How Well Can Young Women Distinguish AI vs. Real Dating Profiles?

Sara Danielsson (sadan@kth.se), Olivia Flodkvist (oliviafl@kth.se), Olivia Martin (oamartin@kth.se), Julia Nordqvist (junordq@kth.se), Jingwen Zhuang (zhuan@kth.se)

Supervisor: Petra Jääskeläinen

Abstract

This study examines how participants detect AI-generated versus real dating profile images by identifying visual cues. Participants achieved high accuracy (87.5%) in recognizing AI-generated profiles and focused on features such as smooth skin, lighting inconsistencies, and symmetrical imperfections. Findings reveal that participants' detection strategies relied on spotting unnatural details, supporting the uncanny valley effect. Future work should explore broader demographics and test evolving AI tools to track changes in detection accuracy and techniques.

Keywords: AI-generated images, AI-hyperrealism, dating profiles, perception, uncanny valley.

1. Introduction

AI development has progressed rapidly in the 21st century, driven by advancements in computing power, data storage, and the availability of large-scale datasets (Tewari and Bose, 2023). One significant area of AI, face synthesis, has achieved remarkable strides, leveraging advanced deep learning techniques and extensive datasets to generate realistic human faces (Huang et al., 2024).

Understanding whether individuals can distinguish AI-generated faces from real ones and assessing the trustworthiness assigned to synthetic faces are critical for addressing potential risks, such as deception and manipulation. This study focuses on these challenges within the context of dating profiles, where profile pictures are central to modern social interactions. Research has shown that profile photos significantly influence first impressions and decisions about trust and compatibility in online dating, often determining the success of forming connections (White et al., 2017). By exploring the ability to discern real and synthetic faces and the trustworthiness attributed to them, this research aims to provide valuable insights into human perception and the societal implications of AI-generated content.

2. Background

2.1 Ability to classify AI faces and real faces

Nightingale and Farid (2022) conducted two experiments to examine the classification of synthetic versus real faces; in the first experiment, participants were asked to classify faces as real or synthetic. The average accuracy was 48.2%, which is almost equivalent to random chance, indicating significant difficulty in distinguishing between real and synthetic faces. In the second experiment, participants performed the same classification task but were provided with trial-by-trial feedback and made aware of rendering artifacts in synthetic faces. The average accuracy improved slightly to 59%. While this improvement suggests that feedback and training had some effect, the accuracy remained only slightly above chance, highlighting the limited impact of feedback. The effect of gender and race on accuracy was consistent across both experiments, with white synthetic faces being the most difficult to classify. Additionally, male white synthetic faces were classified less accurately than female white synthetic faces (Nightingale and Farid, 2022).

Huang and Mittal's (2024) study presents different findings, concluding that humans are relatively skilled at distinguishing between real and fake face images—a conclusion that contrasts with prior research. Their study investigated how individuals perceive real and fake face images using eye-tracking technology. Across 20 participants, the average recognition accuracy was 76.8%, with participants correctly identifying both real and fake images. The primary reason for errors was participants incorrectly classifying fake images as real. The study found that image backgrounds significantly influenced participants' performance. Recognition accuracy was 85% when there was a person in the background. Eye-tracking data revealed that participants scrutinized images more closely when they suspected an image might be fake. However, participants actually spent more effort examining real images than fake ones (Huang et al., 2024).

Miller and Steward (2023) conducted a study using 100 AI-generated faces and 100 closely matched human faces, where participants were tasked with distinguishing between the two. The results revealed that participants were more likely to judge White AI-generated faces as human (65.9%) than actual human faces (51.1%). This phenomenon, referred to as AI hyperrealism, is explained by face-space theory, which suggests that AI algorithms like StyleGAN2 amplify the average facial features present in their training data. This process makes AI-generated faces appear more prototypical and realistic than real human faces (Miller et al., 2023).

2.2 Perceived ability compared to actual ability

Miller and Steward (2023) also provided confidence ratings (on a scale of 0–100) for each judgment. While higher confidence was associated with fewer errors when identifying human faces, the opposite was true for AI-generated faces. Participants who frequently misclassified AI faces as human were overly confident in their incorrect judgments. This finding underscores the risks of overconfidence when identifying AI-generated faces (Miller et al., 2023).

Labajová (2023) also identified a significant gap between individuals' perceived and actual ability to recognize AI-generated content. In her survey, 87 out of 100 participants expressed complete confidence in their ability to differentiate between AI-generated and human-generated material. However, when tested with two visual and two textual examples in the questionnaire, only 20 of the 87 confident respondents accurately identified the source of the content (Labajová, 2023).

2.3 Women on online dating platforms

Online dating platforms have become very popular over the last few years (Chan, 2017). Different genders reveal varying behaviors and interaction patterns on online dating platforms (Abramova et al., 2016). Women prioritize socio-economic factors like income and stability, while men focus on physical attractiveness, reflecting evolutionary theories of mate selection. Young women, as active participants on these platforms, are especially suited for studying how individuals distinguish between

AI-generated and authentic profiles. Their selectivity and ability to interpret social indications offer valuable insights into the evolving dynamics of online dating and the broader impact of technology on interpersonal relationships.

3. Method

3.1 Research question & Aims

The research question of this study was: *"How Well Can Young Women Distinguish AI vs. Real Dating Profiles?*" The aim was to investigate whether young women, aged around 20-30, can distinguish between real and AI-generated dating profiles.

3.2 Participants

All recruited participants were women aged 19–28. Before starting the experiment, they answered questions regarding their gender, sexuality, age and ethnicity. A total of 20 women participated in the study, 18 were straight and two were bisexual. Participants were recruited in two ways: first, through acquaintances of the researchers, and second, by approaching potential participants at KTH Royal Institute of Technology. Although the research aimed to study the age range 20-30, most participants (85.7%) were aged 25 or under. The youngest being 19 and the oldest 28. Participants were from a wide range of countries, including China, Germany, South Korea, Cyprus, Romania, Poland and Sweden. Although the majority of the participants (66.7%) were from Sweden.

3.3 Creating dating profiles

A total of 20 profiles were created, evenly split between real and fake ones. The real profiles were developed using free images from Unsplash.com (Unsplash, 2024). These images were selected based on three criteria:

1) How authentic they appeared
2) The availability of at least three different pictures of the same individual
3) Whether the individuals looked to be between the ages of 20 and 30. Three pictures per person were chosen for consistency.

Afterwards, all the real images were sent to an AI image generation tool that generated AI versions based on the original pictures with certain prompts. The AI image generation tool Getimg.ai (Getimg, 2024) was chosen. It provides the interface to run the model Juggernaut XL, which is the model to generate fake images from real images in this case. The Juggernaut XL image generation model is a prominent AI tool developed through a collaboration between KandooAI and RunDiffusion (RunDiffusion, 2024). It is based on the Stable Diffusion XL (SDXL) architecture and is designed to produce high-quality, photorealistic images from textual descriptions. (Podell et al., 2023) Before the images were generated, two parts of the prompts were prepared. The first prompt was the original image as an image prompt. The second prompt was a text that described the desired picture style. The second prompt also included the age information so that the results from the model could be as close as what was for the original images.

```
Real photo. Phone photo. Young adult man.
Age X. Realistic. Extremely detailed.
```

Second prompt

During the generation process, only images that appeared to have no obvious flaws were being selected, otherwise they were re-generated. This would help lower the chance of other artifacts in the picture to influence the judgement from the participants. Although two sets of obvious flawed images are kept as a feature, since no AI image generation tools are perfect.

To ensure the profiles appeared genuine, Tinderkit.com (Tinderkit, 2024) was used during the creation process, so that every profile looked just like what they would in Tinder.

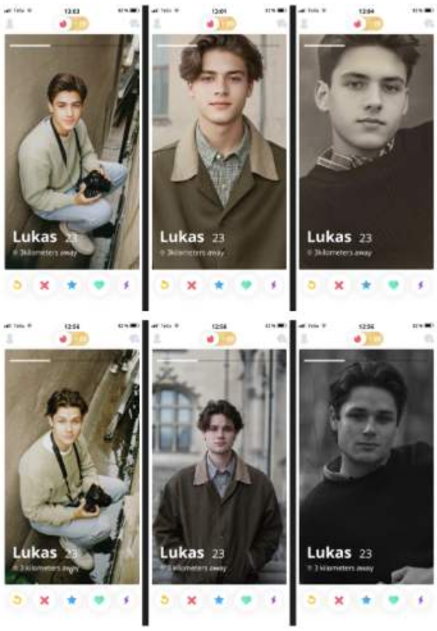


Figure 1. Comparison of a real profile (bottom) and its AI-generated counterpart (top). See all profiles in the Appendix 1.

3.4 Survey

In this study a between-subjects design was conducted, the participants were split into two equal groups, Group A and Group B. Each group answered one of two Google Forms, with everyone in Group A completing the same survey and everyone in Group B completing the other. The only difference between the surveys was which profiles from the pool of 10 were AI-generated and which weren't.

The participants were not informed of the AI generated pictures before answering the survey to avoid influencing the results. The aim of the survey was to gather quantitative data on how both the real and the AI generated dating profiles were perceived and qualitative data on why they "swiped" a certain way.

While answering the survey, each participant was presented with 3 pictures from each of the total 10 profiles. Half of the profiles had real pictures and the other half were AI generated. For each profile the participant was asked to swipe right or left, depending on their preference, with a following question about the reasoning (see figure 2). Swiping right indicated liking a profile, while swiping left meant they did not.

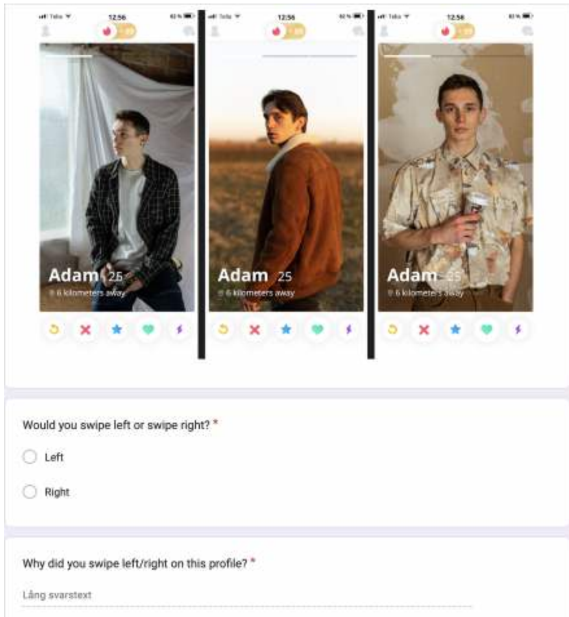


Figure 2. Overview of the swiping section design

The participants were asked which profiles they suspected were AI or not, this was done after the swiping section so the awareness of AI was not prevalent in the swiping part of the survey. The participants were also asked to rate their ability in differentiating real faces from AI faces to get a sense of the participants' own perceived ability (see figure 3).

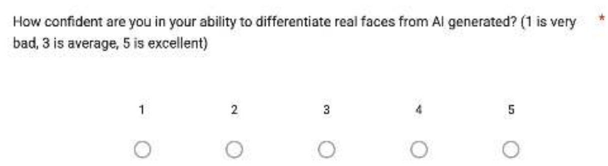


Figure 3. Question from the survey asking the participants how good they are at differentiating real faces from AI faces.

3.5 Post survey interview

In the interview conducted immediately after the survey, participants were asked about their experiences with AI-generated profiles or content. The questions could be found in Table 1. They discussed how they differentiate between real and AI-generated images and how they might interact with a dating profile they suspected was AI-generated. These questions aimed to gain insights into participants' perceptions of AI-generated dating profiles and their experiences in identifying AI-generated images. Additionally, the participants were once again presented with pictures of all the profiles and asked to motivate why or why not they thought that that profile was AI. The reason for asking this question was to allow the participant to elaborate more freely on their decision to mark some profiles as AI.

No.	Interview Question
1	Have you experienced AI generated social media / dating profiles before? If yes, do you want to elaborate?
2	How do you distinguish between AI generated profiles and real-life profiles?
3	If you suspected a profile was AI, would you still swipe right? Why or why not?
4	If you thought a profile was AI-generated, how would it change the way you see or interact with it?
5	What made you notice that it was AI generated? Any specific features?Was it different from profile to profile or was it similar features that gave it away?

Table 1. Interview questions

3.6 Data analysis

3.6.1 Qualitative data analysis - thematic analysis

The thematic analysis (Braun and Clarke, 2006) was based on the text-based responses collected from the survey and post-interview. The process was organized into three distinct parts to ensure a structured and comprehensive approach, see figure 4.

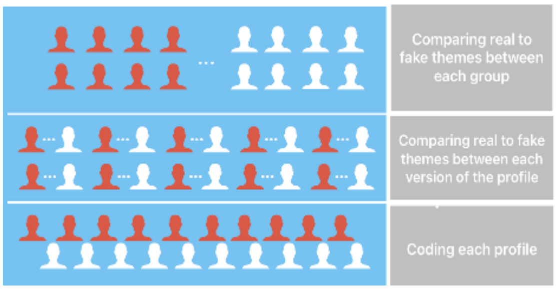


Figure 4. Process for analyzing swiping responses and individual interview questions.

The first thematic analysis examined text responses from the swiping task in the survey, where participants provided reasons for their decision to swipe or not swipe on a particular profile. The second thematic analysis examined a targeted analysis of a single interview question from the post-interview, which specifically addressed AI-generated profiles. The question was: "What made you notice that it was AI generated? Any specific features? Was it different from profile to profile or was it similar features that gave it away?". Both analysis 1 and 2 followed the structure of figure 3.

The third analysis focused on the majority of the interview questions from the post-interview, which were analyzed separately. Each question's answers were initially coded. In the second step, the codes for each question were grouped separately to identify recurring themes.

3.6.2 Quantitative data analysis - classification accuracy analysis

Two sections of our survey involve quantitative data: participants identifying whether profiles are real or fake, and participants self-rating confidence in detecting AI on a scale from one to five. For the 'real or fake profile identification' section, we, the researchers, perform a classification accuracy analysis (Pizer and Marron, 2017), including the use of a confusion matrix. For the confidence ratings, we calculate the mean value.

We use this formula for the classification accuracy analysis:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Where TP means True Positive, TN means True Negative, FN means False Positive and FN means False Negative. Specifically in our case the abbreviations mean:

- **True Positive (TP):** A participant correctly identifies an AI-generated image as "AI."
- **True Negative (TN):** A participant correctly identifies a real image as "real."
- **False Positive (FP):** A participant incorrectly identifies an AI-generated image as "real."
- **False Negative (FN):** A participant incorrectly identifies a real image as "AI."

The classification accuracy measures the proportion of correctly classified images—those identified as "AI" or "real"—out of the total number of images. This metric works well in this case because our dataset is balanced, meaning we have an equal number of AI-generated profiles and real profiles. A balanced dataset ensures that the accuracy metric is not biased toward the majority class (as can happen with imbalanced datasets).

The confusion matrix provides additional insight beyond overall accuracy. It reveals the distribution of errors and allows us to determine whether specific groups were more likely to be misclassified. For example, if there are more false positives than false negatives, it might indicate a bias toward labeling profiles as "real." Similarly, if the true positive or true negative rates are significantly higher, it could reflect participants' tendencies or limitations in distinguishing AI-generated profiles from real ones. This helps us identify patterns in the results and any potential skew in participant responses that might have affected the overall outcomes, improving the depth of our analysis.

We calculated classification accuracy at three levels: for each individual participant, for each survey group (Group A and Group B), and for all participants combined. This multi-level analysis allowed us to identify whether any specific individuals had a disproportionate impact on the results or whether the outcomes were heavily influenced by the group to which they belonged. By examining accuracy across these different levels, we could detect any potential biases or anomalies in the data.

For the confidence ratings, we calculated the mean confidence scores separately for Group A and Group B, as well as the overall mean across all participants. By comparing confidence ratings with classification accuracy on a group-by-group basis, we gain insights into potential differences between groups. This approach allows us to see whether there is a disparity between participants' confidence in their ability to identify AI-generated profiles and their actual accuracy.

4. Results

4.1 Survey - swiping section

Following are the themes identified from the survey's written responses and the post-survey interview questions about dating profiles, along with a description of each theme. The first section contains the themes for the real profiles and the next section has the themes for the AI-generated profiles.

4.1.1 Themes in real profiles

Themes in the real profiles	**Quotes from participants**
Not my type	"Too old", "Not my type"
My type	"Cool guy", "He is my type"
Good Looking	"Cute", "Hot", "Very handsome"
Don't like the photos	"Weird poses" , "Too staged"
Too serious	"Does not smile in the photos", "Photos look too serious"

Table 2. Overview of the themes in the real profiles & quotes from participants

Participants' responses to real profiles reflected personal preferences and perceptions of authenticity. Themes included "Not my type" and "My type," highlighting alignment or misalignment with individual preferences. "Good looking" emphasized physical attractiveness, while "Don't like the photos" pointed to critiques of overly professional or staged images. The theme "Too serious" captured preferences for more approachable expressions, with a lack of smiles often noted.

4.1.2 Themes in AI-generated profiles

Themes in AI-generated profiles	**Quotes from participants**
Not my type	"Looks too young", "Too hipster"
Don't like the photos	"Looks like he is wearing a filter", "Pictures too professional", "His eye gaze looks empty"
Good looking	"Cute", "Nice smile", "I like his hair"
Something is off	"Right picture looks like AI, not sure if it is a real person", "Photos look professional, like AI".

Table 3. Themes in AI-generated profiles & quotes from participants.

For AI-generated profiles, participants identified several critical themes. "Not my type" and "Good looking" echoed real profiles, but "Don't like the photos" was more prevalent due to perceived artificiality or excessive refinement. The theme "Something is off" uniquely captured participants' sensitivity to subtle flaws, such as unnatural lighting or lifeless expressions, often triggering suspicions of AI generation.

4.1.3 Cross comparison

Comparing the themes revealed that while "Good looking" appeared in both real and AI-generated profiles, real profiles were more likely to resonate with participants' preferences through nuanced traits represented by the "My type" theme. Participants were also more forgiving of imperfections in real profiles, which added to their authenticity. In contrast, AI profiles faced greater scrutiny, with "Something is off" dominating critiques and underscoring the uncanny valley effect. Real profiles' approachability and depth of expression set them apart from AI-generated ones, which were often seen as too polished or artificial.

4.2 Survey - follow up section

The participants demonstrated a strong ability to distinguish between AI-generated and real dating profiles, with an overall accuracy of 87.5%. Group B showed a slightly higher accuracy of 90%, while Group A achieved an accuracy of 85%. This suggests that women in their 20s and 30s can generally differentiate between the two types of profiles.

Analysis of the confusion matrix revealed differences in how accurately participants identified each type of profile (see figure 5). Participants correctly labeled real dating profiles as "real" 92 instances, doing slightly better than their ability to recognize AI-generated profiles as "AI", which they did 83 instances.

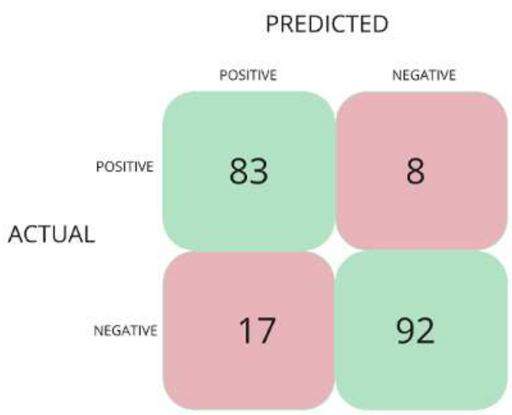


Figure 5. Combined results confusion matrix for both groups.

However, some errors were observed. AI-generated profiles were mistakenly identified as real in 17 instances, meaning participants thought these profiles were genuine when they were actually AI-generated. On the other hand, real profiles were incorrectly identified as AI in 8 instances, indicating participants believed these genuine profiles were created by AI.

These findings suggest that real profiles are somewhat easier to recognize as "real" than AI-generated profiles are to identify as "AI." This variance may highlight the increasing realism of AI-generated content, making it more challenging to distinguish from authentic profiles.

When it comes to participants self-rating their confidence in detecting AI on a scale from 1 to 5, the mean value was 3.3. Group B had a slightly higher confidence with a mean of 3.7, while Group A had a lower average of 2.9. These mean values are somewhat above average, though not significantly so. Given their overall high accuracy, it seems they may have underestimated their ability.

4.3. Interview - general AI questions

Most participants were familiar with AI-generated profiles and content on social media platforms such as Instagram, TikTok, and Snapchat, including AI models, influencers, and various types of memes. However, some participants had no direct experience with AI-generated content, while others had only heard about it without encountering it firsthand.

When it came to distinguishing AI-generated profiles from real ones, participants identified key indicators such as unrealistic eyes, overly filtered images, flawless and poreless skin, and lighting that appeared too perfect or unnatural (see figure 6). These features contributed to an overall appearance that seemed excessively idealized. However, many participants mentioned that it can be hard to distinguish, especially nowadays when AI is getting quite good at generating images.

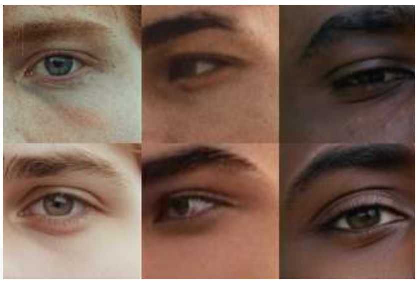


Figure 6. Detail comparison between real profile images (top) and their AI replicas (down)

Most participants said they would swipe left if they thought a profile was AI-generated. Some explained that they would do so because the profile seemed fake or like a catfish, while others felt it lacked authenticity and found it almost a little creepy. A couple of women, however, said they would swipe right out of curiosity if they suspected the profile was AI.

4.4 Interview - differences between AI and real profiles

4.4.1 Themes in real profiles

Themes in real profiles	**Quotes from participants**
Natural picture details	“Grainy picture”, “Different textures”
Normal features	“Hair on stomach”
Imperfect skin	“Not smooth skin”
Good looking	“No filter”, “Contrasts in the face”
Normal background	“No blurred background”

Table 4. Overview of the themes in real profiles

Real profiles were favored for their "Natural picture details" and "Normal features," with participants appreciating imperfections like "Imperfect skin" and contextual realism from "Good lighting" and "Normal backgrounds." These details collectively enhanced authenticity and relatability.

4.4.2 Themes in AI-generated profiles

Themes in AI-generated profiles	**Quotes from participants**
Soft picture details	“Edited”, “Mild edges”, “Soft colours”
Weird features	“Eyes look dead”, “Too many fingers”
Smooth skin	“Skin looks blended”, “Perfect skin”
Fake lightning	“Halo effect”, “Lack of shadow”
Fake eyes	“Crossed eyes”
Different persons	“Three different persons”

Table 5. Overview of the themes in AI-generated profiles

AI-generated profiles faced criticism for being overly polished, as seen in the themes "Soft picture details" and "Smooth skin." Inconsistencies, captured in "Weird features" and "Fake eyes," often gave away artificial origins. "Fake lighting" further detracted from realism, highlighting a lack of natural shadows or depth.

4.4.3 Cross comparison

The analysis showed that participants looked closely at both real and AI-generated profiles, but AI profiles were judged more critically. Real profiles were seen as real because of small details like textured skin, natural lighting, and the way they expressed emotions, which made them seem believable. But AI-generated profiles were often pointed out for being 'too perfect,' with smooth skin, unnatural lighting, and problems with how their eyes and hands looked. Participants also said that realistic backgrounds and consistent visual identities are important. They said that profiles where different people appear together are less authentic. Overall, imperfections and visual consistency are important factors that distinguish real profiles from AI-generated ones. This shows the areas where AI design must improve to create more convincing profiles.

5. Discussion

5.1 Method critique

The chosen method has some limitations that should be acknowledged.To start with, the participants were exclusively female engineering students from KTH, which does not provide a broad representation of women as a whole. Engineering students might have more experience with AI, making them more aware of AI-generated images. This could have led to higher accuracy in distinguishing between AI-generated dating profiles and real ones compared to if the study had included a broader group of women in their 20s and 30s from diverse backgrounds and professions.

Another limitation of our method was the process of creating the dating profiles and selecting suitable images. Since we had to use license-free pictures, the available options were limited. We also had specific requirements for the pictures, such as needing them to look authentic and ensuring the individuals appeared to be in their 20s. As a result, selecting images was challenging since they had to meet our requirements while also being license-free. These pictures didn’t always feel like the best fit for dating profiles. Many details that matter in dating profiles, such as personality, hobbies and lifestyle, were difficult to capture in the available license-free images. The selected pictures often looked too professional, which may have influenced our results. This could have been different if we had been able to use images of “everyday” people and create dating profiles based on those more natural-looking photos.

Another problem with the pictures was that the ones we found were almost always from the same photoshoot. Participants often pointed this out, saying it made the profiles feel staged and unrealistic. Real dating profiles usually have a mix of casual and candid photos, which helps make them feel authentic. This lack of diversity in photos likely influenced participants’ ability to identify both the real and the fake profiles. To address this, future studies could use a broader selection of images, including those sourced from volunteers or collaborations to access more natural looking photos. This would better represent real world dating profiles and reduce the bias introduced by overly polished or repetitive images.

Some of the AI-generated images also had clear flaws, like extra fingers, unnatural eyes, weird lighting, and inconsistent ethnic features in the same profile. Many participants commented on these things and pointed out that those factors were clear giveaways if a profile was AI-generated or not. One profile, in particular, caught the participants' attention: the person in the images seemed to change ethnicity between pictures (see Appendix 1, profile labeled "Aaron"). These flaws likely impacted the results, as they made the task of identifying AI-generated profiles less challenging than it would be with AI-pictures that had been better. To improve this, future studies could use higher-quality AI tools with better checks for errors to ensure that the pictures look more authentic. Evaluating the images with pilot groups before the main study could also help identify and address these flaws.

There were two other examples of ethnicity altering where all the pictures in the profile were affected (see Appendix 1 "Robin" and "Simon"). This is likely due to AI Hyperrealism (Miller et al., 2023), resulting in most of the people of colour in our set of real profiles appearing to be whitewashed after being entered into the AI model. If another AI model had been used, the generated images would likely have been different, as each model is trained on unique datasets with distinct biases.

5.2 Reflection on results

5.2.1 Errors

Across all errors, participants mistakenly identified AI-generated profiles as real in 17 cases, while real profiles were incorrectly labeled as AI in 8 cases. This indicates participants were significantly more likely to misidentify AI-generated profiles as real than to misclassify real profiles as AI. The former suggests that the AI was successful in deceiving some participants. Conversely, the misclassification of real profiles as AI may stem from the fact that the real images were sourced from external websites, which often feature "posed" or stylized elements that might resemble AI-generated attributes.

5.2.2 Difference between group A and B

Group A and Group B showed slightly different accuracy rates (85% compared to 90%), a difference that is most likely attributable to random variation. A closer look at the confusion matrices reveal that the primary discrepancy lies in identifying real profiles as real—Group A misclassified real images as fake more often than Group B (six instances versus two). For AI images misidentified as real, the difference was marginal—nine cases for Group A and eight for Group B. If this variation is related to the test conditions, it suggests that the real images shown to Group B may have exhibited characteristics more commonly associated with AI-generated content.

5.2.3 Accuracy and confidence levels

Participants demonstrated a high level of accuracy (87.5%) in distinguishing between AI-generated and real profiles. This contrasts with earlier studies, such as Nightingale and Farid (2022), which reported accuracy levels closer to random chance (48.2%). This higher accuracy may be attributed to the dating profile context. Participants might have evaluated the images more critically because of the dating context, where trustworthiness is a highly relevant factor. Dating profiles inherently require greater scrutiny, as participants assess not only appearance but also personality cues conveyed through images. Other possible explanations for the high accuracy include the AI tool used, the participant pool consisting mainly of engineering students and the deliberate selection of images. This raises questions about whether AI detection performance improves when participants are motivated by personal stakes, such as evaluating dating profiles.

Interestingly, while participants achieved high accuracy, their confidence levels were moderate with a mean of 3.3 out of 5, which is 66%. Comparing this to the accuracy of 87.5% there is a gap between self reported confidence and actual ability. This suggests a level of self-doubt or cautiousness, even when judgments were correct. Similar patterns of underconfidence in the ability to detect AI content were observed in Miller and Steward's (2023) findings, indicating that participants might have internalized concerns about AI's growing realism and potential to deceive.

Additionally, studies like Labajová (2023) showed participants generally trusted AI-generated content with a mean trust score of 3.6 on a 6-point scale. In contrast, our participants appeared more skeptical, as indicated by the frequency of negative themes such as 'Something is off.' This difference may stem from the more intimate and personal setting of dating profiles, which may prompt stricter evaluations compared to social media contexts.

5.3 Future work

Although none of the participants were aware that the dating profiles were AI-generated, their status as students at KTH Royal Institute of Technology likely makes them more familiar with AI compared to the average woman in their age group. Conducting this study again with a participant group that includes individuals beyond students would likely yield more generalizable results.

Furthermore, expanding the participant demographics to include a wider range of age groups could provide valuable insights, such as comparing how well individuals aged 60 and older recognize AI-generated profiles compared to those in their twenties. This could also be expanded to include men as well to allow for a second comparison.

Moving forward, there is a wide variety of generative AIs available for image creation. The AI used in this study produced different results with each prompt, meaning that other AIs would likely yield varying outcomes as well. Given the rapid advancements in AI-generated images, distinguishing between real and fake images is becoming increasingly challenging. As a result, if this study were repeated, the outcomes could be significantly different. It would then be valuable to explore the specific factors people rely on to identify fake images.

6. Conclusion

This study examined perceptions of trustworthiness in AI-generated versus real dating profile images. Participants demonstrated high accuracy (87.5%) in distinguishing AI-generated profiles but showed moderate confidence in their judgments. Key findings highlighted participants' sensitivity to subtle visual cues, such as smooth skin and inconsistent features, in AI-generated profiles, supporting the uncanny valley effect.

The results underscore the need for AI literacy to mitigate risks of deception and manipulation. Future research should include diverse demographics and platforms to better understand evolving AI technologies and their societal impact.

7. Project assessment

Our group has had excellent teamwork. The five of us have gotten along very well and everyone has had a say in all decisions that have been made. The workload has been distributed fairly between all members, with everyone contributing to all stages in the process; preparations, evaluation and the project report. Over the course of this project we have deepened our knowledge of thematic analysis and learned what a confusion matrix is and how to apply it. In hindsight, we should have chosen a different source for our images, as the ones we used appeared more like photoshoot pictures than typical images found in a dating profile, which likely impacted the results. On the other hand, we are proud as a group that we were able to conduct evaluations with as many as 20 participants.

Finally, a thank you to our supervisor Petra, who’s always been giving us helpful feedback and invaluable support throughout the process.

Reference List

Abramova, O., Baumann, A., Krasnova, H., & Buxmann, P. (2016, January). Gender differences in online dating: What do we know so far? A systematic literature review. In 2016 49th Hawaii International Conference on System Sciences (HICSS) (pp. 3858-3867). IEEE.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

Chan, L. S. (2017). Who uses dating apps? Exploring the relationships among trust, sensation-seeking, smartphone use, and the intent to use dating apps based on the integrative model. Computers in Human Behavior, 72, 246-258.

Getimg (2024), AI-powered image generation platform. Available at: https://getimg.ai

Huang, J., Gopalakrishnan, S., Mittal, T., Zuena, J., and Pytlarz, J., 'Analysis of Human Perception in Distinguishing Real and AI-Generated Faces: An Eye-Tracking Based Study', arXiv, 23 September 2024, arXiv:2409.15498v1 [cs.CV].

Labajová, L., Exploring the Perceptions, Credibility, and Trustworthiness of the Users Towards AI-Generated Content, Master's Thesis, Malmö University, 22 May 2023, supervised by E. Cory, examined by B. Romic.

Miller, E. J., Steward, B. A., Witkower, Z., Sutherland, C. A. M., Krumhuber, E. G., and Dawel, A., 'AI Hyperrealism: Why AI Faces Are Perceived as More Real Than Human Ones', Psychological Science, 34, 12 (2023), pp. 1390–1403, https://doi.org/10.1177/09567976231207095.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. IEEE Robotics & automation magazine, 19(2), 98-100.

Nightingale, S. J. and Farid, H., 'AI-Synthesized Faces Are Indistinguishable from Real Faces and More Trustworthy', Psychological and Cognitive Sciences, (2022), available at: https://doi.org/10.1073/pnas.2117609119

Pizer, S. M., & Marron, J. S. (2017). Object statistics on curved manifolds. In Statistical Shape and Deformation Analysis (pp. 137-164). Academic Press.

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., ... & Rombach, R. (2023). Sdxl: Improving latent diffusion models for high-resolution image synthesis. arXiv preprint arXiv:2307.01952.

RunDiffusion (2025), Juggernaut XL. Available at: https://rundiffusion.com/juggernaut-xl

Saranya, A. and Subhashini, R., 'A Systematic Review of Explainable Artificial Intelligence Models and Applications: Recent Developments and Future Trends', Decision Analytics Journal, 7 (2023), 100230.

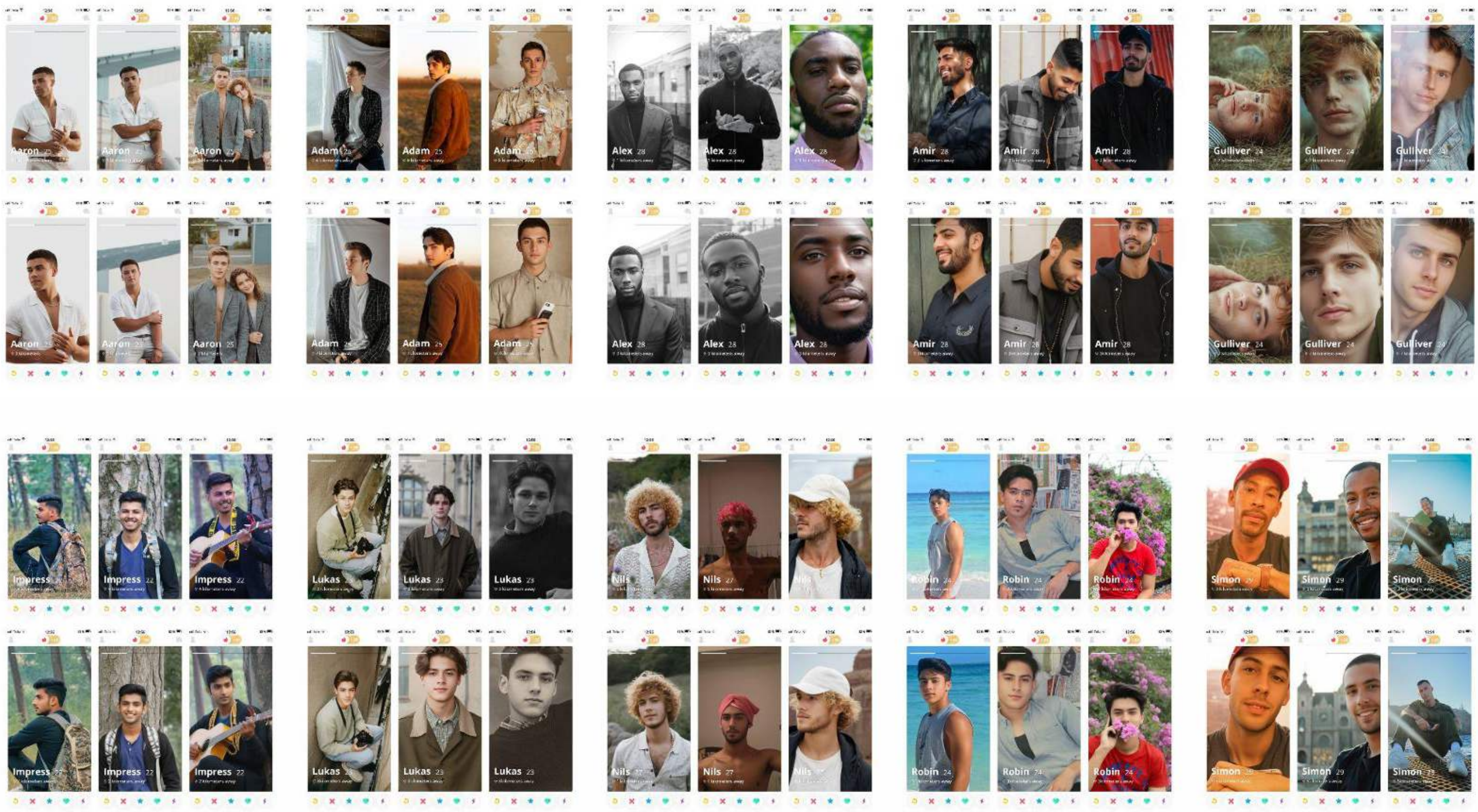
Tewari, J. and Bose, M., 'History of Artificial Intelligence', Indian Journal of Law & Legal Research, 5, 2 (2023), 1.

TinderKit (2024), Build fake Tinder profiles & prank your friends. Available at: https://tinderkit.com/

Unsplash, *Homepage*, Unsplash. Available at: https://unsplash.com/

White, D., Sutherland, C.A.M. & Burton, A.L. Choosing face: The curse of self in profile image selection. Cogn. Research 2, 23 (2017). https://doi.org/10.1186/s41235-017-0058-3

Appendix 1: All dating profiles and their corresponding AI-generated profiles (real profiles are on the first and third rows)



Appendix 2: Reference list for the images

Aaron: https://unsplash.com/@kenziekraft

Adam: https://unsplash.com/@alessiac_jpg

Alex: https://unsplash.com/@aliixar

Amir: https://unsplash.com/@awmirjim

Gulliver: https://unsplash.com/@kenziekraft

Impress: https://unsplash.com/@imimpress

Lukas: https://unsplash.com/@blakecheekk

Nils: https://unsplash.com/@jasrolyn

Robin: https://unsplash.com/@iamsupergilbert

Simon: https://unsplash.com/@instagramfotografin