**Perception Maze:**

**A Comparative Study of Human and Artificial Intelligence on the Perception of Authenticity in Artistic Images**

https://www.youtube.com/watch?v=SkvKRpdk1uA

Danqi Wang

21015048

Supervisor: Jasper Zheng  
Word count：7012

Abstract

This thesis investigates the perceived differences between humans and Generative Adversarial Networks (GANs) in recognising the authenticity of images, with a focus on artistic images. The study analyses the Metfaces dataset using StyleGAN3's discriminator and compares the ability of AI and humans to recognise authentic and generated portraits by employing quantitative survey methods and qualitative interviews. The results shed light on the complexity of authenticity judgements, suggesting that while the AI can achieve baseline performance levels, human evaluators demonstrate higher accuracy, especially in cases where the AI misjudges. The study highlights the importance of personal experience and intuitive judgement in the authenticity assessment process and calls for AI systems to be able to integrate human-like emotional and aesthetic judgements. The study contributes to a broader exploration of the integration of AI in the arts, suggesting directions for future research, including enhancing AI's understanding of the nuances of human judgement and exploring the impact of the presentation medium on the perception of authenticity.

Content

[1. Introduction 4](#_Toc151616664)

[2. Related Work 5](#_Toc151616665)

[2.1 Information Dissemination and Subjectivity of Images 5](#_Toc151616666)

[2.2 Artistic Image Generation in Machine Learning 7](#_Toc151616667)

[2.3 GAN Detector 8](#_Toc151616668)

[2.4 Assessing Image Authenticity: Human vs. Computer Vision Systems 10](#_Toc151616669)

[2.5 Chapter Summary 11](#_Toc151616670)

[3. Methodology 13](#_Toc151616671)

[4. Implementation 14](#_Toc151616672)

[4.1 Extracting Output Data from the Discriminator 14](#_Toc151616673)

[4.2 Data Analysis and Conclusions 15](#_Toc151616674)

[5. A Project Write-Up 16](#_Toc151616675)

[5.1 Theoretical Background 16](#_Toc151616676)

[5.2 Experimental Design and Sample Selection 16](#_Toc151616677)

[5.3 Results Analysis 17](#_Toc151616678)

[5.3 Brief Discussion and Conclusion 22](#_Toc151616679)

[6. Evaluation 24](#_Toc151616680)

[6.1 Interview Format 24](#_Toc151616681)

[6.2 Thematic Analysis 25](#_Toc151616682)

[7. Discussion 28](#_Toc151616683)

[7.1 Integration of Survey and Thematic Analysis Results 28](#_Toc151616684)

[7.2 Reflection 30](#_Toc151616685)

[7.3 Future Work 30](#_Toc151616686)

[8. Conclusions 31](#_Toc151616687)

[9. Appendix 32](#_Toc151616688)

[10. Reference 33](#_Toc151616689)

## 1. Introduction

With the advent of sophisticated artificial intelligence techniques, the understanding of the nature of image authenticity has reached a new frontier. As these techniques continue to evolve, particularly in the field of art, recognising what is real and what is artificially generated has become a key challenge. This thesis delves into the heart of this challenge, exploring the perceptual dichotomy between human cognition and GANs, particularly through the lens of the capabilities of StyleGAN3 discriminators and their interaction with human judgement.

The genesis of this research lies in a critical exploration of how technological advances (particularly in the field of AI image generation) are redefining the benchmarks of authenticity. It seeks to shed light on the subtle ways in which humans perceive and judge the authenticity of art images - an area where the human eye and AI intersect but diverge significantly.

In the sections that follow, the paper describes a mixed-method approach that combines rigorous quantitative analysis with insightful qualitative interviews to illuminate the multidimensional nature of image authenticity judgements. Thematic analysis of the interview data provides insights into the human factors that influence these judgements, while the findings provide empirical evidence of the differences between AI and human capabilities in this area.

The journey of this thesis runs through the intricate interplay between emotion, art and technological prowess, ultimately aiming to enrich the discourse on the role of AI in the arts and guide the development of more empathetic and culturally aware AI systems. Through this exploration, the thesis positions itself at the intersection of technology, art, and human perception, offering a unique perspective on evolving narratives of authenticity in the digital age.

## 2. Related Work

This chapter provides a comprehensive examination of the use of images in information dissemination, the subjectivity of image-generation techniques, and methods for assessing the authenticity of images. Firstly, we discuss the role of the intuitive and subjective nature of images in the communication of information in modern society, particularly as carriers of historical and cultural memory. By analysing fictional images on social media, we reveal the potential misleading nature of the information conveyed by images. Secondly, this paper delves into the application of machine learning, especially Generative Adversarial Networks (GAN), in art image generation. The achievements of StyleGAN and its iterations in creating hyper-realistic images are analysed, as well as their use in deep forgery techniques. Further, the development of GAN detectors is explored, tools that are crucial in distinguishing generated images from real ones, despite their limitations in the recognition of art images. Finally, the differences between human and computer vision systems in the assessment of image authenticity are compared, especially when dealing with images generated by advanced techniques like StyleGAN3.

Overall, this chapter provides a comprehensive perspective for understanding and evaluating the complexity and impact of images in the current digital age by examining the role of images in information dissemination, the application of machine learning in the generation of artistic images, and different approaches to image authenticity assessment.

### 2.1 Information Dissemination and Subjectivity of Images

In modern society, images play a key role in online information dissemination as an important medium for conveying information. According to Mitchell (2011), images are intuitive, vivid, and easy to understand, can quickly attract people's attention and convey a large amount of information in a short period of time. This highlights the powerful benefits and convenience of images as visual elements. In Adam Curtis's (2016) documentary, he mentions that images are not only tools used to display or record historical events and people, they also carry traces of history and cultural memories. This side by side supports the versatility and readability of the role of images. At the same time, in HyperNormalisation, he also demonstrates through archival images the reality of the data processing machines of large corporations such as Google and Facebook, which represent materialised manifestations of existing socio-political control structures (Utterson, 2023), which include the viral spread of false and misleading information. This confirms the motivation of this paper, which is that images are 'subjective'.

Because of this, images, as an extremely important and frequent tool of information transmission, represent objective meanings that are embodied by specific objects or people, but the subjectivity of an image can be endowed with its creator. The term "endowed with" is explained in the Cambridge Dictionary (2023) as follows: If someone or something is endowed with a particular quality or feature, the person or thing naturally has the right to be endowed with it. Thus, the subjectivity of an image is influenced by the maker, which allows the image to express meanings other than its subject, an expression that is often used in works of art, and is no exception in generative art.

Some time ago, a set of 50 artworks of Donald Trump ranging from a fictional arrest to a fictional prison break was widely circulated on social media platforms (Kahn, 2023). The set of images was created by one of the founders of Bellingcat using Midjourney, and it was created to raise awareness of the impact of images generated by artificial intelligence at a certain point in time: this occurred when the Justice Department filed criminal charges against Trump for allegedly bribing pornographic actresses into silence (Gonzalo, 2023). Some researchers have pointed out that the human brain first perceives a visual image as a unified whole and then analyses its parts (Liu et al., 2022). However, after seeing the whole, it may ignore the fine details. This has led many people who are not aware of generative imagery to misinterpret the set of images as a real arrest scene. Even more worrying is the possibility that the generative images could be used for malicious purposes. German artist Boris Eldagsen won a major photography prize but declined it after revealing that his work was created using AI (Glynn, 2023); and just over a month into the war in Ukraine, a deep forgery of Zelensky emerged (Higgins, 2023). According to various voices, including some leading AI researchers (Acerbi, Altay and Mercier, 2022), generative AI will make it easier to create truthful but false or misleading content on a large scale, with potentially catastrophic consequences for people's beliefs and behaviours, the information public sphere and democracy (Simon, Altay and Mercier, 2023). Perceiving and judging the authenticity of images with subjectivity in today's cyberspace is a tricky problem for both humans and AI.

### 2.2 Artistic Image Generation in Machine Learning

In the current "misinformation age", artificial intelligence plays a pivotal role. Machine learning, as one of the branches of AI, and in particular the development of deep learning techniques within it, has led to major breakthroughs in the field of image generation. Generative Adversarial Networks (GAN) is one of them, which can extensively learn the style of a set of images and can better replenish similar features (Yang, 2022).GANs are capable of generating works of art such as photographs, oil paintings, etc. by learning a large number of image datasets through neural networks (metfaces dataset/lhq). Relatively, this can also lead to some negative social discussions, as in the example mentioned in 2.1.

The principle mechanism by which GAN operates is that a generative model is confronted with a discriminative model and learns to determine whether the samples are from the model distribution or from the discriminative model (Yang, 2022). Due to the clever adversarial design of GAN, the generative network has more powerful data generation capabilities than other generative models, which has made it one of the models that have been the focus of academic attention in recent years. Here, we can pay special attention to the discriminator (D). In traditional GANs, the discriminator is optimised to classify real data as real and generated data as fake. This is because, after constant confrontation, the hidden layer of D is updated to heel level, constantly looking for differences between the new output image and the real one. While the discriminator must also improve its capabilities in order to train these larger networks, it is usually rarely, if ever, completely discarded and explored.

Broad, Leymarie, and Grierson (2020) delved into the application of Generative Adversarial Networks (GANs) to deep forgery image production. They go beyond traditional uses by freezing discriminator weights and tuning generators to produce increasingly abstract images, highlighting the sensitivity of deep forgery parameters and demonstrating the malleability of the technique. Their research not only exemplifies the fusion of technology and art, but also raises concerns about the potential misuse of deep forgery techniques. Porres (2021) goes a step further by using features learnt by the discriminator to create new images, challenging the traditional use of GANs primarily aimed at enhancing image resolution. The use of StyleGAN and its advanced versions in these studies signals its strong potential and influence in the fusion of art and AI. The body of research suggests that while studies using discriminators exist, they tend to focus more on the generation of new images than on improving the discriminator's ability to recognise authenticity. The consistent use of StyleGAN3 and its iterations in these studies emphasises the utility of StyleGAN3 in the field of art generation.

### 2.3 GAN Detector

With the development of digital technologies, discriminating between real and fake images becomes increasingly difficult, especially with the advancement of advanced techniques such as Generative Adversarial Networks (GANs) (Goodfellow et al., 2014). Based on this, GAN detectors have become an important tool for discriminating the authenticity of visual content in the field of machine learning-based image generation.A GAN detector is a system designed to differentiate between real images and those synthesised by Generative Adversarial Networks (GANs). This distinction is crucial in various fields such as digital forensics, media integrity and visual information protection.GAN models such as GP-GAN (Karras et al., 2017), BigGAN (Brock, Donahue and Simonyan, 2018), StyleGAN (Karras, Laine and Aila, 2018) and StyleGAN2 (Karras et al., 2020)) recent advances significantly enhance the synthesis of realistic facial images by embedding rich semantic information in intermediate features to generate high-quality latent spaces (Wang et al., 2023).GAN detectors typically employ discriminative models that are trained to recognise the subtle patterns and artefactual features in GAN-generated images. These patterns may be imperceptible to the human eye, but can be detected by sophisticated machine learning algorithms. discriminators in GANs that were initially used to train generative models can now be repurposed as detectors. Alternatively, standalone models, such as Convolutional Neural Networks (CNNs), can be trained specifically for the detection task. It is worth noting that StyleGAN2 and StyleGAN3 have made great strides in improving the quality of the generated faces, which has made it substantially more difficult for the GAN detectors to determine whether they are real or fake.

Based on a review of previous studies, the authors found that there are no GAN detectors that specialise in detecting art-based images. However, it has been mentioned that the core challenge of visual recognition is the same for both faces and objects (Tsao and Livingstone, 2008). So here in this paper we briefly discuss the past detection for GAN generated faces.

Detecting GAN-generated faces (especially faces that resemble human faces) is still a complex and challenging task. Various methods have been developed, including: deep learning methods for pattern recognition, but with insufficient interpretability; physical methods for detecting differences in ambient light reflection, but with the risk of false positives; physiological methods for analysing facial feature anomalies; and insufficiently researched human vision for recognising fake faces, which needs to be further explored. (Tsao and Livingstone, 2008).

Despite the increasing advances in GAN technology for image generation and the fact that researchers have developed a variety of detection methods, the detection of GAN-generated art images is still an unexplored area.

### 2.4 Assessing Image Authenticity: Human vs. Computer Vision Systems

There are perceptual differences between human and computer vision systems when assessing image veracity. Human intelligence, consisting of complex functional integration of brain evolution interacting with the environment, supports abstract thinking and problem solving (Baeza-Yates and Villoslada, 2022). The human visual system exhibits a high degree of specialisation and complexity in facial recognition and processing (Liu et al., 2022). The human brain has specialised mechanisms for facial recognition and tends to process as a whole rather than breaking down faces into separate parts (Baeza-Yates and Villoslada, 2022). Comparatively, AI systems, while making significant advances, lack the concrete experience of human brain, body and environment interaction (Lyu et al., 2022). Computer vision systems, on the other hand, rely on algorithmic models to automatically assess realism. For example, StyleGAN3 solves the problems of traditional GANs, such as texture sticking, by improving signal processing and enhancing image detail control to produce more natural and coherent images, which are particularly suitable for video and animation. It introduces Fourier features and a design that takes into account variability such as rotation to ensure image realism (Karras et al., 2021). This shows the contrast between the overall perception of human vision and the details of AI processing. At the same time, the adaptability of human facial processing may still outperform existing computer vision algorithms.

Lu et al. (2023) conducted HPBench experiments using the newly developed Fake2M dataset to measure the ability of humans and AI models to discriminate between real and AI-generated images. Whilst the human's 61.3% accuracy in distinguishing between these images suggests a lack of familiarity with AI-generated content, it also highlights an educational gap that should be addressed. Furthermore, the persistent 13% error rate of the AI model under similar conditions reveals the shortcomings of current AI in authenticity detection and the need for further research into algorithmic enhancements to improve accuracy. Furthermore, the profound influence of social and cultural contexts on human visual perception opens the way to investigate how humans process visual information and how AI systems are sensitive to these factors. These perspectives are closely related to the research gaps that my research intends to fill, in particular an in-depth study of the effectiveness and limitations of GAN discriminators in recognising and evaluating art images, as well as their potential adaptations in different socio-cultural landscapes.

The results of existing research are crucial for understanding the potential of AI in image authenticity assessment and the profound differences between human and machine perceptions of image authenticity. Such comparisons provide a basis for investigating differences in the perception of image verisimilitude between humans and AI. However, there is no specific research on artistic images, which is a relatively unexplored area.

### 2.5 Chapter Summary

The research in this chapter delves into the role of images in modern information dissemination, the application of machine learning in artistic image generation, and the differences in the assessment of image authenticity between human and computer vision systems. Through these discussions, this paper reveals several notable research gaps.

First, there is a lack of in-depth research on the dichotomous scoring function of Generative Adversarial Network (GAN) discriminators. Although GAN discriminators play a key role in image generation and validation, their efficacy and limitations in art image recognition and evaluation have not been fully explored. This constitutes one of the major gaps in my research.

Secondly, the effectiveness of recognising art-based images is an uncharted territory. Current techniques and research have focussed on the generation and detection of non-artistic images, while there is a relative lack of assessment of the authenticity of artistic images.

Given these research gaps, the goal of my research focuses on exploring the differences between humans and Generative Adversarial Networks (GANs) in the perception of image authenticity, particularly in the comparison of the perception of authenticity of images of portraits. This thesis aims to provide researchers with new directions on how to extract knowledge of the physical world from human perception through this research.

Another core of this research lies in the expression of the work, exploring how humans should adapt and act in the face of unstoppable technological advances. As AI continues to advance in the field of art making, we are challenged to redefine art, authenticity and the role of humans. This study aims to provide insights into these important issues and to guide future research directions.

## 3. Methodology

In this study, I adopt a mixed research approach which combines quantitative and qualitative techniques to comprehensively assess and compare the ability of machines and humans in recognising image authenticity.

Specifically with respect to machine discrimination of image authenticity, I focus on analysing the Metfaces (Karras et al., 2021) dataset using the discriminator provided by StyleGAN3. In a pre-trained model based on this dataset, I randomly selected 100 real images of people's portraits and also generated 100 synthetic images of people's portraits using the StyleGAN3 generator. These 200 images were discriminatively analysed and computational metrics were applied to classify and interpret the results.

In the human discrimination research section, I designed a series of questionnaires to collect quantitative data. And python was used to conduct a comprehensive analysis of these data, which reflect the performance and basis of judgement of different individuals in recognising the authenticity of images. In addition, I collected qualitative data through one-on-one interviews which helped me to understand the participants' thinking patterns and mental processes during image authenticity judgements. I used a thematic analysis approach to interpret the interviews to reveal the intuition and logic that humans rely on when judging the authenticity of images.

The application of this mixed-methods approach aims to capture the multidimensional nature of both machine and human judgements of image authenticity. Through this approach, I attempt to assess the effectiveness of machine learning models in mimicking human visual and cognitive abilities and explore their potential limitations. At the same time, it allows me to explore how humans adjust their judgement criteria when confronted with images produced by advanced generative algorithms.

## 4. Implementation

### 4.1 Extracting Output Data from the Discriminator

This section describes the implementation of the discriminator (D) of StyleGAN3 in assessing image authenticity. Firstly, under the guidance of our mentor, we have successfully extracted the code for the authenticity scoring part of the discriminator on the input image from the public source code of StyleGAN. The details of this code are given in the Appendix. We used the ffhq dataset (Karras et al., 2021) for our experiments, comparing real images from the dataset with images generated by StyleGAN3 using pre-trained models from the same dataset.

In our experiments, we found that the output scores of D ranged from -1 to 1, which is different from the general expectation (that the scores of D should be between 0 and 1, and that the closer to 1 means that the image is more realistic). By further testing the ffhq dataset, we unexpectedly found that images are more likely to be considered real when D's output is close to 0, while outputs close to 1 are considered generated. This is contrary to the description in Dimitrov and Luo (2018).

To explore this phenomenon, I tested multiple StyleGAN3 models trained using different datasets, including metfaces, WikiArt (Pinkney, 2021), and Lhq (Pinkney, 2021a). The results showed that the underlying theory of D did not apply to these datasets and that the D output data varied significantly across the datasets.

Further research pointed out that different datasets and models result in different output scores for the StyleGAN discriminator, with higher scores representing images as real and lower scores representing images as faked (GitHub, n.d.). The results of my experiments verified this. Therefore, I chose the metfaces dataset for in-depth analysis due to the fact that it is officially released by Nvidia and its artistic portraits represent faces to a certain extent, which has less impact on the results.

During the analysis process, I generated 100 synthetic images of character portraits using StyleGAN3 generator and numbered them as seed0001-0100.Similarly, I randomly selected 100 real images of character portraits from the metfaces dataset and numbered them as MF1-100.Subsequently, I used the discriminator on each of these 200 images to scored and recorded the results.

I collated this data, mixing the generated and real images, and arranged them in descending order of score. The results of the data collation are presented in the attached table.

### 4.2 Data Analysis and Conclusions

In the Data Analysis and Conclusion section, I cleaned and analysed the data. From the 200 valid numerical scores, an average score of approximately -1.34 was calculated, with scores ranging from a low of -5.24 to a high of 3.31. Negative scores accounted for 82 per cent of the total number of scores, indicating that despite the fact that most scores were negative, a significant proportion of the overall scores were still below zero. Based on this it is concluded that the output scores of D show some regularity in a single dataset, despite the fact that the output scores of the discriminator are not consistent with its underlying theory. In particular, in models based on the metfaces dataset, a lower output score for D means that the image is more likely to be generated and vice versa. With the help of the scoring threshold theory in Computed Metrics (Drakesmith et al., 2015), I determined a score threshold, i.e., the median of D's output scores for these 200 images, to differentiate between real and generated images for this dataset. Thus, in my experimental setup, images with scores higher than -1.6163 are considered real, while images with scores lower than -1.6163 are considered fake.

## 5. A Project Write-Up

In order to comprehensively assess human and AI performance in image recognition and judgemental authenticity, this paper adopts a research design that combines quantitative and qualitative methods. This involved not only collecting and analysing quantifiable data through questionnaires, but also included person interviews as a complementary means of gathering qualitative information to gain deeper insights.

### 5.1 Theoretical Background

Human perceptions of image verisimilitude arise from a complex interplay of visual cues and cognitive processes. Research has shown that humans rely on certain features to judge image verisimilitude, such as facial symmetry, texture detail, and contextual coherence (Freire, A. & Lee, K. 2001). Also, this paper refers to the visual cues used in the questionnaire used by Tahir et al. in conducting a comparison of perceptual differences between human and deepfake generated videos.

For the question section of the experimental design, several aspects that are crucial for judging the authenticity of an image were taken into account in the spirit of readability but without losing important researchable information: facial features and symmetry, naturalness of expression, skin texture and detail, lighting and shadows, background and environment, clothing and accessories, hairstyles and texture, and overall coherence.

### 5.2 Experimental Design and Sample Selection

Based on the theoretical underpinnings mentioned above, I designed a questionnaire (see Appendix) to conduct a human experiment to facilitate the acquisition of quantitative data. Participants were not restricted by age, gender or educational background. The experiment was presented as a blind test to ensure that participants did not know whether they were evaluating real images or StyleGAN3 generated images.

In terms of experimental sample (image) selection, based on the findings in 4.2, I extracted 20 images from the table. These 20 images are 10 real and 10 generated respectively. Meanwhile, five of the 10 real images are images that the machine considers fake. And vice versa. I then tested a simple human experiment with these 20 images and found that because of the large sample size, it reduces the population's participation and concentration in the experiment. So I randomly selected 10 of these images as the final sample for the experiment.

一群人的照片

描述已自动生成

Figure 1: Experimental Samples and Their Numbering

### 5.3 Results Analysis

This study conducted a detailed analysis of 59 survey responses using Python, uncovering patterns and differences in image authenticity perception between humans and AI. The conclusions benefited from multi-angle data processing and visual presentation of the results.

R1: Human vs. AI Accuracy

图表

描述已自动生成

Figure 2: Accuracy of AI and Human judges

In order to assess the consistency and differences in the perception of authenticity between human judges and AI, a comparison of the accuracy of the two was made. The results showed that the AI system had achieved benchmark levels of image authenticity judgement, but did not outperform the best human judges. Among human judges, there was considerable variation in the ability to discern image authenticity, which may reflect individual differences in processing visual cues.

R2: Correlation between Reviewers' Backgrounds and Accuracy

图表, 条形图

描述已自动生成

Figure 3: Analysis of the Correlation between Reviewers' Backgrounds and Accuracy

In exploring the association between a reviewer's background and the accuracy of their judgement, we found that those reviewers with a background in art and experience with AI-generated images performed best in determining the authenticity of art images. This may reflect their deeper understanding of the artistic elements and characteristics of AI-generated images.

R3: Visual Cues Relied Upon by Humans for Judging Image Authenticity

图表

描述已自动生成

Figure 4: Visual Cues Analysis

After a detailed analysis of the visual cues used by each participant to determine the authenticity of an image, we found that humans rely heavily on the overall impression or harmony of an image to determine its authenticity in this judgement process. As shown in Figure 4, participants' judgements of image authenticity were based primarily on overall impression and artistic elements, while also considering specific visual details. The reliance on "Overall coherence" suggests that participants were intuitively assessing the harmony between various elements in the image, rather than relying on any single defining feature.

R4: Frequency of Cues Used in Authenticity Image Recognition

图表, 条形图, 直方图

描述已自动生成

Figure 5: Frequency of cues used in authenticity image recognition

The analysis of Figure 5 indicates that "Overall coherence" is pivotal for accurately identifying authentic images. It reflects the participant's ability to evaluate the harmony of an image's elements. Features like "Skin texture", "Light and shadow", "Hair", and "Hair texture" are also crucial, as they directly contribute to the image's realism. The significance of "Facial expression" and "Eyes" aligns with the understanding that these are essential in reflecting authenticity.

However, "Overall coherence" also leads the cues in misidentified images, implying that while it's a vital indicator, it can be misleading if misinterpreted. Incorrect judgments often involved "Skin texture" and "Light and shadow", suggesting that errors in interpreting these aspects can result in false authenticity assessment. Misinterpretations related to the "Face" and "Content" indicate that these elements, while important, can be ambiguous and lead to deception.

The findings imply that the accuracy of authenticity detection is not merely about the presence of certain visual cues but also how they are perceived in the image's overall context. Elements like 'accessories', 'clothing', and 'hairstyle' were less significant in determining authenticity, suggesting they are not as influential in the judgment process. This underscores the complexity of image authentication, where evaluative skills must extend beyond surface details to a more nuanced interpretation of the image's coherence and context.

R5: Human vs. AI Performance in Specific Scenarios

图表, 条形图

描述已自动生成

Figure 6: Human vs. AI Performance Comparison in Specific Scenarios

A comparison of human and AI performance in specific scenarios revealed that while the AI recognised real images with only 50% accuracy, human judges performed better when the AI incorrectly judged the generated images as real. This trend may reflect the superiority of humans in recognising certain features of AI-generated images.

### 5.3 Brief Discussion and Conclusion

This chapter demonstrates the capabilities and limitations of humans and AI in judging the authenticity of images by analysing data from multiple perspectives. The analyses show that although the AI achieves the benchmark level in some cases, human reviewers show higher accuracy in specific scenarios, especially in the case of misjudgments. The background of the reviewers, particularly their artistic background and experience with AI image generation, significantly influenced the accuracy of their judgements. In addition, the study highlighted the importance participants placed on overall coherence and artistic details when assessing image authenticity, suggesting that the overall impression is as important as specific details in the perception of image authenticity. Through these in-depth analyses, this chapter provides important insights into understanding the interplay between humans and AI in the assessment of image authenticity. It also emphasises the need for humans to be critically aware when encountering machine-generated content.

一群人的照片

描述已自动生成

女人看着电脑

描述已自动生成

Figure7 & 8: Interviews VCR Screenshots

## 6. Evaluation

In order to provide a more in-depth and comprehensive perspective of the research, this thesis supplemented the collection and analysis of questionnaire data with interviews to gather non-numerical information to gain qualitative insight. In analysing the transcribed transcript data from the audio recordings of the interviews, I used Thematic Analysis (TA), which is a method used to develop, analyse, and interpret patterns in qualitative data sets (Braun & Clarke, 2021). This method was chosen because it not only provides rich descriptive detail, but also because it is a technique that is operationally robust for individual researchers (Bryan-Kinns, Wang & Wu, 2018). The initial data processing steps included using Word's transcription function to transcribe the interviews, followed by initial coding of the transcripts using inductive thematic analysis to categorise the data into themes, and reviewing and defining these themes (Ford & Bryan-Kinns, 2022). In addition, I attempted to use TA to analyse how people from different professional backgrounds judge the authenticity of images of portraits and the effect of different media on judgements of image authenticity.

### 6.1 Interview Format

In selecting the survey format, I initially considered converting the questionnaire to an interactive web-based format, similar to a participatory game, with the aim of widening the scope of the survey and increasing participation. Utilising the Framer application, I successfully created a basic interactive web-based game, details of which are provided in the Appendix.

I also attempted to draw on Actor-Network Theory (ANT) and Susan Lepselter's analysis referred to by Dányi (2017) to frame the interview format, in particular the way in which presenting images of experiments uses different mediums.ANT theory suggests that all elements (including non-human entities such as technology or mediums) are actors 'actors' in a network, which means that any medium used to present the images is not just a tool to convey information, but they also influence the transmission and reception of information. Based on this theory, I chose different media to display the experimental images in order to investigate how different ways of viewing influence people's judgement on the authenticity of the images.

In contrast to ANT is Susan Lepselter's analysis, which argues that the art of perception lies in how to connect images in a way that is situated in chaos, rather than merely looking for the hidden integrity behind the fragments. I liken this analysis to a kind of jigsaw puzzle, where the questionnaire acts as a display medium, as if one were standing outside the puzzle trying to find the meaning of the whole. If people are to feel that they are at the centre of these chaotic jigsaw pieces, the images need to be presented in a holistic manner. Therefore, I chose to print out the physical experimental images and present them in a chaotic format to see how people's judgements about the authenticity of the images differed when they were at the centre of the jigsaw story, experiencing the chaos and complexity of it.

The finalised interview format involved inviting people from a variety of backgrounds to participate in interviews, first making judgements about the physical printed images and undergoing a basic qualitative interview, and then experiencing a web-based version of the image game. At the end of this process, interviewees were asked about the difference in perception between the images presented in the two media and how this difference affected their judgement and understanding of the content of the images.

### 6.2 Thematic Analysis

The interviews with the four participants, which I collated into 21 initial codes using inductive TA, were reviewed and categorised into seven themes that illuminated the multifaceted nature of judgements about the authenticity of portrait images. Here, I present an analytical narrative of these themes, synthesising insights gained from participants with diverse artistic backgrounds and exposure to AI-generated images.

T0: First impressions and self-assessment of authenticity judgements

The assessor's first impressions form the basis of his or her authenticity judgements. These impressions often stemmed from the direct gaze exhibited by the AI-generated portraits, which lacked the emotional depth characteristic of real artworks. Participants felt that genuine artworks exude the emotion of the artist, suggesting a deeper connection between the creator's intention and the viewer's perception, which is absent in AI-generated images.

T1: Emotional engagement

Emotion played a crucial role when participants engaged with the artwork. They reported instinctive reactions to the real artwork, which seemed to resonate emotionally and cognitively. In contrast, AI-generated images were described as having a robotic and lifeless aura that lacked the emotional power that participants associate with authentic artistic expression.

T2: Artistic elements

The identification of authenticity is heavily influenced by artistic elements such as brushwork, texture, colour and context. Authentic artworks have a narrative and stylistic coherence that cannot be convincingly replicated in AI-generated images. Participants noted that AI-generated images have a uniform background and lack the storytelling quality inherent in real paintings.

T3: Common Sense Knowledge

Participants relied on common sense knowledge to recognise inconsistencies in human features and lighting in the images. Of note were discrepancies in age-appropriate features; for example, a child's face should not have wrinkles. This mismatch demonstrates the inability of AI to understand and replicate the complexity of human aging.

T4: Understanding of AI

Perceptions of AI were shaped by its portrayal of overly perfect portraits, suggesting that its vision is too idealised to capture the reality of human imperfections. Participants felt that the AI was unable to replicate the subtle details that professional artists incorporate into their work, such as the brushstrokes that reveal the artist's technique.

T5: Personal background

Participants' backgrounds, especially artistic backgrounds, determined their ability to recognise real images from AI-generated images. Professional artists or those with extensive artistic experience were able to recognise clues to authenticity that others may not have been able to recognise based on their understanding of artistic techniques and styles.

T6: Images and Media

The medium (digital or physical) in which an image is presented significantly influences the judgement process. Digital media, with its clarity and resolution, made it easier for participants to detect fakes, while printed images posed more challenges due to the quality of the print that could lose detail.

The thematic analysis illuminates the complex interplay between human intuition and technical expertise in image authenticity assessment. It shows that while AI has made significant advances in image generation, it is still unable to replicate the human qualities and emotional depth that characterise genuine artworks. This analysis not only contributes to the understanding of human-computer interaction in the art domain, but also highlights areas where AI can be further refined to meet the nuanced expectations of human assessors.

## 7. Discussion

### 7.1 Integration of Survey and Thematic Analysis Results

Through Chapters 5 and 6 of this paper, I have come to a number of conclusions, and the findings (R1-R5) and thematic analyses (T0-T6) reveal different but interrelated findings that shed light on the differences between humans and Generative Adversarial Networks (GANs) in the perception of image authenticity, particularly with regard to the perception of authenticity in artistic images. Here is how these insights intertwine with the overall research questions and objectives:

R1 and T0 reflect the interaction between first impressions and self-assessments of authenticity judgements. Comparisons of the accuracy of authenticity judgements between AI and human participants echo the reliance on first impressions, which are formed based on the presence or absence of direct gaze and perceived emotional depth. This aspect highlights the research question by emphasising the different cues used by humans and GANs in perceiving image verisimilitude.

R2 and T5 show that individuals with a background in art and experience with AI-generated images have an increased ability to recognise the authenticity of artistic images. This finding is consistent with the goal of providing direction for researchers to extract knowledge of the physical world from human perception, emphasising the impact of individual artistic expertise on the ability to distinguish between real and AI-generated images.

R3's reliance on overall coherence and skin texture as key visual cues for judging authenticity intersects with T2's artistic elements and T3's common sense. These findings suggest that overall impression and specific visual details are critical for authenticity judgements, supporting the research objectives by demonstrating how human perception of artistic images provides support for knowledge extraction.

T6 and R4 jointly demonstrate that the medium and visual cues play a pivotal role in the assessment of image authenticity. The clarity provided by digital versus physical media directly affects the accuracy of interpreting these cues. The inherent high resolution of digital media enables participants to discern subtle details with greater ease, aiding in the judgment of overall Coherence and the veracity of textures. In contrast, the physical medium, prone to detail loss, may lead to erroneous judgments as these cues become less conspicuous and more challenging to interpret.

R5 and T4 showed that humans outperformed the AI in cases where the AI misjudged the generated image as real, reflecting an understanding of the limitations of the AI and human intuition in capturing flaws not recognised by the AI.

Each theme in the thematic analysis adds depth to the quantitative findings in the analysis of results, providing insights into the reasons behind the patterns observed in the data. For example, individuals with artistic backgrounds and AI experience had better judgement accuracy, possibly due to their nuanced understanding of artistic elements and AI capabilities, as reflected in T2 and T4. Similarly, the reliance on overall coherence and specific visual details (e.g., skin texture) is consistent with the qualitative finding that individuals' authenticity assessments are based not only on individual deterministic features, but also on the overall impression of the image associated with T2 and T3.

These connections illustrate the complexity of judging the authenticity of artistic images and highlight the importance of considering human intuition and technical aspects when assessing the performance of AI compared to human judgement. The findings emphasise the need for critical awareness when encountering machine-generated content and suggest areas where AI can be improved to meet the subtle expectations of human assessors.

### 7.2 Reflection

Embarking on this scholarly inquiry, I ventured into the blurred lines at the intersection of human cognition and artificial intelligence, particularly in discerning the authenticity of images. This intellectual journey illuminates the inherent complexity of distinguishing the real from the artificially generated - a task that requires the analytical and emotional capacities of the human mind. The endeavour is as much about understanding the evolving capabilities of AI as it is about appreciating the unparalleled human capacity for synthesis - the way we weave visual cues into a tapestry of judgement. These insights have not only enriched my understanding, but have also given me a profound respect for the depths and subtleties of human perception that remain elusive and mysterious to the binary world of AI.

### 7.3 Future Work

The findings and insights from this study pave the way for multiple avenues of future research. One promising direction lies in the development of AI systems that incorporate quasi-human emotional and aesthetic judgements, possibly through advanced neural network architectures or hybrid models that integrate human feedback. In addition, exploring how cultural and psychosocial factors influence the perception of image authenticity could further enrich AI's understanding of parahuman authenticity judgements. Investigating how different presentation media affect the perception of authenticity and developing AIs that can adapt their judgements to various media also marks an important step forward. Finally, expanding the scope of the research to cover a wider range of demographics and a wider range of image types may help create more versatile and robust AI systems.

## 8. Conclusions

This thesis unravels the complex labyrinth of authenticity perception in which artificial intelligence is a sincere but inexperienced participant. The human capacity to combine emotional empathy, artistic appreciation and situational intelligence in authenticity judgements remains unparalleled. This study highlights the richness of personal experience in shaping perceptions of authenticity, suggesting that AI may mimic but never fully embody the profound range of human judgement. As AI innovation continues to evolve, a nuanced grasp of human perception must guide its progress, ensuring that AI advances in the arts and other fields are not only computationally competent, but also philosophically deep and humanistically resonant.

## 9. Appendix

All appendix can be found at the following links: <https://github.com/wwdddq/MSc-Advanced-Project/tree/main>

## 10. Reference

Acerbi, A., Altay, S. and Mercier, H. (2022). Research note: Fighting misinformation or fighting for information? *Harvard Kennedy School Misinformation Review*. doi:<https://doi.org/10.37016/mr-2020-87>.

Baeza-Yates, R. and Villoslada, P. (2022). Human vs. Artificial Intelligence. doi:<https://doi.org/10.1109/cogmi56440.2022.00016>.

Braun, V. and Clarke, V. (2021). *Thematic Analysis: A Practical Guide*. [online] *Google Books*. SAGE. Available at: <https://books.google.co.uk/books?id=eMArEAAAQBAJ&pg=PT17&hl=zh-CN&source=gbs_toc_r&cad=2#v=onepage&q&f=false> [Accessed 19 Nov. 2023].

Broad, T., Leymarie, F.F. and Grierson, M. (2020). *Amplifying The Uncanny*. [online] arXiv.org. doi:<https://doi.org/10.48550/arXiv.2002.06890>.

Brock, A., Donahue, J. and Simonyan, K. (2018). *Large Scale GAN Training for High Fidelity Natural Image Synthesis*. [online] arXiv.org. Available at: <https://arxiv.org/abs/1809.11096>.

Broderick, R. (2019). *Here’s A Detailed Look At How Russian Trolls Used Social Media To Meddle In The 2016 Election, According To The Mueller Report*. [online] BuzzFeed News. Available at: <https://www.buzzfeednews.com/article/ryanhatesthis/mueller-report-internet-research-agency-detailed-2016>.

Bryan-Kinns, N., Wang, W. and Wu, Y. (2018). Thematic Analysis for Sonic Interaction Design. *BCS Learning & Development*. doi:<https://doi.org/10.14236/ewic/hci2018.214>.

Cambridge Dictionary (2023). *endow*. [online] @CambridgeWords. Available at: <https://dictionary.cambridge.org/dictionary/english/endow?q=endowed> [Accessed 18 Nov. 2023].

Curtis, A. (2016). *HyperNormalisation (2016 + subs) by Adam Curtis - A different experience of reality FULL DOCUMENTARY*. *YouTube*. Available at: <https://www.youtube.com/watch?v=fh2cDKyFdyU>.

Dányi, E. (2017). *Is everything connected?*.

Dimitrov, D.M. and Luo, Y. (2018). A Note on the *D*-Scoring Method Adapted for Polytomous Test Items. *Educational and Psychological Measurement*, 79(3), pp.545–557. doi:<https://doi.org/10.1177/0013164418786014>.

Drakesmith, M., Caeyenberghs, K., Dutt, A., Lewis, G., David, A.S. and Jones, D.K. (2015). Overcoming the effects of false positives and threshold bias in graph theoretical analyses of neuroimaging data. *NeuroImage*, [online] 118, pp.313–333. doi:<https://doi.org/10.1016/j.neuroimage.2015.05.011>.

Ford, C. and Bryan-Kinns, N. (2022). Identifying Engagement in Children’s Interaction whilst Composing Digital Music at Home. *Creativity and Cognition*. doi:<https://doi.org/10.1145/3527927.3532794>.

GitHub. (n.d.). *The pre-trained sytleGAN2 discriminator can not discriminate fake and real images · Issue #180 · NVlabs/stylegan3*. [online] Available at: <https://github.com/NVlabs/stylegan3/issues/180> [Accessed 21 Nov. 2023].

Glynn, P. (2023). Sony World Photography Award 2023: Winner refuses award after revealing AI creation. *BBC News*. [online] 17 Apr. Available at: <https://www.bbc.co.uk/news/entertainment-arts-65296763>.

Gonzalo, M. (2023). *La próxima vez no te servirá contar los dedos para saber si una imagen es falsa en internet*. [online] Newtral. Available at: <https://www.newtral.es/como-detectar-imagenes-videos-audios-deepfakes-generados-ia/20230331/>.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014). *Generative Adversarial Networks*. [online] arXiv.org. Available at: <https://arxiv.org/abs/1406.2661>.

Higgins, E. (2023). *Making pictures of Trump getting arrested while waiting for Trump’s arrest.* [online] Twitter. Available at: <https://twitter.com/EliotHiggins/status/1637927681734987777>.

Kahn, G. (2023). *Will AI-generated Images Create a New Crisis for fact-checkers? Experts Are Not so Sure*. [online] Reuters Institute for the Study of Journalism. Available at: <https://reutersinstitute.politics.ox.ac.uk/news/will-ai-generated-images-create-new-crisis-fact-checkers-experts-are-not-so-sure>.

Karras, T., Aila, T., Laine, S. and Lehtinen, J. (2017). *Progressive Growing of GANs for Improved Quality, Stability, and Variation*. [online] arXiv.org. Available at: <https://arxiv.org/abs/1710.10196>.

Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J. and Aila, T. (2021). Alias-Free Generative Adversarial Networks. *arXiv:2106.12423 [cs, stat]*. [online] Available at: <https://arxiv.org/abs/2106.12423>.

Karras, T., Laine, S. and Aila, T. (2018). *A Style-Based Generator Architecture for Generative Adversarial Networks*. [online] arXiv.org. Available at: <https://arxiv.org/abs/1812.04948>.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J. and Aila, T. (2020). Analyzing and Improving the Image Quality of StyleGAN. *arXiv:1912.04958 [cs, eess, stat]*. [online] Available at: <https://arxiv.org/abs/1912.04958>.

Liu, S., Huang, S., Fu, W. and Jerry Chun‐Wei Lin (2022). A descriptive human visual cognitive strategy using graph neural network for facial expression recognition. *International Journal of Machine Learning and Cybernetics*. doi:<https://doi.org/10.1007/s13042-022-01681-w>.

Lu, Z., Huang, D., Bai, L., Liu, X., Qu, J. and Ouyang, W. (2023). *Seeing is not always believing: A Quantitative Study on Human Perception of AI-Generated Images*. [online] arXiv.org. Available at: <https://arxiv.org/abs/2304.13023>.

Lyu, Y., Wang, X., Lin, R. and Wu, J. (2022). Communication in Human–AI Co-Creation: Perceptual Analysis of Paintings Generated by Text-to-Image System. *Applied Sciences*, 12(22), p.11312. doi:<https://doi.org/10.3390/app122211312>.

Pinkney, J. (2021a). *alis/lhq.md at master · universome/alis*. [online] GitHub. Available at: <https://github.com/universome/alis/blob/master/lhq.md> [Accessed 21 Nov. 2023].

Pinkney, J. (2021b). *StyleGAN 3*. [online] lambdalabs.com. Available at: <https://lambdalabs.com/blog/stylegan-3>.

Porres, D. (2021). *Discriminator Synthesis: On reusing the other half of Generative Adversarial Networks*. [online] arXiv.org. doi:<https://doi.org/10.48550/arXiv.2111.02175>.

Simon, F.M., Altay, S. and Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. doi:<https://doi.org/10.37016/mr-2020-127>.

Tahir, R., Batool, B., Jamshed, H., Jameel, M., Anwar, M., Ahmed, F., Zaffar, M.A. and Zaffar, M.F. (2021). Seeing is Believing: Exploring Perceptual Differences in DeepFake Videos. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. doi:<https://doi.org/10.1145/3411764.3445699>.

Tsao, D.Y. and Livingstone, M.S. (2008). Mechanisms of Face Perception. *Annual Review of Neuroscience*, [online] 31(1), pp.411–437. doi:<https://doi.org/10.1146/annurev.neuro.30.051606.094238>.

University, S. (2017). *Stanford study examines fake news and the 2016 presidential election*. [online] Stanford News. Available at: <https://news.stanford.edu/2017/01/18/stanford-study-examines-fake-news-2016-presidential-election/#:~:text=Stanford%20study%20examines%20fake%20news> [Accessed 16 Nov. 2023].

Utterson, A. (2023). Software, Self, Society: The Computer Histories of Adam Curtis. *Quarterly Review of Film and Video*, pp.1–18. doi:<https://doi.org/10.1080/10509208.2023.2247312>.

Wang, X., Guo, H., Hu, S., Chang, M.-C. and Lyu, S. (2023). *GAN-generated Faces Detection: A Survey and New Perspectives*. [online] arXiv.org. doi:<https://doi.org/10.48550/arXiv.2202.07145>.

William John Mitchell (2001). *The reconfigured eye : visual truth in the post-photographic era*. Cambridge/Massachusetts ; London: Mit Press.

Yang, K. (2022). *Landscape Art Image Style Reconstruction Algorithm Based on Machine Learning*.