Improving Human AI Image Detection: the Role of Emotional Content, Visual Attention to Detail Skills and Inductive Learning

Tess Walvius

s5490146

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

PSB3-BT_2526-1a-11

Supervisor: Dr. Ben Gützkow

Second evaluator: prof. dr. Ernestine Gordijn

In collaboration with: Lisa Bevers, Maaike de Kruijf, Megan Keane, Moritz Noß and Merwe Oosterhof.

January 26th, 2026

Declaration of AI use

1. No AI use

“No content generated by AI technologies has been presented as my own work.”

Abstract

Recent advances in text-to-image generation have raised concerns about the potential harmful effects of this new technology: scams, manipulation and misinformation. The present study aimed to investigate the factors that influence humans' ability to successfully differentiate between AI and real images. The effect of inductive learning on improving human AI (and real) image detection, as well as the effects of emotional content and visual attention to detail ability were studied using an online image detection task. Based on the literature it was expected that inductive learning, high visual attention to detail skills and fearful pictures (as opposed to happy and angry images) would have a positive effect on the detection of AI generated images. The data of $N = 270$ participants was analysed. The results showed training through inductive learning to be effective, especially when the image contained a happy expression. AI Images with fearful facial expressions were detected more accurately than those containing happy or angry expressions, only when individuals did not receive any training. Visual skills measured by the L-EFT did not show any effect on the detection rate of AI images, nor did it show an effect on the detection of real images. No interaction effects of training were found between visual attention to detail for training and emotional content. These findings suggest that inductive learning can be a promising approach for improving human AI detection and may help with minimizing the risks associated with text-to-image generation.

Keywords: Inductive learning, Human AI detection, Fear Bias

Improving Human AI Image Detection: the Role of Emotional Content, Visual Skills and Inductive Learning

Within only a few years, generative AI has changed how visual information is created, shared and perceived. During this development, generative AI has become more accessible for the general public. With, in the Netherlands alone, 23% of the population (12 years and older) and nearly half (48.7%) of young adults aged 18-25 having used AI in 2024 (Centraal Bureau voor de Statistiek, 2024). The use of AI is not only popular in the Netherlands however, text-to-image generation is being used by millions of people worldwide in an array of different fields, such as: marketing, art creation and education (Oppenlaender, 2024).

While the rapid rise of AI generated content, specifically text-to-image generation, has created opportunities for creativity, efficiency and innovation, it also raises concerns regarding authenticity, misinformation and the potential misuse of AI generated images (Cazzamatta & Sarısakaloğlu, 2025). With these concerns, along with AI generated images becoming more indistinguishable from real images (Yang et al., 2024) and the increase in the use of AI (Centraal Bureau voor de Statistiek, 2024), understanding how humans perceive and recognise AI generated images therefore becomes increasingly important. Research shows that humans struggle to accurately detect artificially generated images, often performing no better than chance, especially when the images contain human faces (Cooke et al., 2025; Greenspan & Bergold, 2025; Sapronov & Gorbunova, 2025; Yang et al., 2024). This inability could have serious implications, such as: the increase of realistic deepfakes (Becker & Laycock, 2023), financial fraud (Cooke et al., 2025), and the reinforcement of negative stereotypes (Chou et al., 2025; Gao et al., 2025).

While these risks alone form a potential threat, the use of emotion in these AI generated images could significantly add to the problem. In 2024, AI generated pictures depicting a frightened little girl in the aftermath of hurricane Helene (Colorado, US),

circulated the internet (Daniel, 2024), creating emotional responses amongst thousands of viewers, with some viewers even using this image to spread political misinformation (Jingnan, 2024). Previous research has shown that (negatively) emotionally charged images can increase people's willingness to donate money to a cause, especially when this image is shown on social media (Bak et al., 2024). However, when such images are artificially generated (i.e. image of a girl during hurricane Helele; Daniel, 2024), this increased willingness to donate could potentially be abused by scammers through evoking empathy to prompt donations to a nonexistent cause. In addition to that, the increase of fake media could lead to fewer donations and volunteering attempts to actual causes, by eliciting skepticism in individuals towards all images of disasters, including real ones (Daniel, 2024).

These potentially harmful use cases raise an important question: 'Is everyone as susceptible to getting tricked by these AI generated pictures? And which characteristics of AI generated images make them more easily detectable?' Could, for example, pictures depicting certain emotions get detected better than others? Real-world examples already demonstrate the potential impact of emotionally charged AI generated images (Daniel, 2024; Jingnan, 2024), highlighting the importance of further investigation. While some studies indicate that individual differences (e.g. age; Cooke et al., 2025) or image characteristics (i.e. emotional content; Park et al., 2024) can influence the detection rate of artificially generated pictures, the amount of quantitative research on this topic remains poor. Given the increase of realism in AI generated images and the potential negative consequences associated with this development, especially for emotional pictures, even small insights into how humans can better detect AI generated pictures can be of great importance.

The present study tries to address these concerns, by testing whether training through inductive learning can improve participants' detection rate of AI generated pictures as opposed to non-AI generated images. In addition to that, the relationship between emotional

facial expression of the AI generated image, individual differences in visual attention (to detail) in relation to the detection rate are also addressed.

Characteristics of AI Generated Images and the Role of Inductive Learning

With recent developments, artificially generated pictures seem more lifelike than ever before, especially for images containing human faces (Yang et al., 2024). While early AI generated images depicting humans often contained flaws, such as too many fingers, blurry skin texture or too many teeth (Chen et al., 2023), these inconsistencies have become less frequent and less detectable to the human eye. As a result, successfully detecting AI pictures now relies on more subtle cues that reveal the picture to be artificially generated, as opposed to depending on explicit rules for differentiation.

Inductive Learning

In order to train individuals to recognise these subtle differences between AI generated and real images, training should focus on seeing multiple examples and naturally forming generalisations about the different categories, without specific rule instruction. One promising approach based on these principles is inductive learning, in which individuals independently learn to recognise patterns across categories (Richter et al., 2022), a process potentially beneficial for successful AI image detection. Kornell and Bjork (2008) found that this learning effect was especially strong when examples of these categories were interleaved. This means that instead of seeing all the examples from one category followed by all the examples of another category, it is more effective to interleave examples from different categories in a shuffled sequence. The reason behind interleaving being effective can be explained by the discriminative-contrast hypothesis, which states that by interleaving categories, the contrasting differences between the different categories become more noticeable to an individual (Kang & Pashler, 2011; Birnbaum et al., 2012). By highlighting fine-grained differences between different categories, this approach could not only improve

AI detection accuracy, but also reduce the potential harmful risks of AI image generation (e.g., emotional manipulation; Daniel, 2024). Training individuals to recognise these patterns on their own through inductive learning could therefore minimize risks associated with AI text-to-image generation becoming more accessible and widespread.

In sum, it can be expected that training individuals to recognise AI generated images with inductive learning will have a positive effect on the detection rate of AI generated images, real pictures and the AI and real pictures combined (overall detection rate), by highlighting subtle differences between AI generated images and real images without external rule coaching, potentially minimizing the harmful risks associated with artificially generated images (Bak et al., 2024; Cazzamatta & Sarısakaloğlu, 2025; Daniel, 2024; Jingnan, 2024).

H1: Inductive learning training will improve the overall image detection rate of AI-generated and real images.

Human AI Detection

Beyond training, various factors could influence the ability for humans to successfully detect artificially generated pictures. For example, emotional images may be particularly likely to manipulate viewers (Bak et al., 2024), making AI generated images a great target for scams and manipulation. However, susceptibility to these images does not depend on content alone. Research suggests that individual differences of the viewer can also play a role in successful AI image detection (Cooke et al., 2025). Together these findings point to an important interaction between the content of an image and the characteristics of the person viewing it. Understanding how these factors can interact is therefore crucial for understanding successful human AI image detection.

A few studies have investigated which individual differences or image characteristics make AI pictures more easily identifiable. For example, Cooke et al. (2025) found that age influenced one's ability to detect AI-generated images, with older participants showing higher

error rates. Another study found that artificially generated images depicting positive emotions were more easily detectable than images depicting angry or neutral facial expressions (Park et al., 2024). However, it should be noted that this study was published in 2024, when AI systems were less developed, and had a small sample size ($N = 18$). Another important limitation of this study is that fear was not included as a negative emotional category, despite its central role in human emotion processing.

Fearful faces are known to capture attention rapidly (Stein et al., 2013), which could potentially positively influence AI detection. In addition, considering the potential harmful consequences of AI pictures depicting fearful faces (Bak et al., 2024c; Daniel, 2024; Jingnan, 2024), inclusion of fear as an emotional category is especially important. Overall, research on which factors influence the detection of AI generated images remains scarce, leaving unanswered questions about how specific emotional content, especially fear, and individual (perceptual) differences can influence AI image detection performance.

Emotion Recognition

Attentional Fear Bias

Humans detect fearful faces quicker than happy or neutral expressions, commonly referred to as an attentional fear bias (Stein et al., 2013). This bias appears to be automatic and may serve an adaptive function, contributing to socioemotional development (Eskola et al. 2023; Segal & Moulson, 2020; Stein et al., 2022). This view of the bias being an automatic mechanism, is also supported by neurological evidence showing that individuals with lesions in the primary visual cortex are still able to subconsciously process fearful faces when they are shown in their blind visual field (Bertini et al., 2017). This evidence suggests that fear recognition depends on highly fundamental processes, which may heighten sensitivity to subtle irregularities, as seen in AI images (Chen et al., 2023).

Researchers have speculated that this attentional fear bias could be related to the fearful faces being perceived as a threat (Stein et al., 2013). However, this bias in attention for fear is even prevalent when compared to angry faces, which can also be classified as threatening (Engen et al., 2015). This evidence suggests that fearful faces have a unique ability to capture human attention that goes beyond a general response to threatening stimuli. To gain a clearer understanding of the attentional fear bias and its potential role in AI image detection, the underlying visual mechanisms that contribute to it should be explored.

Emotion Recognition, Processing and AI Generated Images

Quick and accurate emotional recognition depends on the integration of individual senses and interaction of different brain regions (Choo et al., 2025). Emotion processing occurs through two different pathways: top-down processing and bottom-up processing, which both contribute to visual attention to detail skills. Recent evidence highlights the importance of top-down processing, showing that factors such as: context, emotional attention and earlier experiences can also influence emotion detection (Mohanty et al., 2025). Research from Lecker et al. (2025) further supports these findings with the addition that especially context seems to play a grand role in fear detection.

Bottom-up processing for emotion detection relies more on the physical aspects of a stimulus, like color or facial expressions (Mohanty et al., 2025). Facial perception through bottom-up processing uses two types of visual information: low-spatial frequencies (global visual features: e.g. the contours of the face; Menzel et al., 2018) and high-spatial frequencies (finer details: e.g. lashes, position of the brows) (Wylie et al., 2022).

Traditionally, it was believed that mostly low-spatial frequencies were associated with rapid fear detection by bypassing the visual cortex and operating mostly unconsciously (Stein et al., 2013). Yet, an increasing amount of studies have shed a light on the importance of high-spatial frequencies for this attentional fear bias (Stein et al., 2013). These high-spatial

details tend to be features in which generative AI models make small errors (Chen et al., 2023). High spatial frequency information could therefore play an important role in identifying AI generated pictures, by revealing small deformities that set apart AI images from real ones.

These fine grained details are not processed the same way by all individuals however. Individual differences in visual attention to detail ability can play an important role in perceiving high spatial frequencies (Nassar & Troiani, 2020). With attention to detail being defined as the ability to focus on fine-grained detail as opposed to the global picture. In conclusion, the value of high-spatial frequencies for AI image detection should be acknowledged and might even be more significant than the value of low-spatial frequencies. In addition to that, individuals with more pronounced visual attention to detail skills, are better at perceiving high-spatial frequencies. Therefore it can be hypothesized that individuals with more developed attention to detail skills are better at AI image detection than those with less developed visual attentional to detail skills.

H2: Individuals with better visual attention to detail ability will detect AI-generated images more accurately.

The tendency for fearful facial expressions to grasp humans' attention (attentional fear bias; Stein et al., 2013) could have an effect on an individual's ability to detect AI generated images depicting this facial expression. Since fearful faces automatically capture humans' attention (Stein et al., 2013), individuals might be more sensitive to inconsistencies in AI generated images that depict this emotion. In addition to that, because the detection of fearful facial expressions relies heavily on high-spatial frequency information and AI models sometimes struggle to generate these features perfectly (Chen et al., 2023), AI generated images depicting fearful expressions may be more easily recognisable than those showing positive or angry faces. Overall, these factors suggest that artificially generated images

depicting fearful faces may be detected more accurately than those showing happy or angry faces.

H3: AI generated faces displaying fear will be detected more accurately than those displaying positive emotions and angry emotions.

Individual's visual attention to detail ability (H2) and detection rate of fearful faces (H3) may also interact. If the detection of especially fearful stimuli depends on the perception of high frequency information (Stein et al., 2013), individuals that are detail oriented may be even better at the detection of fearful faces as opposed to happy or angry faces. Therefore it can be hypothesized that individuals with more pronounced attention to detail skills will detect fearful AI generated faces more accurately than participants with lower attention to detail abilities. Furthermore, receiving inductive learning training is expected to enhance participants' sensitivity to these high spatial frequencies for fearful expressions (Richter et al., 2022).

H4: Participants who receive inductive learning training will show a stronger relationship between visual attention to detail skills and detection accuracy for fearful AI faces as compared to individuals who do not receive this training.

Method

Participants

A total of $N = 291$ participants completed the study. All participants had to be at least 16 years of age to participate in this study. Some participants were recruited through the SONA platform belonging to the University of Groningen (n = 242), consisting mostly of first-year psychology students, who received course credits for their participation. Other participants were recruited through the personal networks of the researchers (n = 49); these individuals did not receive any compensation for participating. All participants received an informed consent form, which they read and signed prior study participation. This consent

form explained their rights as participants in a research study and how their data would be handled.

To ensure data quality, a mandatory completion time was used. The median was used to compute this completion time. For the control condition, the median completion time was 1003.5 seconds. Participants that completed the study within less than half of this duration or more than three times the median duration were excluded from analysis. Based on these criteria, one participant was excluded for taking too little time and four participants were excluded from analysis for taking too long.

In the training condition the median completion time was 1796 seconds. Applying the same exclusion criteria as before, sixteen participants from the training condition were excluded for taking too long to complete the study. No participants from the training condition were excluded for taking too little time.

After applying these criteria a total of $N = 270$ participants (203 female, 60 male, 5 non binary, 2 undisclosed) were left in the final sample. Of these participants, n = 148 were left in the control condition and n = 122 participants remained in the training condition. Most participants (n = 231) were around the age of 18-24. The remaining age categories were represented as follows: 25-34 (n = 14), 35-44 (n = 4), 45-54 (n = 11), 55-64 (n = 3) and 65+ (n = 7).

The study was deemed low risk by the guidelines set up by the Ethics Committee and therefore exempt from full ethical review. Data collection commenced on November 6th of 2025 and ended on November 16th of 2025 under code PSY-2526-S-0030.

Design

This study used a mixed experimental design and was conducted online using the Qualtrics platform. The design included one between-subjects factor (training vs. control) and three within-subjects factors (facial emotion, attractiveness and image type).

Participants were randomly assigned to either the experimental condition (i.e., training) or the control condition (no training). The independent variables included the training condition, facial emotion (happy; angry; fearful), level of attractiveness (high or low), and image type (AI-generated or real photo). The dependent variables measured were classification accuracy, confidence ratings, theory of mind and attention to detail.

Materials and Procedure

The experimental design was set up to examine whether inductive learning improves people's ability to distinguish AI-generated faces from real ones and how individual differences in cognitive and perceptual abilities affect their classification accuracy. All participants completed a test where they had to determine if a picture was AI generated or a real photograph. These images varied systematically in terms of emotional expressions (happy, angry and fearful) and attractiveness, and they were balanced for gender (women and men) and race (Black and White). The real pictures came from the Chicago Face Database (CFD) (S et al., 2015a; S et al., 2020). This database was chosen as all faces were labeled and categorized for the emotions, gender, race and attractiveness, the latter based on a ranking by a large sample (N = 100). All pictures in the database were unified, meaning the background was removed and all jewelry was removed. The AI images were generated through Midjourney version 7 (Midjourney, 2025) and Dall-e version 3 (OpenAI, 2025). These models were chosen because they were accessible to the research team and they are widely used (Bhattacharya, 2025), therefore creating a representative picture sample. Example prompts that were used can be found in appendix A, in addition to that, examples of AI images used in the study can be found in appendix B. The decision on which pictures to include was made on a joint agreement (n = 6) based on photorealism, expressed emotion and attractiveness. The pictures were matched to comparable images from the Chicago face database, to relatively match the gender, race and attractiveness. Both AI models on occasion added jewelry or a

coloured background. Using prompts to remove or alter the image did not result in a change in jewelry or a colored background. Therefore the AI images were edited to remove the background and remove jewelry to match the appearance of the real photographs.

The test contained 48 unlabeled pictures, 24 AI and 24 real images. These were further divided into 12 males and 12 females, each with the 3 types of emotions. For each emotion, happy, angry and fear, there were 4 pictures, of which half are considered attractive and the other half are considered unattractive. The labeling of the attractiveness was based on previous ratings for the real photographs (S et al., 2015; S et al., 2020) and based on prompt instructions for the artificially generated images. In the inductive learning condition, there were a total of 96 images shown with a description of whether this is AI or a real image.

Several measures were administered as part of the broader study. Measures relevant to the current research (visual attention to detail skills and demographic characteristics) are thoroughly reported below, whereas the remaining measures (theory of mind and confidence) are reported in short.

Visual Attention to Detail Skills. The participant's visual attention to detail ability was measured using the L-EFT (M = 8.008, SD = 1.806). The L-EFT measures the ability to disembed information from context (de-Wit et al., 2017). A higher score on this test indicates better visual-attention to detail skills in the disembedding of visual information. This test originally consists of 64 questions, however including the whole test would have caused concerns for the total length of the questionnaire. The L-EFT is a newer version of the original embedded figures test by Witkin (1971), which has been validated in short form with twelve and even six questions (Mumma, 1993). While this did not exist for the L-EFT, the data set including accuracy results was made publicly available by de-Wit et al. (2017). Based on this data of 255 participants, the ten most difficult items were selected, as Schlooz &

Hulstijn (2014) found that increased difficulty showed better increased sensitivity. The items had an accuracy range of 0.49-0.70, with an average of 0.59. While the validation of the short version of the L-EFT has yet to be determined, the L-EFT is validated to test visual attention to detail ability without major influences from broader cognitive abilities (Huygelier et al., 2018). The questionnaire first showed a figure at the top of the screen. The figure was presented together with three response options, among which only one contained the embedded figure. Participants were asked to choose the image in which they believed the figure appeared.

Demographic Characteristics. Participants were asked to indicate their gender, age group, and preferred language (Dutch/English). The corresponding question for gender was; ‘‘what is your gender?’’ with the response options *male, female, non-binary, other* or *prefer not to say.* Age was also asked and categorized into groups; *18-24, 25-34, 35-44*, and so on, up to *65+*. At last ‘‘what is your native language?’’ was asked. Participants could give multiple answers from the response options *English, Dutch, German, other* or *prefer not to say.* For each option, the opportunity to not disclose this type of information was included for ethical reasons.

Theory of Mind. This measure was not included in the analysis of the present study and is therefore only briefly described. The “Reading the Mind in the Eyes Test” (RMET; Baron-Cohen et al., 2001) was used to assess participants’ Theory of Mind (ToM; Carlson et al., 2013). Participants viewed eight images depicting eyes and were asked to select the emotion that best matched the expression from four options (M = 5.822, SD = 1.368).

Confidence. This measure was not included in the analysis of the present study and is therefore only briefly described. Participants’ confidence in their ability to classify images was assessed using an ad-hoc measure during multiple time points. Responses were given on a 6 point Likert-scale (1 = not confident at all, 6 = very confident) to avoid neutral responses

Image Classification. For the classification of faces generated by AI or real faces we asked with every shown image: “This image is:” with two response options. “AI-generated” or “A real photo” with an allowed response time of 15 seconds. In the inductive learning condition, participants completed a training session where they viewed a series of faces correctly labelled as AI-generated or real. These were shown in an alternating pattern to help the participants notice the visual distinctions between both categories. Participants in the control condition did not receive this training and instead went straight to the test.

After the survey participants had the opportunity to write any remark or feedback about the experiment if they wished to. Finally, they could see a message thanking their participation, which marked the end of the procedure.

Results

Preliminary Analysis

All data was analysed using JASP version 0.95.4 (JASP Team, 2025). Performance scores were calculated by creating an additional scoring variable for each test question and assigning either a 0 (incorrect) or 1 (correct) value based on the answer of the specific test question. These additional scoring variables were then added up in different categories to determine performance scores: happy AI score, angry AI score, fearful AI score, AI detection rate (total of AI pictures correctly identified), real pictures detection rate (total of real pictures correctly identified) and overall detection rate (total of all pictures correctly identified).

The effect of training on the detection rate (H1) was analyzed using an independent sample t test. To analyse the effect of the LEFT scores on the detection rate (H2), a linear regression model was used. Finally, to analyse the difference in detection rates for the different emotional categories (H3) and the relationship between training and the LEFT scores and AI depicted emotional category (H4), a RM (repeated measures) ANOVA with post hoc t tests were used.

All model assumptions were checked and corrected for if necessary. Any corrections done on the model/data are mentioned in the corresponding parts.

Main Analyses

Regarding Hypothesis 1, stating that inductive learning will improve the detection rate of all images (overall detection rate), one-sided t-tests were performed using condition (training vs. control) as the independent variable and the different picture detection rates (overall detection rate, AI detection rate and real pictures detection rate) as dependent variable. Training improved participants' overall detection rate ($t(247.7) = -9.014$, $p < .001$, $d = -1.077$, $M_{diff} = 6.273$), as well as their AI detection rate ($t(219.6) = -9.024$, $p < .001$, $d = -1.070$, $M_{diff} = 3.361$) and the detection rate of real images ($t(263) = 6.243$, $p < .001$, $d = -0.751$, $M_{diff} = 2.911$).[1] Although Cohen's d is negative, this reflects the coding in the analysis and does not indicate the direction of the effect; performance was higher in the training condition for all variables. These differences are also illustrated in Figure 1, which shows the overall detection rate increasing in the training condition. Therefore, the results support Hypothesis 1.

To analyse the second Hypothesis, which states that individuals with high visual attention to detail scores will show higher performance in AI picture detection, a linear regression model was conducted to predict AI detection rates from visual attention to detail (LEFT-scores), while controlling for condition (training vs. control). According to this analysis, there did not seem to be any relationship between the L-EFT score of participants and the AI detection rate when controlled for condition type (B = .151, SE = .107, $p = .158$). As illustrated in Figure 2, the LEFT scores also showed a ceiling effect with the mean score (

[1] There were indications of normality being violated for all three variables: overall detection rate (W=.93, p<.001), AI detection rate (W=.93, p<.001), and real pictures detection rate (W=.96, p<.001). However, because of the large sample size, violations of normality do not necessarily warrant any concern, as the central limit theory ensures that the test statistics remain robust (Field, 2017). A Mann-Withney test did not yield different results (all p's < .001)

$\underline{x}$=8.008, sd=1.806) being only two points away from the maximum score of ten, with the most common score being the maximum score (mode = 10). In conclusion, the results do not support Hypothesis 2.

Hypothesis 3, stating that fearful AI generated pictures will be detected more accurately than angry or happy AI generated pictures, was examined using an RM ANOVA and post hoc t tests with the three different AI emotional categories (happy, fearful and angry) as the repeated measures cells and condition (training vs. control) as a between subjects factor. Since Mauchly's test indicated that the assumption of sphericity was violated (W = .963, X^2 = 10.12, p = .006), the degrees of freedom were corrected using the Greenhouse-Geisser correction (ε = .964). There seemed to be an effect of the type of emotional AI content on the detection rate of these images, F(1.924; 515.680) = 19.798, p = < .001, η^2_p = .069, as well as an interaction effect between emotional content and condition type (training vs. control), F(1.924; 515.680) = 4.769, p = .01, η^2_p = .017. In the control condition, AI faces displaying fear were more easily detectable than AI faces displaying happiness (t(267) = 6.912, p_{bonf} < .001) or anger (t(267) = 3.653, p_{bonf} = .005). However, in the training condition this difference disappeared (fear v.s. happiness: t(267) = 2.173, p_{bonf}=.460, fear v.s. anger: t(267) = 1.525, p_{bonf} = 1). Notably, training led to the largest improvement in the detection rate of happy AI generated faces (M_{exp}-$M_{contro\,l}$ = 1.396, t(249.7) = 7.832, p_{bonf} < .001, d = 0.937). This interaction effect between training and emotion shown in the AI pictures is visible in Figure 3, in which the increase in the detection rate from control to the training condition varies mostly for the AI pictures displaying happiness. Therefore the results partially support Hypothesis 3, as the effect of fearful AI pictures being easier to detect, was only visible in the control condition.

The final hypothesis (H4) expected a stronger interaction effect between the detection rate of fearful AI pictures and experimental condition for individuals with high visual

attention to detail. This was analysed using an RM ANOVA, using the three different AI emotional categories (happy, fearful and angry) as the repeated measures cells, condition (training vs. control) as a between subjects factor and the LEFT score as a covariate. According to Mauchly's test, the assumption of sphericity was violated ($W = .963$, $X^2 = 9.782$, $p = .008$), which is why the Greenhouse-Geisser correction was used ($\varepsilon = .964$). Moreover, there was also no relationship between the LEFT score and the emotional content of the pictures with relation to condition ($F(1.929; 503{,}412) = .533$, $p = .580$). In sum, the results do not support Hypothesis 4.

In conclusion, some hypotheses were supported. Training through inductive learning had a positive effect on the overall detection rate, AI detection rate and real pictures detection rate (H1). AI generated pictures depicting fearful faces were more easily detectable than AI generated images of happy and angry faces (H3), only in the control condition. And lastly, there was no relationship between the LEFT scores and the detection rates (H2) and this relationship did not differ over condition types and visual attention to detail scores (H4).

Explorative Analysis

An additional exploratory analysis focussed on the effects of age on the overall detection rate with regards to the conditions. As previous research identified, older participants usually had more difficulties detecting AI generated images (Cooke et al., 2025) and the effect of training on this detection rate has not yet been explored. This explorative research aims to examine potential areas for further research without making any inferential claims.

To examine the effect of age in relation to condition type (control vs. training), an ANOVA was used with the overall detection rate as the dependent variable and the condition type (training vs. control) and their age group (older vs. younger adults) as independent variables. After this analysis, post hoc independent t tests were also used to examine the

results more closely. Because of the limited number of individuals in the older age groups, a division was made between younger adults (18-24, 25-34, 35-44) (n = 249) and older adults (45-54, 54-64, 65+) (n = 21).

An individual's age did seem to have an effect on the overall detection rate ($F(1; 266) = 50.25$, $p < .001$). Interestingly, the analysis also suggested an interaction effect between one's age and the effect of training through inductive learning, $F(1; 266) = 18.67$, $p < .001$, $\eta^2_p = .185$, with older adults showing greater improvements in their overall detection rate after training (M_{exp}-$M_{control}$ = 15.806, $t(266) = 6.834$, $p_{bonf} < .001$, d = 3.013). For the control condition, younger adults performed better than older adults ($\underline{x}_1$-$\underline{x}_2$ = 13.733, $t(266) = 8.694$, $p_{bonf} < .001$, d = 2.618). However, in the training condition this difference is no longer significant ($\underline{x}_1$-$\underline{x}_2$ = 3.332, $t(266) = 1.834$, p_{bonf} = .406, d = 0.635). This effect is also visible in Figure 4. This result suggests that different age groups might respond differently to inductive learning training.

Discussion

The main aim of this study was to investigate the effects of inductive learning training, the type of emotion depicted in AI images and visual attention to detail skills on successful AI image detection. By identifying the factors that influence the detection of artificially generated pictures, this study aimed to contribute to minimizing the risks associated with the widespread distribution of AI pictures (Bak et al., 2024; Cazzamatta & Sarısakaloğlu, 2025; Daniel, 2024; Jingnan, 2024). At first, training through inductive learning was associated with higher detection rates of both AI-generated and real pictures, only in the control condition (H1). In contrast, better visual attention to detail skills measured by the L-EFT were not associated with higher detection rates for either image type (H2). Notably, AI pictures demonstrating a fearful expression were detected more accurately than those displaying a happy or an angry expression in the control condition (H3). Finally, no interaction effect was

found between visual-attention to detail skills and the type of emotion depicted in the AI images, nor was this relationship moderated by training (H4).

Interpretation of Results

Hypothesis 1. The finding that inductive learning improved the overall detection rates suggest that the participants were able to learn to find small differences between the real images and those generated by artificial intelligence. The overall training effect was also large (d = -1.077), hinting that training had a substantial impact on detection ability. Through repeated exposure and continuous comparison of varied image examples, as opposed to explicit instruction, individuals were able to learn to make their own generalisations for successful differentiation between AI and real pictures. As a result, inductive learning as a process may be particularly effective in distinguishing AI images as it supports perceptual sensitivity to small differences that distinguish these closely resembling categories.

Hypothesis 2. No support could be found for the expected relationship between visual attention to detail skills and the detection rates. One possible explanation for this null-effect could be that the finding of training through inductive learning suggests successful AI image detection may rely more strongly on perceptual learning than on individual differences, such as visual-attention to detail skills. Another possible reason could be rooted in the measure used to determine visual attention ability (L-EFT; De-Wit et al., 2017), however this point is discussed further in the limitations section.

Hypothesis 3. In control condition, detection accuracy was lower for happy and angry AI-faces, as compared to fearful AI-faces. This emotional advantage of fear disappears however, when inductive learning training is applied, implying an interaction effect between emotion displayed in the AI picture and training. One possible explanation for this interaction could be that baseline detection rates for fearful AI faces were already elevated, limiting the improvement from training compared to other emotional categories. A potential reason for

this heightened baseline (only visible in control condition) detection ability for fearful AI faces might come from an inherent fear bias (Stein et al., 2013), which makes humans more perceptive to fearful faces, in turn recognising artificially generated fearful faces more accurately. The benefit of this main effect then appears to be reduced when inductive learning is applied. On the other hand, since AI generated images are based on already existing images and the demand of the individuals generating these images (Yadav et al., 2024), the ability for humans to detect fearful AI faces more accurately compared to happy and angry artificially generated expressions in the control condition, could also be due to the AI systems being less able to generate this type of emotion accurately. Text-to-image generation is widely used in the field of marketing (Institute for Human-Centered Artificial Intelligence, 2025), which usually relies on happy images to sell products (Sengupta et al., 2025). Therefore, it can be argued that text-to-image generators are just not trained well enough to produce accurate fearful faces. Nevertheless, the increase of the AI detection rate for fearful faces after training through inductive learning suggests that this effect can not fully be attributed to general limitations of AI systems alone. Previous research found happy AI faces to be easier to detect than angry AI faces (Park et al., 2024), stating that positive emotions were therefore easier to detect than negative emotions. However, the current study results suggest that emotion-related AI detection advantages may be more specific than previously assumed by Park et al. (2024), with fearful faces (negative emotion) being the easiest to detect compared to both happy and angry faces. This finding highlights the importance of distinguishing different emotion types rather than treating emotion as a binary category.

Hypothesis 4. There was no interaction found between visual attention to detail skills and detection of fearful AI generated faces. This relationship also did not differ when participants had received inductive learning training. These findings suggest that visual attention to detail ability, as measured by the L-EFT, did not successfully moderate the

detection of fearful AI faces. However, the interpretation of this null-effect should be taken with caution as the results could be due to limitations regarding the L-EFT measure, which are discussed further in the limitations section and were also addressed in relation to Hypothesis 2.

Explorative Analysis. Additional exploratory analysis found a main effect of age on the overall detection rate, with older adults (aged ≥45yo) scoring generally lower than their younger counterparts (aged ≤44yo). However, in the training condition, this effect disappeared, suggesting an interaction effect between age and inductive learning training. This means that older adults improved more after training through inductive learning than younger adults. One possible explanation for this effect could be that older adults have less prior experience with identifying text-to-image generated pictures, because they use it less (Centraal Bureau voor de Statistiek, 2024), making their baseline detection rates lower. Training then has an even greater effect, by allowing them to get experience with the different categories. In contrast, younger adults could already possess more familiarity with AI generated images, limiting their learning ability. It is important to note that this interpretation remains speculative, since experience with AI images was not measured, the sample size of the older adults was small. In addition, these findings should be taken with caution because they were exploratory, the study was not made to study age-related effects on detection rates and there were no hypotheses formulated to test this beforehand.

Implications

Theoretical Implications. The support for Hypothesis 1, stating that inductive learning will improve participant's ability to successfully distinguish between AI generated and real images, provides theoretical support for the inductive learning theory (Kornell & Bjork, 2008) as well as the discriminant-contrast hypothesis (Kang & Pashler, 2011; Birnbaum et al., 2012). Due to the fast development of text-to-image generation, successful

detection of artificially generated pictures depends on noticing fine details rather than obvious distortions (Chen et al., 2023). Interleaving of the examples may have facilitated this noticing process by making the contrast between the pictures more apparent (Kornell & Bjork, 2008). This contrast in turn enhances participant's ability to differentiate between the different categories, as supported by the discriminative-contrast hypothesis (Kang & Pashler, 2011; Birnbaum et al., 2012). This finding extends the inductive learning theory with new domains in which it has proven to be effective. The finding that AI images containing fearful faces get detected better than happy or angry faces, indirectly supports the theory of an attentional fear bias capturing human's attention (Stein et al., 2013). The higher detection accuracy of these AI images with faces depicting fear indicate that prioritized emotional stimuli processing can enhance sensitivity to artificially generated pictures. Even though the study design does not allow for any direct conclusions about attentional mechanisms, the results do indirectly support the role of fear-related attentional processing in detecting AI pictures.

Practical Implications. The large positive effect of inductive learning training on distinguishing artificially generated images from real ones, gives hope towards human's ability to identify AI pictures. This can have implications as an intervention to protect individuals from potential harmful consequences of AI image generation, such as scams (Bak et al., 2024; Daniel, 2024) and misinformation (Cazzamatta & Sarısakaloğlu, 2025; Jingnan, 2024). As known from previous research, the elderly are frequently scammed on the internet (Burnes et al., 2017). In combination with the results of the explorative analysis suggesting that older individuals benefit most from training, interventions based on inductive learning in this particular group could be especially effective. However, as previously mentioned, these results need to be taken with caution and further research is necessary to make any inferential claims.

Strengths, Limitations and Future Directions

Strengths. The study had a number of strengths. To start off, the study had a large sample size of $N = 270$ participants, which increases statistical power and reliable parameter estimates. Secondly, because of the use of standardized images, the internal validity of the study is high, limiting the influence of outside factors on the results. The inclusion of multiple emotions was also a strong point of the present study, as previous research had not yet examined the effect of fearful emotions on AI image detection and only focussed on happy vs. angry emotions (Park et al., 2024). The use of inductive learning as a training method can also be considered a strength in this study, as the approach has been well supported by previous research on learning patterns and categorization (Richter et al., 2022), strengthening the interpretation of the training effects for AI image detection.

Limitations. As with any experimental study, there were also limitations that need to be taken into account when interpreting the study results. The primary limiting factor for this study is the fast development of AI systems. With AI models evolving rapidly (Cooke et al., 2025), the findings of the present study reflect a snapshot in time and may not be generalizable to newer text-to-image models. In addition to the previous point, it is unclear how long the inductive learning training effect actually stays, especially when taking the rapid development of text-to-image models into account. In the present study, participants were tested on their detection skills immediately after the training. Therefore, no claims can be made about the long term effects of inductive learning training on AI image detection. Furthermore, the use of only two text-to-image models to generate the AI pictures, is also a limitation. These models were chosen because they are one of the most widely used (Bhattacharya, 2025) and were accessible to the research team, however the findings of this study are therefore only generalizable to images created by these exact text-to-image models: Midjourney (Midjourney, 2025) and Dall-e (OpenAI, 2025). The sample that was used in the study was a limiting factor for the generalizability of the results. The sample consisted of first

year psychology students from the University of Groningen and participants recruited from the personal networks of the research team, making this sample a convenience and WEIRD (Western, Educated, Industrialized, Rich, Democratic; Kanazawa, 2020) sample. Consequently, the study sample is not representative of all individuals. The final limitation addresses the null-effect of the visual attention to detail scores. The L-EFT might not have been the most sensitive measure to measure attention to detail in this study. The scores on the L-EFT showed a strong ceiling effect ($\underline{x}$=8.008, sd=1.806, mode = 10), meaning a large number of participants achieved high scores on the test, therefore the variability on these scores was limited. This could have had an impact on the ability to detect a relationship between visual attention to detail skills and detection rates. In addition to this, the short version of the L-EFT has not yet been validated for testing visual attention ability (De-Wit et al., 2017). Therefore the absence of a main effect of visual attention to detail skills on AI image detection could be due to study limitations rather than the absence of a correlation.

Future Directions. In retrospect, the L-EFT might not have been the best measure for differentiating participant's visual attention to detail skills in this study. Therefore, future research could focus on using other measures with higher sensitivity and differentiability to conceptualize visual attention to detail skills in relation to the detection of AI images. More sensitive measures may reduce ceiling effects and show more variability between individuals, thereby providing a more accurate assessment of how visual attention to detail ability relates to AI image detection. For example, a visual search task with increasing difficulty could be used for measuring attention to detail skills (Palmer & Davis, 2004), with a specific focus on recognising subtle differences. Taken together, future research is necessary to investigate if visual attention to detail skills contribute to successful AI image detection. Beyond these methodological points, future research could also explore the inductive learning effects across age groups, AI models and focus on the long term effects of training.

Conclusion

To conclude, the main takeaway from this experiment is that training through inductive learning appears to be an effective method for improving individuals' ability to successfully distinguish between artificially generated and real images. In addition, at baseline, individuals seem to be better at detecting fearful AI faces as opposed to happy or angry AI faces. This emotional advantage diminishes when inductive learning is applied, implying that humans can be trained to detect happy and angry AI generated faces as accurately as AI images containing fearful expressions. In contrast, no evidence was found for a relationship between visual-attentional to detail skills and detection ability. These findings may have implications for minimizing the potential harmful risk associated with text-to-image generation, such as scams (Bak et al., 2024; Daniel, 2024) and misinformation (Cazzamatta & Sarısakaloğlu, 2025; Jingnan, 2024). Future research could focus on the long term effects of inductive learning training, the training effect in older adults, the detection rate of artificially generated videos and/or using other AI models.

References

Aziz, M., Rehman, U., Danish, M. U., & Grolinger, K. (2025). Global-Local Image Perceptual Score (GLIPS): Evaluating photorealistic quality of AI-Generated images. *IEEE Transactions on Human-Machine Systems*, 1–11. https://doi.org/10.1109/thms.2025.3527397

Bak, S., Yeu, M., Min, D., Lee, J., & Jeong, J. (2024). Charitable crowdfunding donation-intention estimation depending on emotional project images using fNIRS-based functional connectivity. *PLoS ONE*, *19*(5), e0303144. https://doi.org/10.1371/journal.pone.0303144

Becker, C., & Laycock, R. (2023). Embracing deepfakes and AI‐generated images in neuroscience research. *European Journal of Neuroscience*, *58*(3), 2657–2661. https://doi.org/10.1111/ejn.16052

Bertini, C., Cecere, R., & Làdavas, E. (2017). Unseen fearful faces facilitate visual discrimination in the intact field. *Neuropsychologia*, *128*, 58–64. https://doi.org/10.1016/j.neuropsychologia.2017.07.029

Bhattacharya, J. (2025, October 27). *AI Image Generator market Statistics*. SEO Sandwitch. https://seosandwitch.com/ai-image-generator-market-statistics/

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2012). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, *41*(3), 392–402. https://doi.org/10.3758/s13421-012-0272-7

Burnes, D., Henderson, C. R., Sheppard, C., Zhao, R., Pillemer, K., & Lachs, M. S. (2017). Prevalence of Financial fraud and Scams among Older adults in the United States: A Systematic Review and Meta-Analysis. *American Journal of Public Health*, *107*(8), e13–e21. https://doi.org/10.2105/ajph.2017.303821

Carlson, S. M., Koenig, M. A., & Harms, M. B. (2013). Theory of mind. *Wiley Interdisciplinary Reviews Cognitive Science*, *4*(4), 391–402. https://doi.org/10.1002/wcs.1232

Cazzamatta, R., & Sarısakaloğlu, A. (2025). AI-Generated Misinformation: A case study on emerging trends in Fact-Checking practices across Brazil, Germany, and the United Kingdom. *Emerging Media*, *3*(2), 214–251. https://doi.org/10.1177/27523543251344971

Centraal Bureau voor de Statistiek. (2024, September 3). Bijna kwart Nederlanders gebruikt kunstmatige intelligentie zoals ChatGPT. *Centraal Bureau Voor De Statistiek*. https://www.cbs.nl/nl-nl/nieuws/2024/36/bijna-kwart-nederlanders-gebruikt-kunstmatige-intelligentie-zoals-chatgpt

Chen, Z., Sun, W., Wu, H., Zhang, Z., Jia, J., Min, X., Zhai, G., & Zhang, W. (2023). Exploring the naturalness of AI-Generated images. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2312.05476

Choo, C. M., Bai, S., Privitera, A. J., & Chen, S. A. (2025). Brain Imaging Studies of Multisensory Integration in Emotion Perception: A scoping review. *Neuroscience & Biobehavioral Reviews*, 106118. https://doi.org/10.1016/j.neubiorev.2025.106118

Chou, W. S., Gaysynsky, A., Everson, N. S., Muro, A., Schrader, K., & Iles, I. (2025). Portrayal of cancer patients in the era of AI: a content analysis of images produced by generative AI tools. *Health Communication*, 1–11. https://doi.org/10.1080/10410236.2025.2537807

Cooke, D., Edwards, A., Barkoff, S., & Kelly, K. (2025). As good as a coin toss: Human Detection of AI-Generated Content. *Communications of the ACM*. https://doi.org/10.1145/3729417

Daniel, L. (2024, October 5). How Hurricane Helene Deepfakes Flooding social media hurt real people. *Forbes*. https://www.forbes.com/sites/larsdaniel/2024/10/04/hurricane-helena-deepfakes-flooding-social-media-hurt-real-people/

De-Wit, L., Huygelier, H., Van Der Hallen, R., Chamberlain, R., & Wagemans, J. (2017). Developing the Leuven Embedded Figures Test (L-EFT): testing the stimulus features that influence embedding. *PeerJ*, *5*, e2862. https://doi.org/10.7717/peerj.2862

Dueck, K. G. (1976). Mathemagenic mechanisms in inductive learning. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, *8*(1), 78–87. https://doi.org/10.1037/h0081936

Engen, H. G., Smallwood, J., & Singer, T. (2015). Differential impact of emotional task relevance on three indices of prioritised processing for fearful and angry facial expressions. *Cognition & Emotion*, *31*(1), 175–184. https://doi.org/10.1080/02699931.2015.1081873

Eskola, E., Kataja, E., Hyönä, J., Nolvi, S., Häikiö, T., Carter, A. S., Karlsson, H., Karlsson, L., & Korja, R. (2023). Higher attention bias for fear at 8 months of age is associated with better socioemotional competencies during toddlerhood. *Infant Behavior and Development*, *71*, 101838. https://doi.org/10.1016/j.infbeh.2023.101838

Field, A. (2017). *Discovering statistics using IBM SPSS statistics*. https://bvbr.bib-bvb.de:443/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=029907115&sequence=000001&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA

Gao, F., Xia, L., & Zhong, W. (2025). Stereotypes in artificial intelligence-generated content: Impact on content choice. *Journal of Experimental Psychology Applied*. https://doi.org/10.1037/xap0000548

Greenspan, R. L., & Bergold, A. N. (2025). Can AI-generated faces serve as fillers in eyewitness lineups? *Memory*, 1–14. https://doi.org/10.1080/09658211.2025.2467134

Huygelier, H., Van Der Hallen, R., Wagemans, J., De-Wit, L., & Chamberlain, R. (2018). The Leuven Embedded Figures Test (L-EFT): measuring perception, intelligence or executive function? *PeerJ*, *6*, e4524. https://doi.org/10.7717/peerj.4524

Institute for Human-Centered Artificial Intelligence. (2025). The 2025 AI Index Report. In *Artificial Intelligence Index Report 2025*. Stanford University. https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf

JASP - a fresh way to do statistics. (2025, October 15). JASP - Free and User-Friendly Statistical Software. https://jasp-stats.org/

Ji, S. (2025). #MeToo in an AI-generated deepfake sexual violence era in South Korea. *Women S Studies International Forum*, *112*, 103146. https://doi.org/10.1016/j.wsif.2025.103146

Kanazawa, S. (2020). What do we do with the WEIRD problem? *Evolutionary Behavioral Sciences*, *14*(4), 342–346. https://doi.org/10.1037/ebs0000222

Kang, S. H. K., & Pashler, H. (2011). Learning Painting Styles: Spacing is Advantageous when it Promotes Discriminative Contrast. *Applied Cognitive Psychology*, *26*(1), 97–103. https://doi.org/10.1002/acp.1801

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories. *Psychological Science*, *19*(6), 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

Kramer, R. S. S., Jones, A. L., Fitousi, D., & Tree, J. J. (2025). AI-generated images of familiar faces are indistinguishable from real photographs. *Cognitive Research Principles and Implications*, *10*(1). https://doi.org/10.1186/s41235-025-00683-w

Lecker, M., Hallock, S., Danielson, A., Van Aertrickc, M., Kindt, M., & Aviezer, H. (2025). Real-life intense fear is communicated through context, not facial expressions.

Proceedings of the National Academy of Sciences, *122*(11). https://doi.org/10.1073/pnas.2414677122

Lu, Z., Huang, D., Bai, L., Liu, X., Qu, J., & Ouyang, W. (2023). Seeing is not always believing: Benchmarking Human and Model Perception of AI-Generated Images. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2304.13023

Menzel, C., Redies, C., & Hayn-Leichsenring, G. U. (2018). Low-level image properties in facial expressions. *Acta Psychologica*, *188*, 74–83. https://doi.org/10.1016/j.actpsy.2018.05.012

Midjourney. (2025). Midjourney. https://www.midjourney.com/home

Mohanty, A., Freeman, J., & Jin, J. (2025). Top-down influences on the perception of emotional stimuli. *Nature Reviews Psychology*. https://doi.org/10.1038/s44159-025-00446-w

Mumma, G. H. (1993). The embedded figures test: internal structure and development of a short form. *Personality and Individual Differences*, *15*(2), 221–224. https://doi.org/10.1016/0191-8869(93)90029-3

Nassar, M. R., & Troiani, V. (2020). The stability flexibility tradeoff and the dark side of detail. *Cognitive Affective & Behavioral Neuroscience*, *21*(3), 607–623. https://doi.org/10.3758/s13415-020-00848-8

Negreiro, M. & European Parliamentary Research Service. (2025). Children and deepfakes. In *EPRS | European Parliamentary Research Service* (Report PE 775.855). https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775855/EPRS_BRI%282025%29775855_EN.pdf

OpenAI. (2025). OpenAI. https://openai.com/nl-NL/index/dall-e-3/

Oppenlaender, J. (2024). The cultivated practices of Text-to-Image generation. In *Springer eBooks* (pp. 325–349). https://doi.org/10.1007/978-3-031-66528-8_14

Palmer, J., & Davis, E. (2004). Visual search and attention: An overview. *Spatial Vision*, *17*(4), 249–255. https://doi.org/10.1163/1568568041920168

Park, H., Kim, G., Lee, D., & Kim, H. K. (2024). Can you spot the AI-Generated Images? Distinguishing fake images using signal detection theory. In *Lecture notes in computer science* (pp. 299–313). https://doi.org/10.1007/978-3-031-60913-8_21

Richter, T., Nemeth, L., Berger, R., Ferri, R. B., Hänze, M., & Lipowsky, F. (2022). Using interleaving to promote inductive learning in educational contexts. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *54*(4), 164–175. https://doi.org/10.1026/0049-8637/a000260

S, D., MA, Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

S, D., MA, Kantner, J., & Wittenbrink, B. (2020). Chicago Face Database: Multiracial expansion. *Behavior Research Methods*, *53*(3), 1289–1300. https://doi.org/10.3758/s13428-020-01482-5

Sapronov, F. A., & Gorbunova, E. S. (2025). Comparing AI-Generated Stimuli and Photos: Visual Search Study. *Moscow University Psychology Bulletin*, *48*(2), 109–131. https://doi.org/10.11621/lpj-25-14

Schlooz, W. A., & Hulstijn, W. (2014). Boys with autism spectrum disorders show superior performance on the adult Embedded Figures Test. *Research in Autism Spectrum Disorders*, *8*(1), 1–7. https://doi.org/10.1016/j.rasd.2013.10.004

Segal, S. C., & Moulson, M. C. (2020). What drives the attentional bias for fearful faces? An eye‐tracking investigation of 7‐month‐old infants' visual scanning patterns. *Infancy*, *25*(5), 658–676. https://doi.org/10.1111/infa.12351

Sengupta, S., Nagral, G., Sawantdesai, M., Yadav, K., & Mulwani, Y. (2025). Leveraging Text-to-Image generation models for automated creative processes in corporate marketing. In *Lecture notes in networks and systems* (pp. 563–576). https://doi.org/10.1007/978-981-97-8602-2_50

Stein, T., Jusyte, A., Gehrer, N. A., Scheeff, J., & Schönenberg, M. (2022). Intact prioritization of fearful faces during continuous flash suppression in psychopathy. *Journal of Psychopathology and Clinical Science*, *131*(5), 517–523. https://doi.org/10.1037/abn0000753

Stein, T., Seymour, K., Hebart, M. N., & Sterzer, P. (2013). Rapid fear detection relies on high spatial frequencies. *Psychological Science*, *25*(2), 566–574. https://doi.org/10.1177/0956797613512509

Witkin, H. A. (1971). Group Embedded Figures test [Dataset]. In *PsycTESTS Dataset*. https://doi.org/10.1037/t06471-000

Wylie, J., Tracy, R. E., & Young, S. G. (2022). The effects of spatial frequency on the decoding of emotional facial expressions. *Emotion*, *23*(5), 1423–1439. https://doi.org/10.1037/emo0001115

Yadav, N., Sinha, A., Jain, M., Agrawal, A., & Francis, S. (2024). Generation of Images from Text Using AI. *International Journal of Engineering and Manufacturing*, *14*(1), 24–37. https://doi.org/10.5815/ijem.2024.01.03

Yang, K., Singh, D., & Menczer, F. (2024). Characteristics and prevalence of fake social media profiles with AI-generated faces. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2401.02627

Tables and Figures

Tables

Table 1

Descriptives of the Score Variables of the Test

	Total_Scared_AI	Total_Angry_AI	Total_Happy_AI_	Total_AI_Correct	Total_Real_Correct	Total_Pictures_Correct
Valid	270	270	270	270	270	270
Missing	0	0	0	0	0	0
Mean	6.926	6.644	6.374	19.94	18.59	38.54
Std. Deviation	1.294	1.360	1.671	3.639	4.188	6.713
Minimum	0.000	2.000	0.000	4.000	6.000	16.00
Maximum	8.000	8.000	8.000	24.00	24.00	48.00

Table 2

Correlations Between the Major Variables

Variable		LEFT score total	Total_AI_Correct	Total_Real_Correct	Total_Pictures_Correct	Age
1. LEFT score total	n	—				
	Pearson's r	—				
	p-value	—				
2. Total_AI_Correct	n	264	—			
	Pearson's r	0.070	—			
	p-value	.258	—			
3. Total_Real_Correct	n	264	270	—		
	Pearson's r	0.004	0.469***	—		
	p-value	.943	< .001	—		
4. Total_Pictures_Correct	n	264	270	270	—	
	Pearson's r	0.040	0.834***	0.878***	—	
	p-value	.517	< .001	< .001	—	
5. Age	n	264	270	270	270	—
	Pearson's r	−0.013	−0.423***	−0.270***	−0.398***	—
	p-value	.831	< .001	< .001	< .001	—

* p < .05, ** p < .01, *** p < .001

Table 3

RM ANOVA of the Interaction Effect Between AI Emotion Type (Happy, Fearful and Angry) and Condition (Training vs. Control)

Within Subjects Effects

Cases	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2_p
Emotion shown AI	Greenhouse-Geisser	11.825	1.928	6.133	6.337	.002	0.023
Emotion shown AI ✻ Condition	Greenhouse-Geisser	8.900	1.928	4.616	4.770	.010	0.018
Emotion shown AI ✻ Age	Greenhouse-Geisser	5.049	1.928	2.619	2.706	.070	0.010
Residuals	Greenhouse-Geisser	498.211	514.790	0.968			

Note. Type III Sum of Squares

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated (p < .05).

Table 4

Post-Hoc Comparisons t Tests of The Effect of AI Emotion Type (Happy, Fearful and Angry) in the Different Conditions (Training vs. Control)

Post Hoc Comparisons - Condition ✻ Emotion shown AI ▼

		Mean Difference	SE	df	t	p_{bonf}
Control, Scared	Training, Scared	−0.875	0.135	267	−6.474	< .001
	Control, Angry	0.372	0.102	267	3.653	.005
	Training, Angry	−0.704	0.141	267	−5.007	< .001
	Control, Happy	0.783	0.113	267	6.912	< .001
	Training, Happy	−0.604	0.156	267	−3.871	.002
Training, Scared	Control, Angry	1.247	0.140	267	8.934	< .001
	Training, Angry	0.171	0.112	267	1.525	1.000
	Control, Happy	1.658	0.153	267	10.872	< .001
	Training, Happy	0.271	0.125	267	2.173	.460
Control, Angry	Training, Angry	−1.076	0.145	267	−7.429	< .001
	Control, Happy	0.411	0.121	267	3.401	.012
	Training, Happy	−0.976	0.160	267	−6.106	< .001
Training, Angry	Control, Happy	1.487	0.157	267	9.451	< .001
	Training, Happy	0.100	0.133	267	0.751	1.000
Control, Happy	Training, Happy	−1.387	0.171	267	−8.100	< .001

Note. P-value adjusted for comparing a family of 15 estimates.

Table 5

RM ANOVA of the non Significant Interaction Between L-EFT Scores and Emotional Content and the non Significant Effect of L-EFT Scores on AI Detection Rate

Within Subjects Effects

Cases	Sphericity Correction	Sum of Squares	df	Mean Square	F	p
Depicted Emotion AI Picture	None	5.230[a]	2.000[a]	2.615[a]	2.778[a]	.063[a]
	Greenhouse-Geisser	5.230	1.929	2.712	2.778	.065
Depicted Emotion AI Picture ✻ Condition	None	7.879[a]	2.000[a]	3.939[a]	4.184[a]	.016[a]
	Greenhouse-Geisser	7.879	1.929	4.085	4.184	.017
Depicted Emotion AI Picture ✻ LEFT score total	None	1.005[a]	2.000[a]	0.502[a]	0.533[a]	.587[a]
	Greenhouse-Geisser	1.005	1.929	0.521	0.533	.580
Residuals	None	491.460	522.000	0.941		
	Greenhouse-Geisser	491.460	503.412	0.976		

Note. Type III Sum of Squares

[a] Mauchly's test of sphericity indicates that the assumption of sphericity is violated (p < .05).

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p
Condition	232.172	1	232.172	71.027	< .001
LEFT score total	6.541	1	6.541	2.001	.158
Residuals	853.153	261	3.269		

Note. Type III Sum of Squares

Table 6

Exploratory Analysis: Descriptives of Older and Younger Adults.

Descriptives - Total_Pictures_Correct

Condition	Above_45_y/o	N	Mean	SD	SE	Coefficient of variation
Control	0	136	36.82	6.026	0.517	0.164
	1	12	23.08	5.213	1.505	0.226
Training	0	113	42.22	4.050	0.381	0.096
	1	9	38.89	5.904	1.968	0.152

Note. Age is coded as 0 = younger than 45 years, 1 = 45 years or older

Table 7

Exploratory Analysis: Post Hoc Comparison t Tests for Differences Between Older and Younger Adults with Regards to Condition

Post Hoc Comparisons - Condition ✻ Above_45_y/o ▼

		Mean Difference	SE	df	t	Cohen's d	p_{bonf}
Control 0	Training 0	−5.405	0.668	266	−8.096	−1.031	< .001
	Control 1	13.733	1.579	266	8.694	2.618	< .001
	Training 1	−2.073	1.805	266	−1.148	−0.395	1.000
Training 0	Control 1	19.138	1.592	266	12.018	3.649	< .001
	Training 1	3.332	1.817	266	1.834	0.635	.406
Control 1	Training 1	−15.806	2.313	266	−6.834	−3.013	< .001

Note. P-value adjusted for comparing a family of 6 estimates.

Note. Age is coded as 0 = younger than 45 years, 1 = 45 years or older

Figures

Figure 1

The Effect of Training on Overall Detection Rate

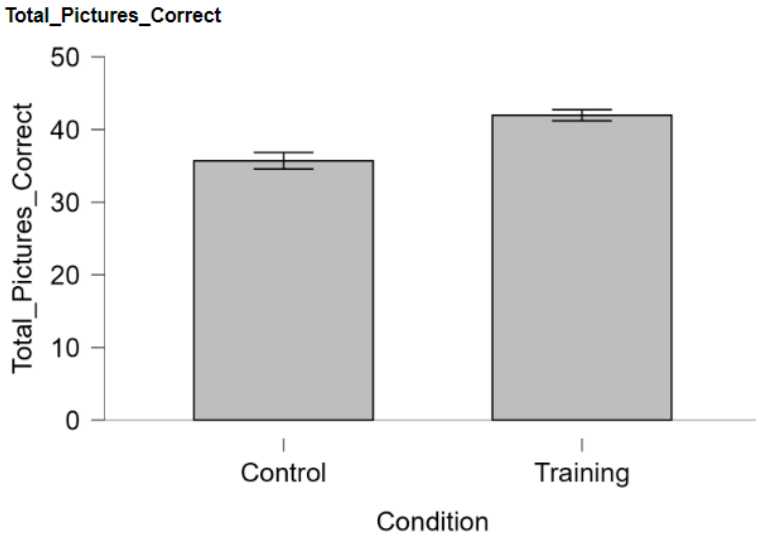


Figure 2

Histogram Illustrating the Distribution of the LEFT Scores

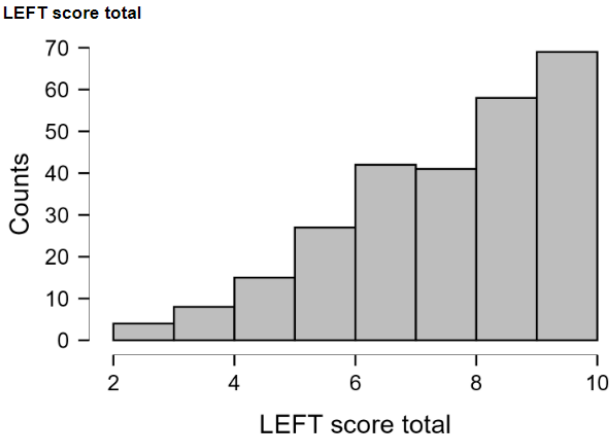


Figure 3

Interaction Between Condition and Depicted AI Emotion Type (Happy, Fearful and Angry)

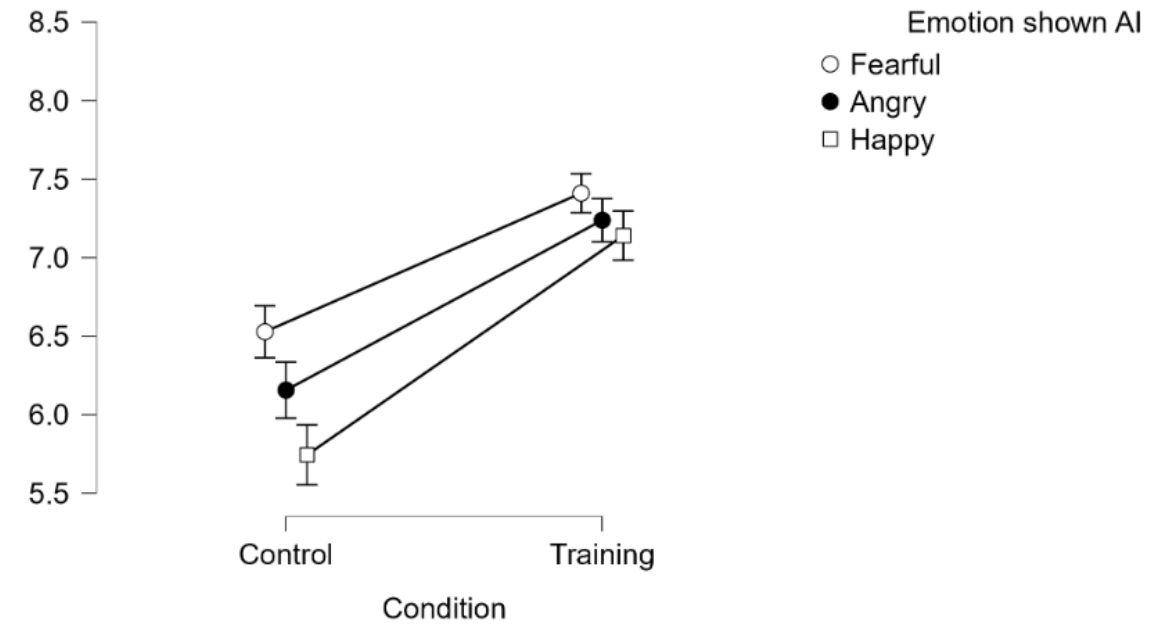
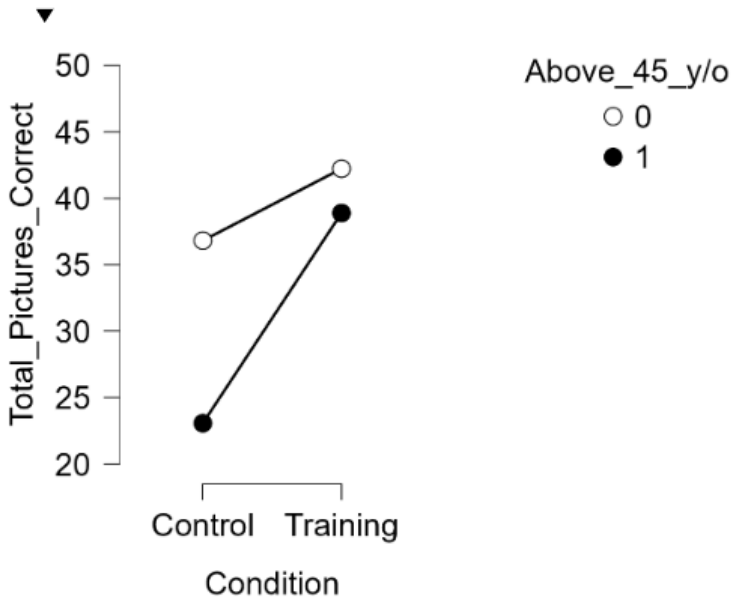


Figure 4

Exploratory Analysis: Interaction Between Age group and Condition



Note. Age is coded as 0 = younger than 45 years, 1 = 45 years or older

Appendix A: Examples of Prompts

The following are examples of prompts used in this study, organised by category (male vs. female, attractiveness and emotional type). These prompts (and variations of these prompts) were used to create the AI generated images.

Females

Angry (Attractive)

Make a realistic photo of a 40 year old white woman. She has an angry expression, wearing a grey t-shirt (round neck) against a pure white background. Natural skin texture, minimal makeup, unstyled hair, and realistic lighting.

Angry (Unattractive)

Make a frontal photograph of an ugly-looking black woman (45 years old) with an irritated expression, wearing a grey t-shirt (round neck) against a white background. Uneven skin texture, tired eyes, unstyled long hair, no makeup, realistic lighting and colors, round, chubby face.

Happy (Attractive)

Make a front-facing photograph of a middle aged attractive black woman with a happy expression (closed mouth), wearing a grey t-shirt (round neck) against a white background. Minimal make-up, no jewelry and realistic lighting.

Happy (Unattractive)

Create a frontal photograph of an ugly-looking young black woman with a happy expression (open mouth), wearing a grey t-shirt with a round neck against a completely white background. Her face must be round and full, she should not be wearing make up and her teeth should be a little crooked.

Fear (Attractive)

Create a picture of an attractive middle-aged white woman which could be used for a database for emotions. The picture should have a white background and she should be

wearing a grey shirt with a round neck. She should look scared, but realistically. She should have her hair up and have no jewelry. Her shirt and shoulders should be visible.

Fear (Unattractive)

Make a picture of an unattractive adolescent white woman with acne (23 years old) who is considered ugly exhibiting the emotion fear, she has to be terrified. Completely white background. Gray shirt (round neck). Should not be too close up (shirt and shoulders visible). Realistic human with facial asymmetry, and add some blemishes in the skin. Not too pretty looking. Not too much light and reduce shininess on the face.

Males

Angry (Attractive)

Make a realistic photo of a 40 year old white man. He has an angry expression, wearing a grey t-shirt with a round neck against a pure white background. Natural skin texture and realistic lighting.

Angry (Unattractive)

Create a frontal photograph of an unattractive adolescent black man with an angry/irritated expression, wearing a grey t-shirt with a round neck against a completely white background. He is overweight and has skin blemishes. Shirt and shoulders should be visible.

Happy (Attractive)

Make a front-facing picture of an attractive white man (25 years old) who is smiling (mouth closed) exhibiting the emotion of being happy. Emotion recognition task theory of mind database. Grey shirt with a white neck and pure white background. Not too zoomed in, his shirt and shoulders should be visible. No additional shine on the face.

Happy (Unattractive)

Produce an image of an ugly middle-aged white man who is happy (smiling with mouth open). He should have facial asymmetry, eye bags, skin blemishes, a grey shirt with a

round neck and a fringe. The background needs to be white. Should be indistinguishable from reality.

Fear (Attractive)

Image of an attractive black man (30 years old) looking scared against a white background. The picture should be usable in a database for emotion recognition. Gray shirt with a rounded neck. Shoulders should be visible. Realistic lightning and reduce shininess on the face.

Fear (Unattractive)

Make a front facing image of a black man (20 years old) who is considered ugly and unattractive. He should be looking frightened. Give him facial asymmetry and acne. He should be wearing a grey shirt with a round neck and his shoulders should be in the picture. Do not add any jewelry or additional facial shininess.

Appendix B: Examples of AI Images Used in the Study

AI Images

Angry Attractive Woman

An example of an AI image of an attractive woman with an angry expression.

Note. This image was generated through Midjourney.

Angry Unattractive Woman

An example of an AI image of an unattractive woman with an angry expression.

Note. This image was generated through Dall-E.

Happy Attractive Woman

An example of an AI image of an attractive woman with a happy expression.

Note. This image was generated through Midjourney.

Happy Unattractive Woman

An example of an AI image of an unattractive woman with a happy expression.

Note. This image was generated through Midjourney.

Fearful Attractive Woman

An example of an AI image of an attractive woman with a fearful expression.

Note. This image was generated through Midjourney.

Fearful Unattractive Woman

An example of an AI image of an unattractive woman with a fearful expression.

Note. This image was generated through Dall-E.

Angry Attractive Man

An example of an AI image of an attractive man with an angry expression.

Note. This image was generated through Midjourney.

Angry Unattractive Man

An example of an AI image of an unattractive man with an angry expression.

Note. This image was generated through Midjourney.

Happy Attractive Man

AI image of an attractive man with a happy expression.

Note. This image was generated through Midjourney.

Happy Unattractive Man

An example of an AI image of an unattractive man with a happy expression.

Note. This image was generated through Midjourney.

Fearful Attractive Man

An example of an AI image of an attractive man with a fearful expression.

Note. This image was generated through Midjourney.

Fearful Unattractive Man

An example of an AI image of an unattractive man with a fearful expression.

Note. This image was generated through Dall-E.