, and the assets acquired for free on the internet were not

always representative of what they wanted. However, several "unprofessionally" drawn assets were still rated high because "they were funny" (participant #21) and were "original artwork and design" (participant #24), highlighting that it is subjective and the artists have an emotional attachment to their art. The main challenge in drawing was lack of skill and expertise, and for those who did, a lack of time. Participants who choose to find the assets online faced issues such as inconsistent style across assets and difficulty finding exactly what they were looking for.

Overall, participants were pleased with the images generated by Midjourney: "really detailed and pretty" (participant #17). However, some struggled to get exactly what they wanted or in the style they wanted. Sometimes the styles might mismatch, or it would be difficult to generate a full body character rather than a portrait. Participants observed a learning curve for learning suitable prompts using Midjourney, but no major difficulties and in general very straightforward to use. A few mentioned the Midjourney servers going down during the experiment. Some nevertheless "preferred the hand drawn characters" (participant #24). The ease of use is illustrated in 3.7, we can see Midjourney being very easy to use ( $M = 8.21$ ,  $SD = 1.37$ ), compared to manual creation ( $M = 5.15$ ,  $SD = 2.62$ ). Regardless of the skill level, Midjourney is relatively easy to use, especially when contrasted to manually developing visual assets, where the facility increases with the skill level. It is vital to acknowledge that "funny" and "creative" were adjectives that appeared frequently in the positive descriptions of assets that were drawn manually. "They had charm that the generated ones didn't" (participant #7).

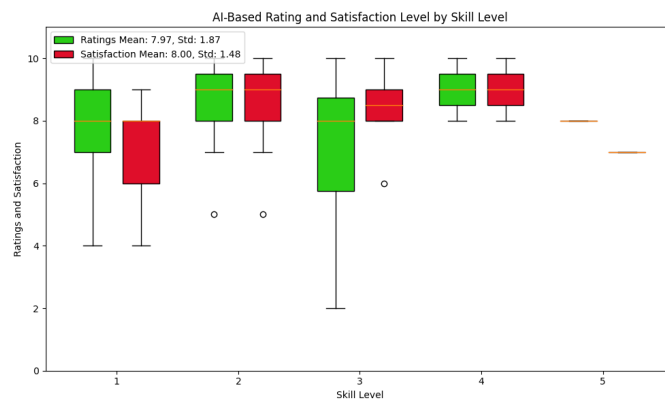
Overall, Midjourney provides an easy way to generate detailed and interesting assets. However, the main complaint is the consistency of the art style and the ability to constantly provide what the user wants. Although the assets are good, they are also "soulless" (participant #33). Making custom assets by hand is a skill that not everyone possesses, but it allows for truly unique and consistent visual assets. As participant #32 puts it, it is like the difference between a "printed letter and a handwritten one, handwritten is more special".

### **3.6.4 Comparison**

Further comparing the two methods, people ranked AI higher than the traditional approach of creating written assets in all three ratings: asset rating, satisfaction level and creation ease, as seen



(a)



(b)

Figure 3.6: Skill level vs satisfaction level of visual assets created by the participants manually (a) and using Midjourney (b)

in Figure 3.8. The violin graph represents the data distribution and helps visualize the variability, spread, and center. The rating of manual creation ( $M = 4.76$ ,  $SD = 2.39$ ), compared to ChatGPT ( $M = 7.94$ ,  $SD = 1.92$ ) Figure 3.8a, the satisfaction level ( $M = 5.24$ ,  $SD = 2.31$ ) versus ( $M =$

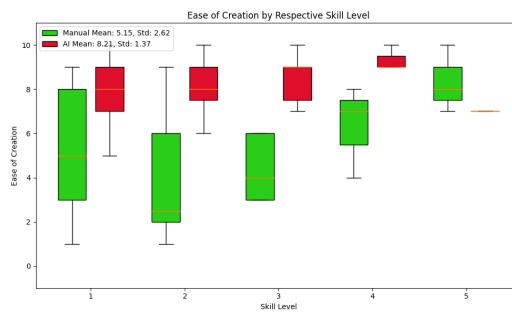


Figure 3.7: Ease of creation of visual assets (1-very difficult, 10-very easy), by respective skill level of participants

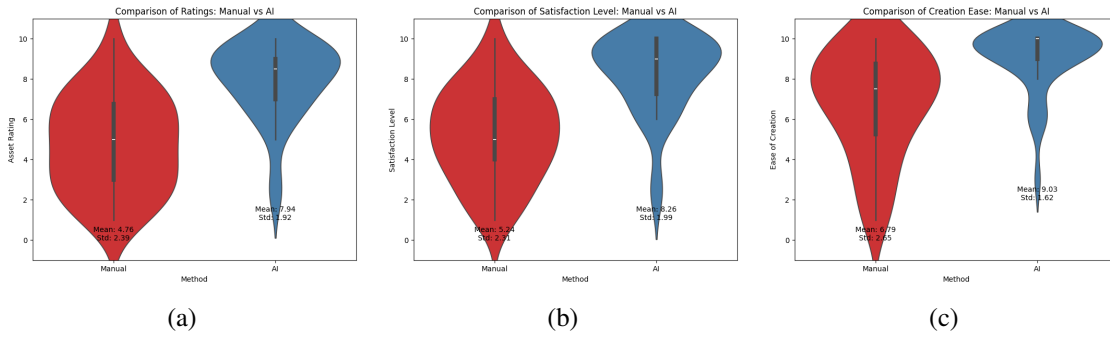


Figure 3.8: Comparison of rating (a), satisfaction level (b), and ease of creation (c) of written assets created using different methods as reported by participants

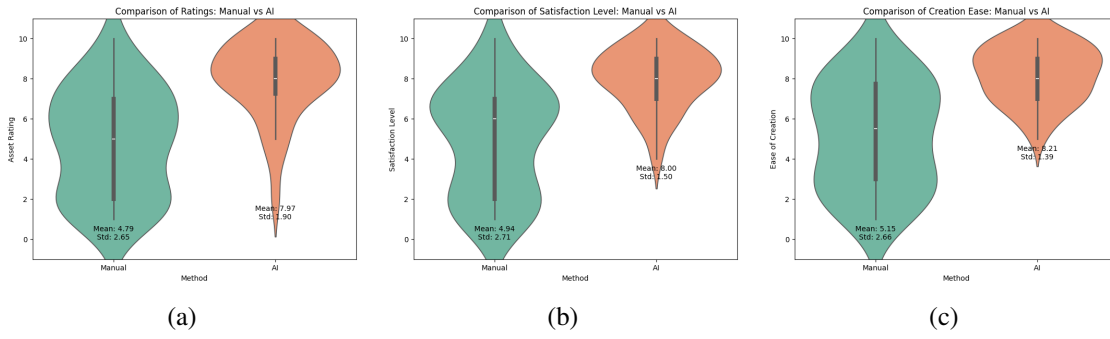


Figure 3.9: Comparison of rating (a), satisfaction level (b), and ease of creation (c) of visual assets created using different methods as reported by participants

8.26, 1.99) for AI (Figure 3.8b), and finally the ease of manual methods (M = 6.79, SD = 2.65) and AI respectively (M = 9.03, SD = 1.62) (Figure 3.8c). Overall, ChatGPT appears to be a better option for creating dialogues and narratives most of the time, with less volatility in ranking for the aforementioned criteria. Participant #33 was one of the few to give a low rating of 3/10 for AI-generated written dialogues and 2/10 for satisfaction level, citing that "the tone used by Chat GPT for buying a concert ticket for a metal show was unrealistic. I tried many times to make it sound normal". It is likely that ChatGPT is unaware of modern niche issues, such as it being complicated to buy tickets and the prices being extremely high and a joyful tone is not appropriate, or that the user lacks the ability to create more tailored instructions. Another participant (#6) who rated the AI-generated dialogues as 2/10 mentioned they were "well thought out but too long and wanted more witty". For reference, a sample dialogue of the participant can be seen in Figures 3.12a and 3.12b. It is certainly possible to ask ChatGPT to have the dialogue as long as the user wants, however the wittiness is subjective and may be lacking. Other participants were highly satisfied with the

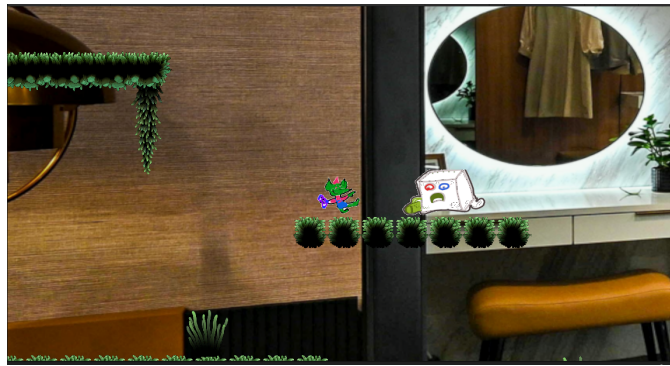


dialogues. Participant #1 mentions that between the two methods, although he also highly rated the dialogues created manually, he would choose "ChatGPT because it is very easy to use and can be modified to adjust the dialogues to the specific needs". Participant #25 also mentions that "Chatgpt is very good at writing elaborate text. I often write very short to the point things". A sample of AI dialogues can be seen in Figures 3.14a and 3.14b, while the ones created by the participant itself in Figures 3.15a and 3.15b. Manually created dialogues had a more even distribution, some participants were happy with it, some were not. In fact, 23/34 participants mentioned they preferred ChatGPT over manually written assets. Even though ChatGPT scored very good overall, more than half of the participants would use a mix of ChatGPT and their own ideas for the dialogues, compose them fully by themselves or hire a professional. This goes to show that given time and a goal of creating something worthwhile, AI might not always be the best answer.

The results are roughly similar for visual assets: Overall Midjourney was ranked higher in rating ( $M = 7.97$ ,  $SD = 1.90$ ) versus ( $M = 4.79$ ,  $SD = 2.65$ ), satisfaction level ( $M = 8.00$ ,  $SD = 1.50$ ) versus ( $M = 4.94$ ,  $SD = 2.71$ ) and ease of creation ( $M = 8.21$ ,  $SD = 1.39$ ) versus ( $M = 5.15$ ,  $SD = 2.66$ ) than the manual approaches. In this case too, the deviation is lower for all of the aforementioned factors while using Midjourney, implying that it is more consistently easy and rewarding. However, as previously stated, higher ratings do not necessarily mean that participants would choose that technique: "I enjoyed the precision of the hand-drawn illustrations, I could add and remove specific details as desired. I felt immense pleasure from drawing on a tablet, a new and refreshing experience. I am not the best at proportions or color schemes, but my designs were colorful, unique and have my trademark playfulness about them. The AI generated images were pretty to look at, but didn't have those specific aspects." (participant #12). The statement is well-put and summarizes the advantages of hand-drawn assets. Sample screenshots Figures 3.10a and 3.10b, demonstrate both manual and AI scenes of the participant's work. Out of 34 participants, 22 preferred AI to generate visual assets, while the remainder preferring to draw them manually. Mostly this is due to the fact that most participants lack the skills required to draw assets that could be then used in a real video game, as it requires a significant amount of work and practice. However, skilled artists much prefer sticking to traditional approaches. Even those with less skills still may prefer drawing them because they "felt more freedom, despite lack of skills." (participant #18), and,

although "it is much faster to use AI tools once you know how they work .. I preferred to use manual tools (Gimp) because it was quite user friendly and also fun to draw shapes" (participant #1). Outside the scope of this project, participant #1 stated he would "buy those that are not too expensive and I would create some of them if I cannot buy them cheaply, I would use a mix to create them". Many participants mentioned they would use a mix of AI and manual methods to achieve the best of both worlds. Several participants stated they would "hire an artist" (participant #33) or "collaborate with local artist(s)" (participant #34). Participant #25 would also rather "buy them. There are a lot of very good artists out there". Screenshots of the results can be seen in Figures 3.11a and 3.11b. Again, some recurring themes in creating assets manually are that "originality is of great importance" (participant #26), "it takes longer but you make them your own. I would take the help of the AI for the setting, background and stuff but the characters would be drawn" (participant #27). Participant #32 states that "as a visual artist, I still prefer to make my own art [...] sometimes art created by an AI is beautiful but feels a bit robotic, I can tell it was made by an AI and it does not have the same value to me". However, AI is "much less effort" (participant #6) and "quick". Participant #28 rated his visual assets created manually as 8, satisfaction with them as 5 and easy as 5, while rating visual ones as 8, 8 and 7 respectively. The participant chose free visual assets on the Internet, and it was "difficult to find assets with the same styles". AI assets posed a similar issue as well, as it was challenging to "come up with prompts to get the correct style for all assets. Sometimes asking for a 2D asset generates 3D assets that can't be used". The results of both are shown in Figures 3.1a and 3.1c. Ideally, the participant mentioned he would use a "mix of both" techniques, "create them using AI tools and pay someone to come up with different variations with specific changes I have in mind".

AI obviously saves time and is cost-effective. Creators without the aptitude or patience to write or draw could also use it to generate assets. However, many participants agree on a hybrid approach: gain some inspiration from AI, adapt it, and develop some of your own. "I think the AI is helpful for brainstorming" (participant #32). They also mention that "AI would be seen by a gamedev as useful to fill in the skill gap between their current capacity and what they would want to attain" (participant #15) and "AI makes it so much easier to plug holes created by the weaknesses of someone who has no experience doing specific thing in gaming" (participant #13). However, participant #13



(a)



(b)

Figure 3.10: Resulting scene with conventional methods (a) and AI tools (b) of participant #12

also raises an interesting point that "it would undermine the skill acquisition of going through the creation process and then carrying on this knowledge into the next project", that "it lacks a definite artstyle" and "lacking the complexity and unpredictability that form what our hands or brains can generate on a page of paper. Or in a videogame."

For a high-quality professional work, it is still worthwhile to employ professionals or do the work yourself if the skills allow it, as it results in assets that look and sound exactly how the creator intended, and many agree that art created by humans is unique and cannot be fully replaced by AI. Others like to write or draw, thus the fact that AI has more details or can do something "better" is irrelevant. Art is subjective, and people have emotional responses and attachments to it that AI cannot replicate.



(a)



(b)

Figure 3.11: Resulting scene with conventional methods (a) and AI tools (b) of participant #25

### 3.7 Conclusion and Future Work

To create engaging and creative narratives and dialogues, the writer must have original and cohesive ideas, as well as the necessary skills to put them into words. ChatGPT can be a useful tool for brainstorming and organizing one's ideas into coherent dialogues. Professionals can still produce the finest outcomes for visual assets since they have the requisite abilities and can provide original and innovative designs that truly capture the creator's vision. However, there is always a trade-off between the cost (both in term of price and time) and the quality. Depending on the project, it may be worthwhile to employ a hybrid strategy, or entirely utilize AI to save money and time.

It will be interesting to explore how AI would handle and generate animations. As demonstrated in this study, there are already various programs that can create animations and videos based on an image input, and create 3D models. It will be relevant to see how well they (or new applications) would perform for a simple video game, both 2D and 3D. Sora in particular would be an interesting application to test as an asset creating tool for video games as it can generate videos based on a text

Mama Bird: Hurry up, Strappy! My chicks are starving.

(a)

Strappy Sparrow: Found a spot, Drama Mama. Follow me.

(b)

Figure 3.12: Sample dialogue created with AI by participant #6

Dragon a deux pattes : Quelle est cette chose, un ours ? Un ogre?

(a)

Krystale : Il suffit dragon a deux pattes, dis-moi ou trouver l'abeille démoniac?

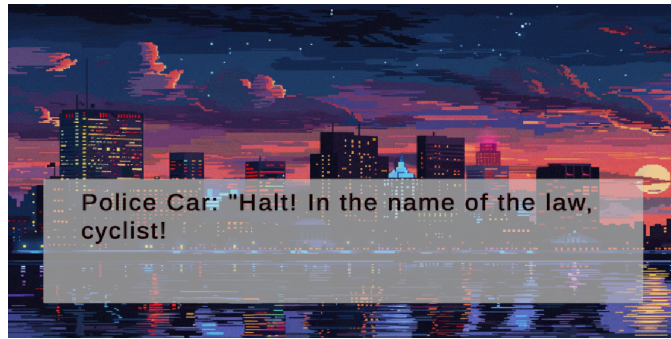
(b)

Tu devras d'abord répondre a une enigme super compliqué de la mort qui tue. .

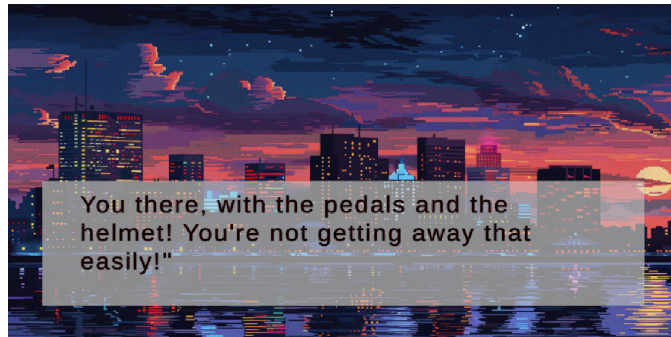
(c)

Figure 3.13: Sample dialogue created with conventional methods by participant #21

input. AI tools are also rapidly developing, so it could be worthwhile to revisit the same study in the future to see whether it would yield different results. Another intriguing issue to explore is how the most recent version of ChatGPT would perform if given or asked to develop a whole narrative



(a)



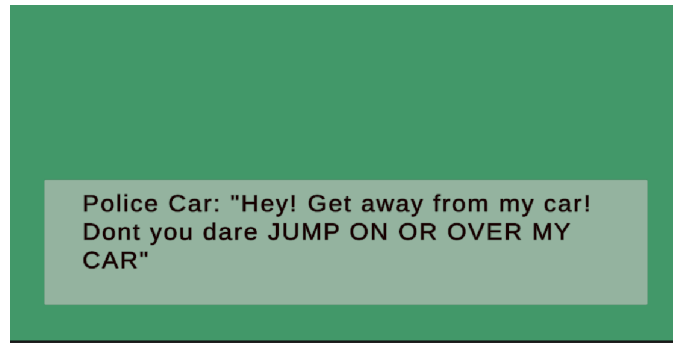
(b)

Figure 3.14: Sample dialogue created with AI by participant #25

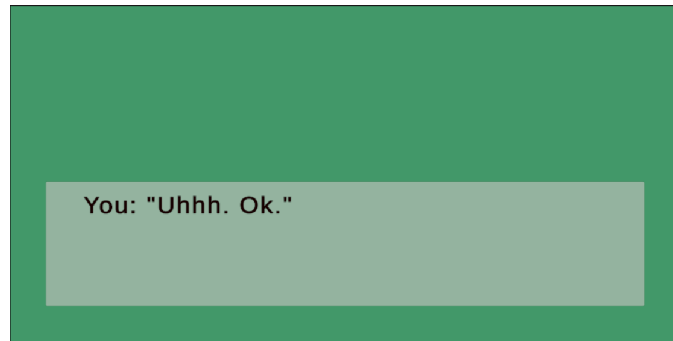
universe for a game and then prompted to produce dialogues or quests. Will it be able to keep track of events while remaining consistent with the content? Perhaps it will be easier to obtain desired art styles and tone from AI; however, it will not be replacing human creativity and originality anytime soon.

## Acknowledgments

This research was undertaken, in part, thanks to funding from the Natural Sciences and Engineering Research Council of Canada Grants RGPIN-2021-03479 (NSERC DG).



(a)

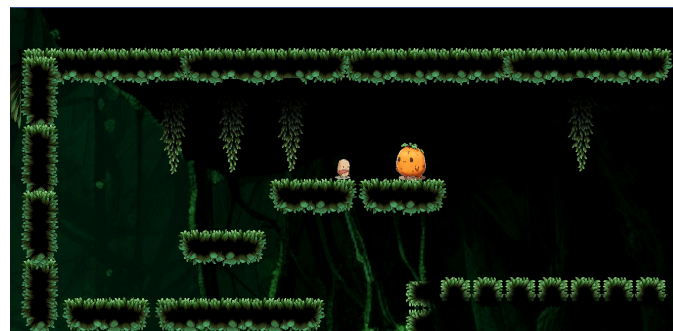


(b)

Figure 3.15: Sample dialogue created with conventional methods by participant #25



(a)



(b)

Figure 3.16: Resulting scene with conventional methods (a) and AI (b) of participant #32

## Chapter 4

# Conclusion and Future Work

In conclusion, video games are an integral part of our culture, more so than ever before. They are a great recreational activity that may also be used for trainings and simulations. Some of them assist players develop various skills such as critical thinking, decision-making, communication, and strategic planning, among others. A lot of effort can go into creating a high-quality video game, requiring a number of talented and driven people with various skill sets to work on it to seamlessly create a product that the users would appreciate. Designers, artists, developers, voice actors, and sound engineers, are just a few of the often required professionals to produce a video game. Generative AI models provide an innovative option to complete specific artistic tasks with fewer resources, generating high-quality work considerably faster and cheaper than traditional methods. However, as discussed previously, this is not without limitations. AI is not yet advanced enough to replace human creativity and individuality. Nevertheless, it can definitely be used by artists to get some ideas and make it work for them rather than replace them. If the quality and originality are less crucial, as they can be in some mobile games, then AI might be the preferred solution.

The very first CNN models and GANs enabled breakthroughs in image and natural-language processing, and were crucial in developing the AI as we know it. New AI tools are constantly being created, and many new technologies have emerged in recent years, if not months. AI is catching the attention of people from numerous domains as it offers additional capabilities to create different types of high quality content quickly, such as simulations and training tools, news articles, video game assets, and personal or professional assistants. In terms of video games, they can prove



to be valuable in aiding with the creation of various visual and written elements, such as stories, dialogues, environments, objects, NPCs, and cutscenes. Aside from static generated objects, they can help create an interesting and immersive world in which the player can chat with lifelike NPCs, have their actions reflected in the surrounding world, and influence the characters' interactions and behaviours based on the choices they make, resulting in a more realistic and unpredictable outcome. It is particularly interesting for sandbox games such as Sims, Role Playing Games (RPGs), and simulations, because they rely a lot on character development and story telling, although all genres can derive benefits from the new AI tools. It will be intriguing to see what the new advancements will bring in the future, as there are several major releases each year, alongside new versions and upgrades to existing tools.

However, one important aspect to keep in mind is the ethical considerations of such tools and how they may affect the artists. Since AI models are trained on available data, they are inherently biased and can perpetuate stereotypes and be misused, either accidentally or on purpose, to create harmful content. The creators must be mindful of the data being used by the model and whether it may infringe on the copyrights of original content authors. They must also review the output to ensure that it is consistent with their fundamental principles and does not promote any negative messages or stereotypes. On another level, it is vital to consider how it will impact artists, as their work remains incredibly valuable, but some companies may choose the AI alternative. It would be tragic to discourage new artists from pursuing their dreams due to a lack of opportunities, and not recognizing the value of existing ones in pursuit of novelty and profits. Hopefully, a balance can be found between AI and humans, resulting in ethical content that combines the best of both worlds.

This research focused on creating 2D visual assets and dialogues for a simple scene of a platformer game, similar to Mario, using conventional methods as well as AI tools (Midjourney and ChatGPT-3.5). The participants were both the creators and the evaluators of these assets, which adds a new and important view on the process and the end result of how these two methods compare.

Some potential future paths to explore include improved versions of ChatGPT, whether they will perform better, maintain better track of all characters and stories while generating relevant

and unique dialogues and quests, as well as be more adaptable to the universe's tone. Image-generating applications are also evolving and new ones are being developed, so it would be pertinent to see whether they can achieve a more consistent artistic style in the future and better understand user prompts. Another interesting feature to investigate is Sora, mentioned previously, which can generate videos based on a text, image, or video input, and how effectively it can be used in games to create scenes with pre-made game characters and environments. As gaming characters are not static, exploring how AI can generate sprites or 3D animations will definitely be of great value.

## **Appendix A**

# **Questionnaire**

The following is a verbatim copy of the questionnaire used in the paper "Comparative Analysis of Game Asset Creation Using Conventional and AI Methods", and the ethics board approval.

### **A.1 Questionnaire on Google forms**

# Project Questionnaire



**B** *I* U ↺ ↻

Form description

1. Group \*

- A
- B

2. Which category below includes your age? \*

- Less than 20
- 20-24
- 25-29
- 30-34
- 35-39
- 40-44
- 45-49
- 50 or older

3. What gender do you identify as? \*

- Male
- Female
- Other...

Figure A.1: Questionnaire part 1

4. What ethnicity are you? \*

- Caucasian
- African/African-American
- Native American
- East Asian
- Latino/Hispanic
- South-East Asian
- Central/South asian
- Other...

5. Rate your knowledge of the following, from 1 to 5, (1 meaning no knowledge at all, 5 meaning you feel comfortable in that field)

Description (optional)

a. Video game creation \*

1                      2                      3                      4                      5

b. Visual assets creation, e.g., characters, environments \*

1                      2                      3                      4                      5

Figure A.2: Questionnaire part 2

c. Written assets creation, e.g., dialogues, quests <sup>\*</sup>

1      2      3      4      5

---

d. Stable diffusion AI tools (Midjourney, OpenAI, etc) to generate visual assets <sup>\*</sup>

1      2      3      4      5

---

e. ChatGPT to generate written assets <sup>\*</sup>

1      2      3      4      5

After section 1   Continue to next section   ▾

**Section 2 of 4**

Project Questionnaire      ✕    ⋮

Manual creation

---

6. On a scale of 1-10, how would you rate the written assets you created **manually**? <sup>\*</sup>

1    2    3    4    5    6    7    8    9    10

Awful                                 Amazing

Figure A.3: Questionnaire part 3

Why? \*

Long answer text

---

7. On a scale of 1-10, how would you rate the visual assets you created **manually**? \*

1 2 3 4 5 6 7 8 9 10

Awful           Amazing

Why? \*

Long answer text

---

8. Were there any main challenges for **manually** written assets? Describe briefly: \*

Long answer text

---

9. Were there any main challenges for **manual** visual assets? Describe briefly: \*

Long answer text

---

10. On a scale of 1-10, how satisfied are you with the result of your **manually** written assets? \*

1 2 3 4 5 6 7 8 9 10

Very dissatisfied           Very satisfied

Why? \*

Long answer text

---

Figure A.4: Questionnaire part 4

11. On a scale of 1-10, how satisfied are you with the result of your **manual** visual assets? \*

1 2 3 4 5 6 7 8 9 10

Very dissatisfied           Very satisfied

Why \*

Long answer text  
.....

12. On a scale of 1-10, how easy was it to create the written assets **manually**? \*

1 2 3 4 5 6 7 8 9 10

Very difficult           Very easy

What made it easy or hard? \*

Long answer text  
.....

13. On a scale of 1-10, how easy was it to create the visual assets **manually**? \*

1 2 3 4 5 6 7 8 9 10

Very difficult           Very easy

What made it easy or hard? Did you mostly create from scratch or used free assets? \*

Long answer text  
.....

Figure A.5: Questionnaire part 5



Section 3 of 4

Project Questionnaire

AI tools

14. Which stable diffusion app did you use for visual assets? \*

Short answer text

15. On a scale of 1-10, how would you rate the written assets you created **with AI Tools**? \*

1 2 3 4 5 6 7 8 9 10

Awful           Amazing

Why \*

Long answer text

16. On a scale of 1-10, how would you rate the visual assets you created **with AI Tools**? \*

1 2 3 4 5 6 7 8 9 10

Awful           Amazing

Why \*

Long answer text

17. Were there any main challenges for written assets made **with AI Tools**? Describe briefly \*

Long answer text

Figure A.6: Questionnaire part 6

18. Were there any main challenges for visual assets made **with AI Tools**? Describe briefly \*

Long answer text

---

19. On a scale of 1-10, how satisfied are you with the result of your written assets **with AI Tools**? \*

1 2 3 4 5 6 7 8 9 10

Very dissatisfied           Very satisfied

Why \*

Long answer text

---

20. On a scale of 1-10, how satisfied are you with the result of your visual assets **with AI Tools**? \*

1 2 3 4 5 6 7 8 9 10

Very dissatisfied           Very satisfied

Why? What would you like to change about the generated assets? \*

Long answer text

---

21. On a scale of 1-10, how easy was it to create the written assets **with AI Tools**? \*

1 2 3 4 5 6 7 8 9 10

Very difficult           Very easy

Figure A.7: Questionnaire part 7

What made it easy or hard? \*

Long answer text

---

22. On a scale of 1-10, how easy was it to create the visual assets **with AI Tools**? \*

1 2 3 4 5 6 7 8 9 10

Very difficult           Very easy

What made it easy or hard? \*

Long answer text

---

After section 3 Continue to next section ▾

**Section 4 of 4**

Comparison ✕ ⋮

Description (optional)

23. Briefly describe which technique you preferred for written assets and why \*

Long answer text

---

24. Briefly describe which technique you preferred for visual assets and why \*

Long answer text

---

Figure A.8: Questionnaire part 8

25. Considering the time it takes to create the visual assets and the cost to buy them from an asset store, would you create or buy them? Explain why, and if you would create them, would you use manual methods or AI Tools? (or a mix): \*

Long answer text

26. Considering the time it takes to create the written assets and the cost pay someone to make them, would you create or buy them? Explain why, and if you would create them, would you use manual methods or AI Tools? (or a mix): \*

Long answer text

27. Any other feedback or comments: \*

Long answer text

Figure A.9: Questionnaire part 9

## A.2 Ethics Review Board Approval

This study received approval from the University Human Research Ethics Committee under the certification number 30018644.



### CERTIFICATION OF ETHICAL ACCEPTABILITY FOR RESEARCH INVOLVING HUMAN SUBJECTS

---

Name of Applicant: Viktorya Markutsa  
Department: Gina Cody School of Engineering and Computer Science\Computer Science and Software Engineering  
Agency: N/A  
Title of Project: Comparative study of asset creation for video games using AI Tools and manual methods  
Certification Number: 30018644

Valid From: October 20, 2023 To: October 19, 2024

The members of the University Human Research Ethics Committee have examined the application for a grant to support the above-named project, and consider the experimental procedures, as outlined by the applicant, to be acceptable on ethical grounds for research involving human subjects.

A handwritten signature in black ink that reads "Richard DeMont".

---

Dr. Richard DeMont, Chair, University Human Research Ethics Committee

Figure A.10: Certification of Ethical Acceptability

### **A.3 Instructions PowerPoint Group A**

Instructions given to participants from Group A (Group B instructions were the same but AI section was first)



# Asset creation instructions

Viktoriya Markutsa

# Goal

Think of a style that you want to use throughout the scene and the setting. Who is the protagonist, what is the story, what does the environment look like? Remain consistent with your ideas.

First, create all assets manually, then generate them using AI Tools. You can have 2 copies of the game as to not overwrite your work.



The image features a horizontal band of abstract graphic design. On the left, there is a dark brown circular shape containing a white silhouette of a pine branch. To its right is a light grey shape with a white outline of a pine branch. Further right is a solid olive green shape. A thin white line curves across the bottom of these shapes. The text 'Manual Creation' is centered within the dark brown circle.

Manual  
Creation

# Visual Assets

Go to the main scene, if not already there: in the Unity editor go to File -> Open Scene -> path/to/the/game/Basic Platformer/Assets/Scenes/MainScene.unity.

There is a simple scene with platforms, a background, decorative environment pieces, and characters. You can also update the assets in the DialogueScene

The current assets are simply placeholders. Your goal is to create your own. The style and setting are up to you.



# Visual Assets

You can either use tools (Photoshop, Gimp, etc) to draw the assets from scratch, find them open source on the internet, or do a mix of the two.



# Visual Assets

Once you have your asset created, you can drag it into the Basic Platformer/Assets/Sprites folder.

To replace the current object, click on it in the hierarchy view on the left of the editor, and in the inspector drag the sprite that you want to replace it with into the Sprite Renderer:

You may also need to adjust the scale if the object is too big or too small.

You can add or move platforms and environment pieces however you want.



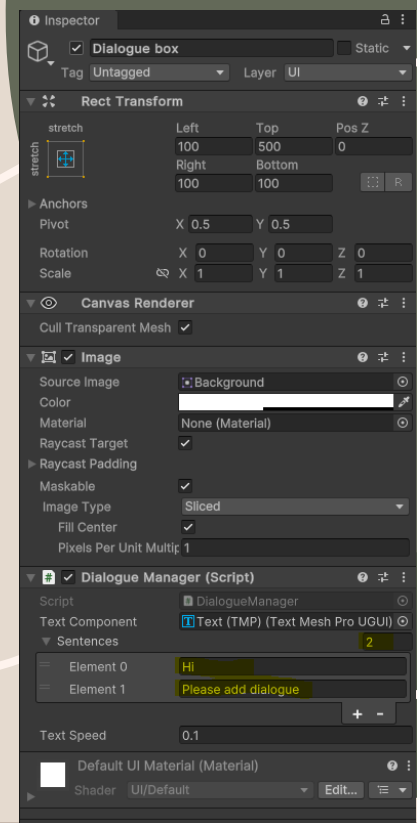
# Written Assets

In this study, the only written asset is the dialogue. It could be between our character and someone else: an npc explaining the mission, a friend, an enemy. It's entirely up to you.

There is 3 dialogue scenes in total: in the beginning, in the middle after interacting with the other character in the scene, and in the end, once you reach the right upper corner. The scenes are named DialogueScene, DialogueScene2 and DialogueScene3. Make sure the story progresses between the scenes.

To access the dialogue scenes, in the Unity editor go to File -> Open Scene -> path/to/the/game/Basic Platformer/Assets/Scenes/DialogueScene.unity

Once the scene is loaded, you can update the dialogues by clicking the DialoguesBox object in the hierarchy at the left and then writing in as many lines of dialogues as you want using the inspector



# Written Assets



The image features a horizontal band of abstract design. On the left, a dark brown circular shape overlaps a light grey shape. To the right, a dark olive green shape overlaps the grey one. A white, thin, wavy line curves across the bottom of the grey and green areas. In the top left corner, a grey silhouette of a pine branch is visible. The text 'AI Tools' is centered within the dark brown circle.

# AI Tools

# Visual Assets

Go to the main scene, if not already there: in the Unity editor go to File -> Open Scene -> path/to/the/game/Basic Platformer/Assets/Scenes/MainScene.unity.

There is a simple scene with platforms, a background, decorative environment pieces, and characters. You can also update the assets in the DialogueScene

The current assets are simply placeholders. Your goal is to create your own. The style and setting are up to you.





# Visual Assets

Midjourney (Discord, recommended method): You can generate images with midjourney using `/imagine`. For example, you can write `/imagine multiple item spritesheet dungeon`

You can be more precise but too much precision is also not always better. If you like one of the 4 generated images but would like something similar, you can make some variations of it by selecting V1/V2/V3 or V4. If you are satisfied with an image, then select U1/U2/U3/U4 and download it.

Bing/Dall-E/Other: You can generate images using prompts such as `multiple item spritesheet dungeon`

You can also specify the style. Again, being precise is good but too much precision can lead to unwanted results. Depending on the prompt, you will have to adjust it to fit your needs, as well as the limitations of the API.

If you created a sprite sheet but without a transparent background, you could fix it on <https://www7.lunapic.com/editor/>

Go to Edit->Transparent Background.

# Visual Assets



# Written Assets

Open ChatGPT and prompt to give the desired dialogues. It will be more tailored to your ideas if you give clear instructions about who is the npc, what is the main idea, the background/ world setting, etc. You can also be less precise to get some ideas and go from there (think to answer Who, What, Where, When, Why, How).

Example: generate a short dialogue between the protagonist and the npc giving him a mission for a game where the protagonist is a plumber and the npc is a dragon

You can then ask ChatGPT to update it in specific ways until you are satisfied or ask a completely different prompt.

generate a short dialogue between the protagonist and the npc giving him a mission for a game where the protagonist is a plumber and the npc is a dragon.

**Protagonist (Plumber):** Whoa, a talking dragon? This is unexpected.

**NPC Dragon:** Indeed, mortal. I am Drakorax, the ancient dragon guardian of these lands.

**Protagonist:** What can I do for you, mighty Drakorax?

**NPC Dragon:** A great darkness has enveloped the kingdom. The once-flowing pipes of our realm have been clogged with a malevolent force, spreading corruption. The balance of our world is at stake.

**Protagonist:** So, you want me to fix some pipes? I'm used to unclogging things, but this sounds serious.

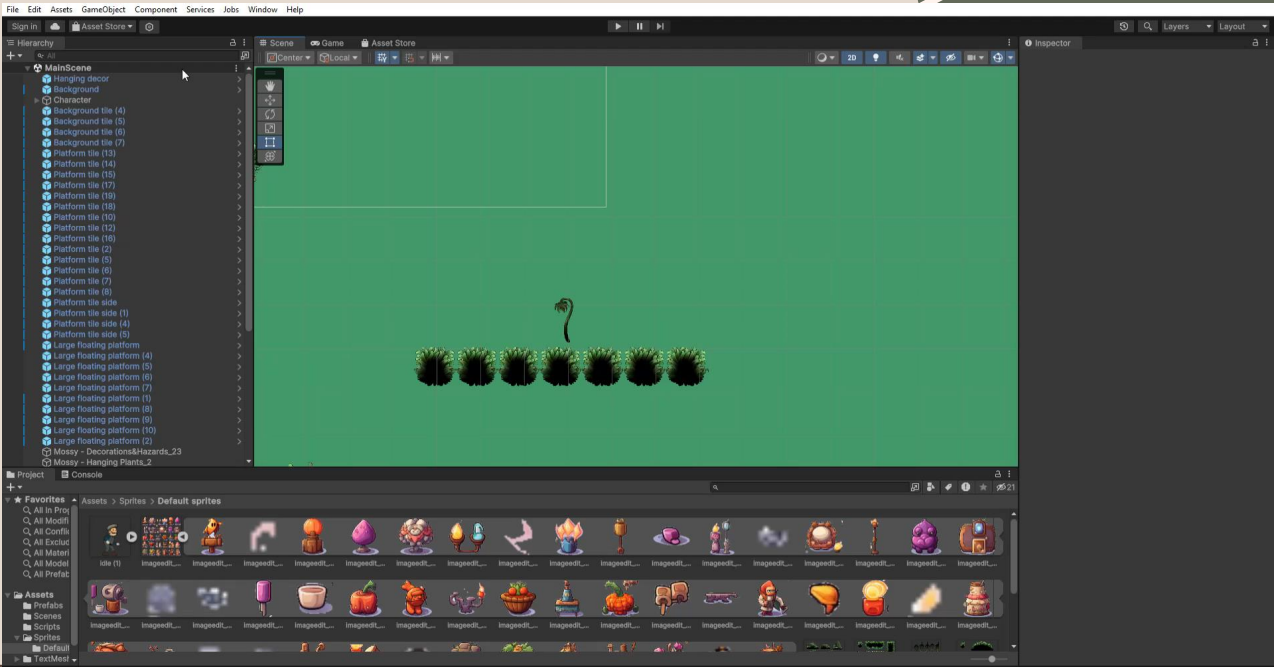
**NPC Dragon:** These pipes are more than they seem. They channel the energy that sustains our realm. I need you to venture deep into the underground plumbing network, confront the corruption, and restore the flow of life-giving energy.

**Protagonist:** Alright, I'm in. But how do I know you won't just gobble me up once I'm down there?

**NPC Dragon:** I am bound by the ancient code of honor that forbids me from harming those who aid our realm. Your safety is assured, but the task is perilous.

**Protagonist:** Got it, Drakorax. I'll unclog those pipes and save the kingdom's energy flow. Where do I start?

# Written Assets



# References

- Bengesi, S., El-Sayed, H., Sarker, M. K., Houkpati, Y., Irungu, J., & Oladunni, T. (2023). *Advancements in generative ai: A comprehensive review of gans, gpt, autoencoders, diffusion model, and transformers*.
- Bhagyashree, Kushwaha, V., & Nandi, G. (2020, 12). Study of prevention of mode collapse in generative adversarial network (gan). In *2020 ieee 4th conference on information & communication technology (cict)* (p. 1-6). doi: 10.1109/CICT51604.2020.9312049
- Buser, J. (2023, 10 6). *The generative ai revolution in the games industry: A path to boundless creativity*. Retrieved from <https://cloud.google.com/blog/products/gaming/generative-ai-in-the-games-industry>
- Cahyadi, M., Rafi, M., Shan, W., Lucky, H., & Moniaga, J. V. (2023). Accuracy and fidelity comparison of luna and dall-e 2 diffusion-based image generation systems. In *2023 ieee international conference on industry 4.0, artificial intelligence, and communications technology (iaict)* (p. 108-112). doi: 10.1109/IAICT59002.2023.10205695
- Caulfield, B. (2024). *Nvidia to reveal new ai innovations at ces 2024*. Retrieved from <https://blogs.nvidia.com/blog/ces-2024/>
- Cerezo-Pizarro, M., Revuelta-Domínguez, F.-I., Guerra-Antequera, J., & Melo-Sánchez, J. (2023). The cultural impact of video games: A systematic review of the literature. *Education Sciences*, 13(11). Retrieved from <https://www.mdpi.com/2227-7102/13/11/1116> doi: 10.3390/educsci13111116
- Chauhan, A., Anand, T., Jauhari, T., Shah, A., Singh, R., Rajaram, A., & Vanga, R. (2024). Identifying race and gender bias in stable diffusion ai image generation. In *2024 ieee 3rd international*

- conference on ai in cybersecurity (icaic)* (p. 1-6). doi: 10.1109/ICAIC60265.2024.10433840
- Chauhan, A., Chauhan, R., Nainwal, A., Arora, A., & Bhatt, C. (2023). Image multidiffusion algorithms for ai generative art. In *2023 6th international conference on contemporary computing and informatics (ic3i)* (Vol. 6, p. 1098-1103). doi: 10.1109/IC3I59117.2023.10397719
- Chen, H.-C. (2023). Harnessing ai for scientific illustration: Exploring tornado dynamics through midjourney. In *2023 12th international conference on awareness science and technology (icast)* (p. 136-141). doi: 10.1109/iCAST57874.2023.10359310
- Chollet, F., Wood, L., & Gupta, D. (2023). *High-performance image generation using stable diffusion in kerascv*. Retrieved from <https://www.tensorflow.org>
- Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018, January). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1), 53–65. Retrieved from <http://dx.doi.org/10.1109/MSP.2017.2765202>  
doi: 10.1109/msp.2017.2765202
- Cui, Y., Ren, J., Xu, H., Pengfei, H., Liu, H., Sun, L., ... Tang, J. (2023). Diffusionshield: A watermark for data copyright protection against generative diffusion models. *arXiv*. Retrieved from <https://ar5iv.labs.arxiv.org/html/2306.04642>
- Dayma, B., & Cuenca, P. (2022). *Dall.e mini- generate images from any text prompt*. Retrieved from <https://wandb.ai/dallemini>
- Douglas Heaven, W. (2024, 2 29). *Google deepmind's new generative model makes super mario-like games from scratch*. Retrieved from <https://www.technologyreview.com/2024/02/29/1089317/google-deepminds-new-generative-model-makes-super-mario-like-games-from-scratch/>
- Dsouza, E. G. (2024). How chatgpt works? technical explanation of its training model. *Edureka*. Retrieved from <https://www.edureka.co/blog/how-chatgpt-works> (Accessed: 2024-05-19)
- Duan, H., Huang, Y., Zhao, Y., Huang, Z., & Cai, W. (2022). User-generated content and editors in video games: Survey and vision. In *2022 ieee conference on games (cog)* (p. 536-543). doi: 10.1109/CoG51982.2022.9893717

- Dubrovinskaya, E., & Tuhtan, J. A. (2023). This fish does not exist: fish species image augmentation using stable diffusion. In *Oceans 2023 - limerick* (p. 1-6). doi: 10.1109/OCEANSLimerick52467.2023.10244720
- Duggan, M. (2015). *Gaming and gamers*. Retrieved from <https://www.pewresearch.org/internet/2015/12/15/gaming-and-gamers/> (Accessed: 2024-05-20)
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2013, 01). G\*power 3.1.7: A flexible statistical power analysis program for the social, behavioral and biomedical sciences, *beh. Res. Meth.s*, 39, 175-191.
- Forouzandehmehr, N., Cao, Y., Thakurdesai, N., Giahi, R., Ma, L., Farrokhsiar, N., ... Achan, K. (2023). Character-based outfit generation with vision-augmented style extraction via llms. In *2023 ieee international conference on big data (bigdata)* (p. 1-7). doi: 10.1109/BigData59044.2023.10416943
- Gaur, V., & Saunshi, N. (2022). Symbolic math reasoning with language models. In *2022 ieee mit undergraduate research technology conference (urtc)* (p. 1-5). doi: 10.1109/URTC56832.2022.10002218
- GIGABYTE. (2024). *Gigabyte's ai-powered gaming laptops: Introducing the future of gaming at ces 2024*. Retrieved from <https://www.gigabyte.com/Press/News/2134>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2020, oct). Generative adversarial networks. *Commun. ACM*, 63(11), 139–144. Retrieved from <https://doi.org/10.1145/3422622> doi: 10.1145/3422622
- Goodman, J., Wallat, A., Perez-Liebana, D., & Lucas, S. (2023). A case study in ai-assisted board game design. In *2023 ieee conference on games (cog)* (p. 1-4). doi: 10.1109/CoG57401.2023.10333138
- Gurucharan, M. (2023). Basic cnn architecture: Explaining 5 layers of convolutional neural network. *upGrad blog*. Retrieved from <https://www.upgrad.com/blog/basic-cnn-architecture/>
- Göring, S., Ramachandra Rao, R. R., Merten, R., & Raake, A. (2023a). Analysis of appeal for realistic ai-generated photos. *IEEE Access*, 11, 38999-39012. doi: 10.1109/ACCESS.2023.3267968

- Göring, S., Ramachandra Rao, R. R., Merten, R., & Raake, A. (2023b). Appeal and quality assessment for ai-generated images. In *2023 15th international conference on quality of multimedia experience (qomex)* (p. 115-118). doi: 10.1109/QoMEX58391.2023.10178486
- Hariffadzillah, T. S. N. T., Fadil, M. Z. M., Harun, A., Ayob, M. A., & Razak, M. R. A. (2023). Exploring ai technology to create children's illustrated storybook. In *2023 4th international conference on artificial intelligence and data sciences (aidas)* (p. 224-228). doi: 10.1109/AiDAS60501.2023.10284665
- Hassija, V., Chakrabarti, A., Singh, A., Chamola, V., & Sikdar, B. (2023). Unleashing the potential of conversational ai: Amplifying chat-gpt's capabilities and tackling technical hurdles. *IEEE Access*, *11*, 143657-143682. doi: 10.1109/ACCESS.2023.3339553
- Hattori, S., Kado, Y., Matsumura, K., & Yamasaki, T. (2024). Stable diffusion for furniture placement with preserved room layout in home staging. In *2024 ieee international conference on consumer electronics (icce)* (p. 1-6). doi: 10.1109/ICCE59016.2024.10444254
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 ieee conference on computer vision and pattern recognition (cvpr)* (p. 770-778). doi: 10.1109/CVPR.2016.90
- Hossain, M. Z., Uz Zaman, F., & Islam, M. R. (2023). Advancing ai-generated image detection: Enhanced accuracy through cnn and vision transformer models with explainable ai insights. In *2023 26th international conference on computer and information technology (iccit)* (p. 1-6). doi: 10.1109/ICCIT60459.2023.10440990
- Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... Adam, H. (2017, 04). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *ArXiv*.
- Hristov, G., & Kinaneva, D. (2021). A workflow for developing game assets for video games. In *2021 3rd international congress on human-computer interaction, optimization and robotic applications (hora)* (p. 1-5). doi: 10.1109/HORA52670.2021.9461355
- Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017, July). Densely connected convolutional networks. In *Proceedings of the ieee conference on computer vision and pattern recognition (cvpr)*.
- Karras, J., Holynski, A., Wang, T.-C., & Kemelmacher-Shlizerman, I. (2023). Dreampose: Fashion



- image-to-video synthesis via stable diffusion. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (p. 22623-22633). doi: 10.1109/ICCV51070.2023.02073
- Krizhevsky, A., Sutskever, I., & Hinton, G. (2012, 01). Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*, 25. doi: 10.1145/3065386
- Kunal, Rana, M., & Bansal, J. (2023). The future of openai tools: Opportunities and challenges for human-ai collaboration. In *2023 2nd International Conference on Futuristic Technologies (Incoft)* (p. 1-6). doi: 10.1109/INCOFT60753.2023.10424990
- Lai, G., Leymarie, F. F., & Latham, W. (2022). On mixed-initiative content creation for video games. *IEEE Transactions on Games*, 14(4), 543-557. doi: 10.1109/TG.2022.3176215
- Lee, J., Moon, S., & Nam, D. (2023). A study on military object detection in panoramic view using stable diffusion. In *2023 14th International Conference on Information and Communication Technology Convergence (ICTC)* (p. 1889-1891). doi: 10.1109/ICTC58733.2023.10392321
- Lopes, P., Liapis, A., & Yannakakis, G. N. (2014). The c2create authoring tool: Fostering creativity via game asset creation. In *2014 IEEE Conference on Computational Intelligence and Games* (p. 1-2). doi: 10.1109/CIG.2014.6932916
- Luccioni, S., Pistilli, G., Rajani, N., Allendorf, E., Solaiman, I., Lambert, N., & Mitchell, M. (2023). Ethics and society newsletter #4: Bias in text-to-image models. *Hugging Face*. Retrieved from <https://huggingface.co>
- Nguyen, L. X., Sone Aung, P., Le, H. Q., Park, S.-B., & Hong, C. S. (2023). A new chapter for medical image generation: The stable diffusion method. In *2023 International Conference on Information Networking (ICOIN)* (p. 483-486). doi: 10.1109/ICOIN56518.2023.10049010
- OpenAI. (2024). *Sora*. Retrieved from <https://openai.com/sora> (Accessed: 2024-05-19)
- Ostuni, D., & Galante, E. T. (2021). Towards an ai playing touhou from pixels: a dataset for real-time semantic segmentation. In *2021 IEEE Conference on Games (Cog)* (p. 1-5). doi: 10.1109/CoG52621.2021.9619112
- Paik, S., Bonna, S., Novozhilova, E., Gao, G., Kim, J., Wijaya, D., & Betke, M. (2023). The affective nature of ai-generated news images: Impact on visual journalism. In *2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII)* (p. 1-8). doi: 10.1109/ACII59096.2023.10388166

- Papa, L., Faiella, L., Corvitto, L., Maiano, L., & Amerini, I. (2023). On the use of stable diffusion for creating realistic faces: from generation to detection. In *2023 11th international workshop on biometrics and forensics (iwbf)* (p. 1-6). doi: 10.1109/IWBF57495.2023.10156981
- 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*.
- Phillips, C., Jiao, J., & Clubb, E. (2024). Testing the capability of ai art tools for urban design. *IEEE Computer Graphics and Applications*, 44(2), 37-45. doi: 10.1109/MCG.2024.3356169
- Phung, V. H., & Rhee, E. J. (2019, 10). A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets. *Applied Sciences*, 9, 4500. doi: 10.3390/app9214500
- Powley, E. J., Colton, S., Gaudl, S., Saunders, R., & Nelson, M. J. (2016). Semi-automated level design via auto-playtesting for handheld casual game creation. In *2016 IEEE conference on computational intelligence and games (CIG)* (p. 1-8). doi: 10.1109/CIG.2016.7860438
- Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. In Y. Bengio & Y. LeCun (Eds.), *4th international conference on learning representations, ICLR 2016, san juan, puerto rico, may 2-4, 2016, conference track proceedings*. Retrieved from <http://arxiv.org/abs/1511.06434>
- Raut, G., & Singh, A. (2024). Generative ai in vision: A survey on models, metrics and applications. *ArXiv, abs/2402.16369*. Retrieved from <https://api.semanticscholar.org/CorpusID:268032428>
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022, jun). High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (p. 10674-10685). Los Alamitos, CA, USA: IEEE Computer Society. Retrieved from <https://doi.ieeecomputersociety.org/10.1109/CVPR52688.2022.01042> doi: 10.1109/CVPR52688.2022.01042
- Singh, A. (2023). A survey of ai text-to-image and ai text-to-video generators. In *2023 4th international conference on artificial intelligence, robotics and control (AIRC)* (p. 32-36). doi: 10.1109/AIRC57904.2023.10303174

- Tan, M., & Le, Q. (2019, 09–15 Jun). EfficientNet: Rethinking model scaling for convolutional neural networks. In K. Chaudhuri & R. Salakhutdinov (Eds.), *Proceedings of the 36th international conference on machine learning* (Vol. 97, pp. 6105–6114). PMLR. Retrieved from <https://proceedings.mlr.press/v97/tan19a.html>
- Tilson, A., & Gelowitz, C. M. (2019). Towards generating image assets through deep learning for game development. In *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)* (p. 1-4). doi: 10.1109/CCECE.2019.8861965
- Turner Lee, N., & White, N. (2024). How openai's sora hurts the creative industries. *Brookings*. Retrieved from <https://www.brookings.edu/articles/how-openais-sora-hurts-the-creative-industries/>
- Understanding the chatgpt architecture: A comprehensive guide. (2023). *Aakit*. Retrieved from <https://aakit.com/chatgpt-architecture-explained> (Accessed: 2024-05-19)
- Vahdat, A., & Kreis, K. (2023). Improving diffusion models as an alternative to gans, part 1. *NVIDIA Technical Blog*. Retrieved from <https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/> (Accessed: 2024-05-19)
- Vartinen, S., Hamalainen, P., & Guckelsberger, C. (2022, 01). Generating role-playing game quests with gpt language models. *IEEE Transactions on Games, PP*, 1-12. doi: 10.1109/TG.2022.3228480
- Vayadande, K., Bhemde, S., Rajguru, V., Ugile, P., Lade, R., & Raut, N. (2023). Ai-based image generator web application using openai's dall-e system. In *2023 international conference on recent advances in science and engineering technology (icrasnet)* (p. 1-5). doi: 10.1109/ICRASNET59632.2023.10420306
- Wang, T., & Kurabayashi, S. (2020). Sketch2map: A game map design support system allowing quick hand sketch prototyping. In *2020 IEEE Conference on Games (Cog)* (p. 596-599). doi: 10.1109/CoG47356.2020.9231754
- Will, & David. (2024). *The current state of ai gaming as we see it: Generative text rpgs, ai storytellers, and games with ai npcs*. Retrieved from <https://www.fables.gg/blog/>

[ai-gaming](#) (Accessed: 2024-05-19)

- Yu, H., Liu, Z., & Guo, Y. (2023). Application status, problems and future prospects of generative ai in education. In *2023 5th international conference on computer science and technologies in education (cste)* (p. 1-7). doi: 10.1109/CSTE59648.2023.00065
- Zambrano, O., & Senouci, B. (2023). Image classification improvement: Text-to-image ai for synthetic dataset approach. In *2023 49th euromicro conference on software engineering and advanced applications (seaa)* (p. 74-77). doi: 10.1109/SEAA60479.2023.00020
- Zhang, J., Wang, Y., Tohidypour, H. R., & Nasiopoulos, P. (2023). Detecting stable diffusion generated images using frequency artifacts: A case study on disney-style art. In *2023 ieee international conference on image processing (icip)* (p. 1845-1849). doi: 10.1109/ICIP49359.2023.10221905
- Zhou, J., Gao, J., Wang, Z., & Wei, X. (2023). Copyscope: Model-level copyright infringement quantification in the diffusion workflow. *arXiv*. Retrieved from <https://ar5iv.labs.arxiv.org/html/2311.12847>