Leadership, Learning, and Business Beyond Boundaries: Managing Change in a Complex World
21–23 May 2026
Ljubljana, Slovenia
make learn TIIM PIConf
Management, Knowledge and Learning International Conference
Technology, Innovation and Industrial Management
Sustainable Development

Perceptual accuracy and confidence in detecting AI-generated images among adolescents in Slovenia

Martina Breg
ISSBS, Slovenia
bregmartina@gmail.com

izr. prof. dr. Andrej Flogie
FNM, Slovenia
andrej.flogie@um.si

Abstract

Recent advances in generative artificial intelligence have significantly changed the way digital images are created and interpreted, raising questions about how reliably people can judge their origin. This study examines how accurately Slovenian adolescents distinguish between human-created and AI-generated images, and whether recognition performance differs across visual categories. In addition, the study explores the relationship between actual performance and self-reported confidence in order to better understand the relationship between perceived and actual performance in this context. The final sample consisted of 98 participants aged 16 to 20. Participants evaluated 18 visual stimuli from three categories—photography, illustration, and graphic design. For each image, they indicated whether it was human-made or AI-generated and rated their confidence in the decision. The data were analysed using a one-sample t-test, paired-samples t-tests, and Pearson's correlation. The results indicate that overall accuracy was significantly above chance level. A weak but statistically significant positive correlation was found between confidence and accuracy. Differences between visual categories were small and not statistically significant. The analysis further showed that participants tended to underestimate their performance, as confidence ratings were significantly lower than actual accuracy. Overall, the findings suggest that adolescents are able to distinguish between human-created and AI-generated images better than chance, although the relationship between perceived and actual performance remains limited. The study contributes to a more nuanced understanding of perceptual accuracy and confidence in the context of generative AI and highlights the importance of developing critical evaluation skills for digital images.

Keywords: generative artificial intelligence, perceptual accuracy, confidence, adolescents, AI-generated images, visual literacy

INTRODUCTION

The rapid development of generative artificial intelligence models, multimodal and diffusion models, and text-to-image systems has significantly transformed the way digital images are created and understood in recent years. Visual images are no longer necessarily the result of human creative intent or a representation of existing reality, but can instead constitute synthetically generated visual constructs. This shift raises important questions about the reliability of human judgment regarding image authenticity, as well as the adequacy of existing concepts of visual literacy in the context of generative artificial intelligence.

Empirical studies show that the human ability to distinguish between artificially generated and authentic images is often limited, and for certain types of images approaches the level of random guessing. At the same time, research highlights a discrepancy between actual performance and subjective confidence in judgment, raising questions about metacognitive calibration in the interpretation of visual content in digital environments.

Despite the growing number of studies, several important gaps remain in the existing literature. First, most studies to date have been conducted on adult populations, while adolescents as a developmentally specific group remain relatively underexplored. Second, existing models of digital visual literacy have primarily been developed in the context of manipulating existing images, and less so in the context of fully AI-generated visual content. Third, the literature suggests that the perception of synthetic images is not uniform across categories of visual material, but it remains unclear whether these effects are similarly expressed among adolescents.

The aim of the study is to examine the extent to which adolescents can distinguish between AI-generated images and those created by humans, and whether the category of the visual product influences recognition performance. The study additionally analyzes the relationship between perceptual performance and confidence in judgment, thereby shedding light on the metacognitive calibration of adolescents in the context of generative artificial intelligence.

2. THEORETICAL BACKGROUND AND LITERATURE REVIEW

Due to the development of TTI, distinguishing between real and AI-generated visual elements is becoming increasingly difficult for both experts and the general public. Automated detection methods can achieve high accuracy in controlled environments, but they face diverse datasets and conflicting conditions, highlighting the need for more robust, adaptable, and ethically guided detection frameworks (Saha et al., 2023; Abate et al., 2025).

Empirical research shows that human observers are often unreliable in determining the origin of images. An experimental study (Högemann et al., 2025) found that participants frequently achieved results close to random guessing when distinguishing between real and AI-generated images, especially when images were created using advanced generative models. The average detection rate was approximately 63.7%, indicating that human judgment remains limited and uncertain.

Similar findings are observed in the perception of art in the age of artificial intelligence. A recent study by van Hees et al. (2025) shows that participants were able to distinguish between human and AI-generated art above chance level; however, they surprisingly showed a greater preference for AI-

generated works. In discussions of authenticity and originality, researchers increasingly refer to the concept of AI-generated media literacy, which relates to the ability to critically evaluate and recognize artificially created visual content (López-Borrull & Lopezosa, 2025).

A comprehensive review of deep learning-based image generation models was conducted by Li et al. (2025), describing advancements and different model types. Particular attention was given to diffusion models, which have recently gained prominence in image generation due to their specific generative processes and high performance in producing visual content. Diffusion models have become a central technology in generative artificial intelligence because of their high visual quality and flexibility. With the development of generative models for creating digital images, concerns about their potential misuse for manipulation and deception have increased (Chen et al., 2025). The improved realism of images generated by diffusion and multimodal models has reduced the presence of visual anomalies that previously served as indicators of artificial origin, making it more difficult for human observers to detect synthetic images based solely on visual cues.

The development of generative models affects not only visual production but also the social perception of authenticity and trust, particularly in the context of media, education, and digital platforms. The issue extends beyond the technological domain, as it directly impacts individuals' ability to critically evaluate visual information. Adolescents today are growing up in a visual environment where digital images are no longer necessarily the result of human creative processes, but of collaboration between humans and algorithms. Consequently, the criteria for aesthetic judgment, authorship, and authenticity are changing. As noted by Zhang & Xu (2025), adolescents are growing up in an environment where digital images are the result of collaboration between humans and algorithms, significantly reshaping standards of aesthetic evaluation, authorship, and authenticity.

Research on the recognition of AI-generated facial images is particularly concerning. Synthetic faces remain indistinguishable from real ones for observers and are often perceived as more trustworthy. This suggests that people frequently cannot reliably determine whether a facial image is real or AI-generated. Findings indicate that recognition performance in some cases is close to random guessing (around 50%) (Nightingale & Farid, 2022). Researchers warn that technological advances in image generation are approaching a point where human ability to detect image origin becomes increasingly limited, nearly indistinguishable (van Hees et al., 2025).

Nevertheless, individuals sometimes rely on specific visual anomalies to improve detection performance, such as irregularities in textures, lighting, or geometry. However, these indicators are becoming less noticeable as models advance (Velásquez-Salamanca et al., 2025). AI-generated images are also described as "too perfect" or slightly unusual, which influences judgments of authenticity. In the field of visual perception, some studies have found that human-created art is valued more highly than AI-generated art, even when people cannot explicitly distinguish between them. This suggests an intuitive preference for human authorship (Oksanen et al., 2023). More recent research (Bianchi et al., 2025) indicates that AI-generated visual content is often misidentified, affecting perceptions of credibility, authenticity, and trust.

Based on these gaps, further research is needed to better understand how adolescents perceive and evaluate AI-generated visual content.

2.1 Visual Literacy and Adolescents

Visual literacy encompasses the ability to interpret, analyze, and critically evaluate visual information. The concept of digital visual literacy, proposed by Avgerinou and Ericson in 1997, emphasizes the integration of visual literacy into digital contexts. This includes the ability to interpret, create, and communicate visual messages using digital tools (Mixer et al., 2008). Visual literacy is considered a key component of digital literacy, enabling learners to engage deeply with content through visual means (Mixer et al., 2008; Marzal García-Quismondo & Parra Valero, 2021). In the context of generative artificial intelligence, this competence is becoming essential, as traditional visual indicators of authenticity are gradually disappearing.

Spalter and van Dam (2008) define digital visual literacy (DVL) as the ability to critically evaluate digital visual materials, make decisions based on visual representations of data and ideas, and use computers to create effective visual communication. They emphasize that visual literacy is not equivalent to other literacies, such as media literacy or digital literacy. The authors highlight the necessity of a critical approach to visual materials in the digital age, as photo manipulation can distort reality. As an example, they cite a photograph from Beirut in 2006, referenced in the article *Digital Literacy*, in which Reuters published a manipulated image that falsely depicted the extent of damage following a bombing. This example underscores the importance of developing critical skills for recognizing manipulative visual content. Long and Magerko (2020) proposed a framework encompassing key components of AI literacy, including technical understanding, practical application, critical evaluation, and ethical considerations (O'Dea & Ng, 2024).

Adolescents represent a particularly relevant group, as they are intensive users of digital visual platforms, while their capacity for critical visual judgment is still developing (Liao et al., 2025). This development is crucial in the context of AI-generated content, where an understanding of basic design principles supports the critical evaluation of visual materials and their effective use across different contexts. Van Hees et al. (2025) further elaborate the aesthetic dimension with findings on human perception of AI-generated works. Their research indicates the importance of distinguishing between the ability to differentiate and aesthetic preferences. Although participants were able to identify AI-generated works above chance level, their aesthetic preferences did not necessarily reflect these distinctions. This highlights the need for a comprehensive approach to visual literacy that encompasses not only the technical ability to distinguish but also an understanding of aesthetic preferences and biases in the context of AI-generated content (Högemann et al., 2025).

Context also plays a key role in the use of AI-generated images. For example, in educational settings, AI-generated visual materials can enhance engagement and accelerate idea generation, although they may also introduce challenges related to authorship and critical digital literacy (Velásquez-Salamanca et al., 2025).

Findings from a study conducted among Slovenian secondary school students (Licardo et al., 2025) reveal notable patterns in the use of generative AI tools. Students most frequently use text-generation technologies, with nearly one third reporting occasional use, while almost half use such tools weekly or even multiple times per week. Among those who specified which tools they use, the most commonly mentioned were ChatGPT, Claude AI, Copilot, and, somewhat unexpectedly, Snapchat AI and Astra AI, although the latter were reported only by a small proportion of respondents.

A markedly different pattern emerges in the use of generative AI for creating visual and audio content. The majority of students—nearly two thirds—never use image-generation tools such as DALL·E or MidJourney. Even more pronounced is the reluctance to create audio and video content, with almost four out of five students reporting that they do not use these functionalities at all. These findings suggest a clear preference among young users for text-based AI applications, while other forms of generative content remain limited to a minority.

Despite the relatively low use of visual content generation tools, an important question arises regarding the long-term effects of exposure to such technologies. The central research problem concerns whether existing models of visual literacy adequately explain the perception, interpretation, and evaluation of generative visual images among adolescents, and whether generative artificial intelligence introduces structural changes in the cognitive and interpretative processes of visual perception.

A discrepancy between confidence and performance is expected, with the possibility of underconfidence, particularly in situations characterized by high uncertainty or task complexity. In the context of generative artificial intelligence, where the boundaries between authentic and synthetic content are increasingly blurred, individuals may express lower confidence despite relatively high actual performance.

2.2 The role of visual content categories

Existing research in cognitive psychology and visual perception suggests that recognition performance is not uniform across all types of images. The human perceptual system processes different visual categories with varying levels of efficiency and accuracy, which has important implications for distinguishing between AI-generated and human-created content. Studies on face recognition have shown that images of human faces require more complex processing than natural landscapes or objects. This finding is particularly relevant in the context of AI-generated portraits, where subtle anomalies in proportions may serve as key indicators of AI-generated origin (Nightingale & Farid, 2022).

AI-generated artworks, especially abstract art, can achieve high aesthetic quality and viewer preference when they are carefully aligned with user-provided inputs (Hou & Huang, 2025). In the context of graphic design, early studies (Chamberlain et al., 2018) have shown that people can perceive design elements such as typographic hierarchy and color harmony within as little as 50 milliseconds. This suggests that, in graphic design, observers may be able to detect deviations from established design principles more quickly—particularly when AI systems fail to reproduce them consistently in complex compositions.

Existing research further indicates that perceptual processing is not uniform across different types of visual content. Different categories of visual material may involve distinct perceptual mechanisms and varying degrees of recognizable cues of AI-generated origin. Photographic images, due to their high level of realism, may make it more difficult to assess their origin, whereas in illustration and graphic design, certain compositional or stylistic features may remain more noticeable.

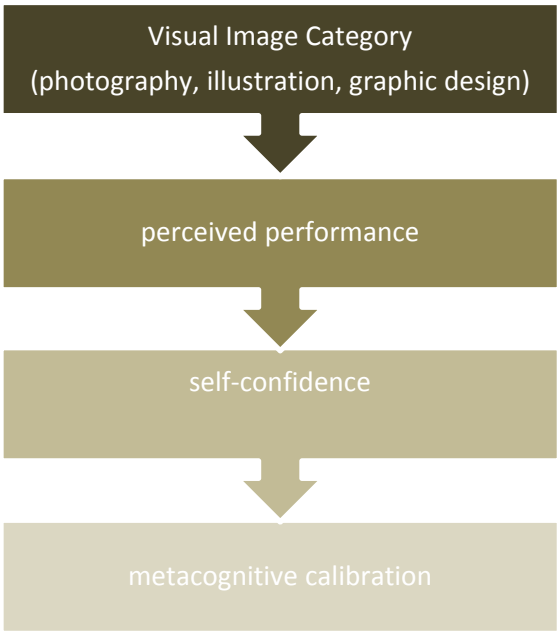
On this basis, the study examines whether the category of visual content influences the ability to recognize the origin of digital images. It does not necessarily assume a strong categorical effect but rather investigates whether statistically significant differences emerge between photography, illustration, and graphic design.

Based on these differences, it can be expected that perceptual difficulty may vary across categories. Photographic images may be more difficult to evaluate due to their high realism, while illustration and graphic design may contain more detectable stylistic or compositional cues.

2.3 Conceptual research model

Based on the literature review, a conceptual research model was developed, which assumes that perceptual performance in judging the origin of digital images depends not only on the individual's perceptual abilities, but also on the category of the visual product. In addition to perceptual performance, the model also includes a metacognitive dimension, expressed in the level of self-confidence and the relationship between perceived and actual performance. Thus, the model connects the perceptual and metacognitive aspects of judging digital images in the context of generative artificial intelligence.

Figure 1: Conceptual research model



Author's own conceptualization based on the literature

3. RESEARCH PROBLEM AND OBJECTIVES

The rapid development of generative artificial intelligence models, particularly diffusion models and text-to-image systems, has significantly reshaped the epistemological status of digital images. An image is no longer necessarily a representation of existing reality or the result of human creative intent, but can instead constitute a synthetically generated visual construct without a referent in the physical world. The research problem arises from the question of whether adolescents can reliably distinguish between AI-generated and human-created images, and whether this ability is influenced by the category of visual material.

Empirical research indicates that human ability to distinguish between synthetic and authentic images approaches the level of random guessing, particularly in the case of photorealistic faces (Nightingale & Farid, 2022; Högemann et al., 2025). At the same time, research in the field of aesthetic judgment highlights a perceptual paradox: observers often express a high degree of confidence in their judgments, even when this confidence is not aligned with actual accuracy (van Hees et al., 2025).

Three key gaps remain in the existing literature. First, most studies have been conducted on adult populations, while adolescents as a developmentally specific group remain relatively underexplored. Adolescents are growing up in an environment where AI-generated images are part of everyday visual experience, which may influence the development of their cognitive evaluation strategies.

Second, existing models of digital visual literacy (Spalter & Van Dam, 2008) were developed in the context of manipulating existing images rather than in the context of fully AI-generated visual realities. It remains unclear whether traditional indicators of authenticity (e.g., detection of anomalies in lighting, texture, or proportions) still represent relevant cognitive strategies in the era of advanced diffusion models.

Third, the literature suggests that perception is not uniform across visual categories. Face recognition studies indicate a high level of perceptual vulnerability (Nightingale & Farid, 2022), while research in graphic design shows that observers can rapidly detect compositional and design principles (Chamberlain et al., 2018). However, it has not been empirically established whether these categorical effects are similarly expressed among adolescents.

Based on the defined research problem, the theoretical framework of visual literacy, and empirical findings on the perception of AI-generated images, the study systematically examines adolescents' perceptual performance and metacognitive calibration when evaluating the origin of digital visual content.

The study contributes to existing research on three levels. First, it focuses on adolescents as a population that has been less frequently included in previous studies. Second, it compares recognition performance across three categories of visual content: photography, illustration, and graphic design. Third, it analyzes the relationship between response accuracy and confidence, thereby providing insight into adolescents' metacognitive calibration in the evaluation of visual content.

3.1 Research questions

RQ1: To what extent does adolescents' perceptual performance in judging the origin of digital images statistically deviate from the level of random guessing?

RQ2: What is the relationship between self-perceived confidence in judgment and the actual accuracy of responses?

RQ3: Does the category of visual content influence the ability to recognize the origin of a digital image?

RQ4: What is the discrepancy between perceptual performance and self-perceived competence, and what does it reveal about adolescents' metacognitive calibration?

3.2 Hypotheses

H1: The average perceptual accuracy of adolescents will be statistically significantly higher than the level of random guessing (50%).

H2: There will be a positive correlation between confidence in judgment and actual response accuracy.

H3: Categorical effect hypothesis:

H3a: Recognition accuracy will be lowest for photographs.

H3b: Recognition accuracy will be higher for graphic design than for photographs.

H4: Average confidence in judgment will be statistically significantly higher than actual perceptual accuracy.

4. METHODOLOGY

4.1 Research plan

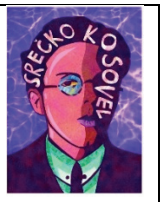
The study was designed as a quantitative experimental study with a within-subject design. Participants evaluated visual stimuli from three categories: photography, illustration, and graphic design. This design enabled a direct comparison of recognition performance across categories while controlling for individual differences.

The stimulus set consisted of a total of 18 digital visual images, divided into three categories: photography (n = 6), illustration (n = 6), and graphic design (n = 6). Within each category, three images were human-created and three were generated using artificial intelligence.

The human-created works were sourced from the researcher's own collection and from the archive of the Secondary School of Design Maribor, where they were created under the researcher's mentorship. The authors of the works consented to their use for research purposes. To reduce potential bias, the authors of the selected works were excluded from the sample.

Synthetic images were generated using generative artificial intelligence tools, including GPT-4o, Imagen, and Midjourney. Although the selection was based on expert judgment, strict criteria of visual comparability (realism, composition, color structure) were consistently applied in order to minimize potential systematic differences between stimulus groups. The aim of the selection process was to ensure the highest possible comparability between human-created and AI-generated images within each category.

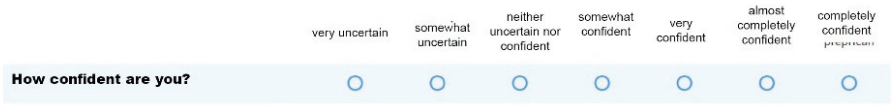
Table 1: Some examples of visual stimuli from the online questionnaire

 Photography	 Photography	 Graphic design (logo)	 Graphic design (logo)	 Digital illustration	 Digital illustration

adapted by author

After each visual stimulus decision, participants also provided a confidence rating. Since the confidence rating was in Slovenian on a 7-point Likert scale, it was translated into English for the purposes of this article.

Figure 2: Indicate how confident you are in your decision



Example of a 7-point Likert scale from an online questionnaire /adapted by author

4.2 Sample

The study involved adolescents aged between 16 and 20 years. The total sample size was N = 112, of which 98 valid responses were included in the final analysis.

Initially, 112 participants were included; however, after data cleaning, the final analyzed sample was reduced to 98 valid responses. The exclusion of 14 cases was based on predefined criteria, including missing or invalid data, incomplete questionnaires, and clearly inconsistent responses. This approach increased the reliability and validity of the subsequent analyses. The mean age of participants was 17.29 years (SD = 0.99), indicating a relatively homogeneous age structure of the sample.

A priori power analysis ($\alpha = .05$, power = .80) indicated that the planned sample size was sufficient to detect medium-sized effects.

Demographic data were also collected, along with information on participants' engagement with visual arts and their prior experience with generative AI tools. Participation was voluntary. The order of stimulus presentation was randomized. For underage participants, appropriate consent was obtained in accordance with institutional and legal procedures. The study was conducted in line with ethical guidelines for research involving human participants.

4.3 Measured variables

Perceptual performance
Responses were coded binarily (1 = correct response, 0 = incorrect response). The proportion of

correct responses was calculated at the level of individual trials and aggregated across categories of visual stimuli. The reference value for comparison was the level of random guessing (.50).

Confidence in judgment
After each classification, participants rated their confidence on a 7-point Likert scale (1 = very uncertain, 7 = completely certain). For individual analyses, mean confidence was calculated. For comparisons with actual performance, confidence ratings were normalized to a 0–1 interval to ensure comparability with proportions of correct responses.

4.4 Procedure

Data collection was conducted in a controlled school environment using an online questionnaire. In the introductory section, participants provided demographic information and answered questions about their experience with generative artificial intelligence.

The experimental part of the study was based on a two-alternative forced choice (2AFC) paradigm. For each visual stimulus, participants determined its origin (human-created or AI-generated) and rated their confidence in the correctness of their decision.

The questionnaire also included an attention check question. Participants responded without time constraints.

The entire procedure lasted an average of 6 minutes and 49 seconds.

4.5 Statistical Analysis

Statistical analyses were conducted using the 1KA survey platform, Microsoft Excel, and the statistical software R and Jamovi. To test hypothesis H1, a one-sample t-test was performed against the reference value of 0.50. To test hypothesis H2, Pearson's correlation coefficient was calculated between confidence levels and response accuracy. To test hypotheses H3a and H3b, paired-samples t-tests were used. Differences between categories were analyzed using paired t-tests, as the aim was to directly compare specific pairs of conditions.

Due to multiple pairwise comparisons across the three categories of visual stimuli, a Bonferroni correction was applied to control for Type I error. The threshold for statistical significance was therefore set at $\alpha = 0.0167$.

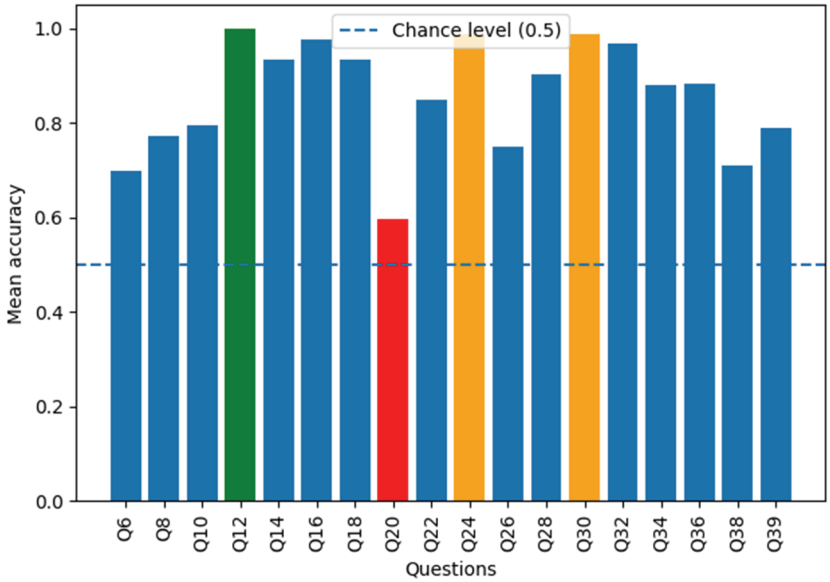
To test hypothesis H4, a paired-samples t-test was also used to compare mean confidence and actual perceptual performance. Perceptual performance was expressed as the proportion of correct responses (0–1), calculated across all tasks or within individual categories of visual stimuli. Confidence was measured on a 7-point Likert scale and, for comparison with actual performance, normalized to a 0–1 interval.

5. RESULTS

The results of the one-sample t-test showed that the average perceptual performance of adolescents was statistically significantly higher than the level of random guessing ($M = 0.76$, $SD = 0.15$), $t(97) = 17.25$, $p < .001$, with a very large effect size (Cohen's $d = 1.73$).

Based on these results, hypothesis H1 was confirmed.

Figure 3: Mean identification across stimuli



adapted by author

The lowest performance was observed for question Q20 in the photography category (M = 0.60).

Figure 4: The photograph was generated using the GPT-4o model, a multimodal artificial intelligence system that enables integrated processing and generation of different types of data.

adapted by author

In contrast, question Q12 (digital illustration) achieved 100% accuracy, meaning that all participants correctly identified its AI origin.

Figure 5: The illustration was also generated using the GPT-4 model.

adapted by author

Among the highest-performing items were questions Q24 and Q30 (M ≈ 0.99), which belong to the graphic design category. The visual images were created using the DALL·E diffusion model and through human design using Adobe Illustrator.

Table 2: Examples of questions Q24 and Q30 in the graphic design category

	
Source of the image generated using the DALL·E diffusion model	Source of the image created by a human (Adobe Illustrator)

adapted by author

Descriptive analysis of perceptual performance across categories of visual stimuli showed relatively small differences between categories. The highest performance was observed for illustration (M = 0.77), followed by graphic design (M = 0.76), while the lowest performance was found for photography (M = 0.75).

Although the differences between categories were small, the descriptive results indicate a slight advantage in recognizing the origin of images in the illustration category. This pattern may suggest that differences between human-created and AI-generated images were somewhat more perceptible in this category compared to others.

To test differences between categories, paired-samples t-tests were conducted. Due to three multiple comparisons, a Bonferroni correction was applied, setting the threshold for statistical significance at $\alpha = 0.0167$. The results showed no statistically significant differences between categories: performance for photographs did not differ significantly from illustrations, $t(97) = 0.20$, $p = .844$; there were also no significant differences between graphic design and photography, $t(97) = 0.46$, $p = .645$, nor between graphic design and illustration, $t(97) = 0.81$, $p = .423$.

Based on these results, hypotheses H3a and H3b cannot be confirmed.

To examine the relationship between confidence and response accuracy, Pearson's correlation coefficient was calculated. The results indicate a statistically significant positive correlation between confidence and accuracy, $r = 0.23$, $p < .001$. The effect size is small, meaning that while higher confidence was associated with a higher likelihood of correct responses, the relationship was weak. Based on these findings, hypothesis H2 can be confirmed.

Additional analysis showed that 23.3% of all responses were cases in which participants were completely confident in their judgment and also provided a correct answer. These results indicate a limited alignment between perceived and actual performance.

To test the difference between confidence and actual perceptual performance, a paired-samples t-test was conducted. The results showed that average confidence was statistically significantly lower than actual perceptual performance, $t(97) = -4.95$, $p < .001$. This indicates a pattern of underconfidence among participants. Based on these findings, hypothesis H4 was not supported, as the direction of the difference was opposite to the one initially hypothesized.

6. DISCUSSION

The results of the study indicate that adolescents achieve relatively high perceptual accuracy when judging the origin of digital images, as their performance was significantly above chance level. This suggests that participants were able to distinguish between human-created and AI-generated images better than would be expected by random guessing. However, this overall level of performance does not necessarily imply consistent perceptual reliability, as the results at the level of individual stimuli revealed considerable variability. The higher accuracy observed in comparison to some previous studies may be partly explained by the specific selection of stimuli or the controlled research context.

Descriptive results showed only minor differences between categories of visual stimuli, with the highest accuracy observed in illustration and the lowest in photography. These differences were not statistically significant, indicating that the type of visual medium cannot be considered a key determinant of perceptual performance. Instead, the difficulty of the task appears to depend more on specific characteristics of individual images (e.g., level of realism, visual complexity, or presence of artifacts) than on broader categorical distinctions. This finding suggests that the initial assumption of systematic category-based differences may have been overly simplified.

In addition to perceptual accuracy, the relationship between confidence and response correctness was examined. The analysis revealed a statistically significant but weak positive correlation, indicating that higher confidence was associated with a slightly higher likelihood of correct responses. However, confidence was not a reliable predictor of accuracy. These findings point to a limited alignment between perceived and actual performance, which may reflect partial or context-dependent metacognitive calibration. This interpretation is further supported by the finding that only 23.3% of responses were both fully confident and correct.

A particularly important result relates to hypothesis H4, which assumed overconfidence. Contrary to expectations, the findings revealed a statistically significant pattern of underconfidence, as participants' average confidence was lower than their actual performance. This finding represents a deviation from the dominant pattern of overconfidence reported in previous studies and suggests that metacognitive calibration may depend on task uncertainty.

This suggests that adolescents did not overestimate their abilities but instead adopted a cautious approach when making judgments. One possible explanation is the high level of uncertainty associated with evaluating AI-generated visual content, where traditional visual cues of authenticity are less reliable.

These findings contrast with a substantial body of literature that often reports overconfidence in perceptual judgments. Instead, the present results indicate that metacognitive biases in this context may be more complex and context-dependent. Under conditions of uncertainty, individuals may become more conservative in their self-assessments, leading to underconfidence rather than overconfidence.

The interpretation of metacognitive processes in this study remains limited and should be considered exploratory, as the research design does not include direct measures of metacognitive sensitivity.

The study has several limitations. First, the sample size is relatively small and gender-imbalanced, which may limit the generalizability of the findings. Second, the selection of visual stimuli was based on expert judgment, which introduces a degree of subjectivity. Third, the study was conducted within a specific cultural and educational context, which may further restrict the applicability of the results to other populations.

Despite these limitations, the findings have important implications for education and the development of visual literacy. The results suggest that perceptual discrimination alone is not sufficient and that it is equally important to develop strategies for evaluating and reflecting on one's own judgments. This is particularly relevant in the context of integrating generative artificial intelligence into educational settings, where the ability to critically assess visual content is becoming an increasingly important competence.

Overall, the results indicate that while adolescents demonstrate relatively high perceptual accuracy, the relationship between perceived and actual performance remains complex. This highlights the importance of further research into metacognitive aspects of visual judgment, particularly in the context of the growing presence of generative artificial intelligence in the visual environment.

7. CONCLUSION

The results of the study emphasize that, although adolescents are relatively successful in distinguishing between AI-generated and human-created images, the relationship between confidence and actual performance remains limited. This indicates the need for the systematic development of critical visual literacy and a deeper understanding of generative technologies, which are increasingly shaping the contemporary visual environment.

The authors acknowledge the use of artificial intelligence tools (e.g., ChatGPT) for language refinement. The authors are fully responsible for the integrity and accuracy of the work.

REFERENCES

Abate, A., Cimmino, L., & Polsinelli, M. (2025). Comparative Evaluation of Synthetic Image Detectors: Insights from Generative Adversarial Networks and Stable Diffusion Generators. In Lecture Notes on Data Engineering and Communications Technologies (Vol. 252, pp. 426–436). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-031-87784-1_39

Bianchi, I., Branchini, E., Uricchio, T., & Bongelli, R. (2025). Creativity and aesthetic evaluation of AI-generated artworks: Bridging problems and methods from psychology to AI. Frontiers in Psychology, 16. https://doi.org/10.3389/fpsyg.2025.1648480

Chamberlain, R., Mullin, C., Scheerlinck, B., & Wagemans, J. (2018). Putting the art in artificial: Aesthetic responses to computer-generated art. Psychology of Aesthetics, Creativity, and the Arts, 12(2), 177–192. https://doi.org/10.1037/aca0000136

Chen, J., Wang, X., He, Z., & Peng, X. (2025). A Comprehensive Exploration on Detecting Fake Images Generated by Stable Diffusion. In Z. Lin, M.-M. Cheng, R. He, K. Ubul, W. Silamu, H. Zha, J. Zhou, & C.-L. Liu (Eds), Pattern Recognition and Computer Vision (pp. 461–475). Springer Nature. https://doi.org/10.1007/978-981-97-8487-5_32

Högemann, M., Betke, J., & Thomas, O. (2025). What you see is not what you get anymore: A mixed-methods approach on human perception of AI-generated images. Frontiers in Artificial Intelligence, 8. https://doi.org/10.3389/frai.2025.1707336

Hou, W., & Huang, R. (2025). Exploring the Aesthetic Judgments of AI-Generated Digital Abstract Arts. In Proceedings of the 2025 2nd International Conference on Digital Society and Artificial Intelligence (pp. 217–221). Association for Computing Machinery. https://dl.acm.org/doi/10.1145/3748825.3748861

Li, J., Zhang, C., Zhu, W., & Ren, Y. (2025). A Comprehensive Survey of Image Generation Models Based on Deep Learning. Annals of Data Science, 12(1), 141–170. https://doi.org/10.1007/s40745-024-00544-1

Licardo, M., Kranjec, E., Lipovec, A., Dolenc, K., Arcet, B., Flogie, A., Plavčak, D., Ivanuš Grmek, M., Bednjički Rošer, B., Sraka Petek, B., & Laure, M. (2025). Generativna umetna inteligenca v izobraževanju: Analiza stanja v primarnem, sekundarnem in terciarnem izobraževanju (1.). Univerza v Mariboru, Univerzitetna založba. https://doi.org/10.18690/um.pef.1.2025

López-Borrull, A., & Lopezosa, C. (2025). Mapping the Impact of Generative AI on Disinformation: Insights from a Scoping Review. Publications, 13(3), 33. https://doi.org/10.3390/publications13030033

Marzal García-Quismondo, M. Á., & Parra Valero, P. (2021). La educación competencial desde visual literacy y gaming para la innovación educativa: Propuesta para un diseño instruccional de curso. Ibersid: Revista de Sistemas de Información y Documentación, 15(1), 75–83. https://doi.org/10.54886/ibersid.v15i1.4717

Mixer, S. J., McFarland, M. R., & McInnis, L. A. (2008). Visual Literacy in the Online Environment. Nursing Clinics of North America, 43(4), 575–582. https://doi.org/10.1016/j.cnur.2008.06.010

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. Proceedings of the National Academy of Sciences, 119(8), e2120481119. https://doi.org/10.1073/pnas.2120481119

O'Dea, X., & Ng, D. T. K. (2024). AI Literacy and Gen-AI Literacy Frameworks. In X. O'Dea & D. T. K. Ng (Eds), Effective Practices in AI Literacy Education: Case Studies and Reflections (pp. 21–27). Emerald Publishing Limited. https://doi.org/10.1108/978-1-83608-852-320241003

Oksanen, A., Cvetkovic, A., Akin, N., Latikka, R., Bergdahl, J., Chen, Y., & Savela, N. (2023). Artificial intelligence in fine arts: A systematic review of empirical research. Computers in Human Behavior: Artificial Humans, 1(2), 100004. https://doi.org/10.1016/j.chbah.2023.100004

Saha, P., Ghosh, S., & Bhandari, D. (2023). Text-Conditioned Image Synthesis—A Review. Conf. Proc. - IEEE Silchar Subsect. Conf., SILCON. Conference Proceedings - 2023 IEEE Silchar Subsection Conference, SILCON 2023. https://doi.org/10.1109/SILCON59133.2023.10404316

Spalter, A. M., & Van Dam, A. (2008). Digital Visual Literacy. Theory Into Practice, 47(2), 93–101. https://doi.org/10.1080/00405840801992256

van Hees, J., Grootswagers, T., Quek, G. L., & Varlet, M. (2025). Human perception of art in the age of artificial intelligence. Frontiers in Psychology, 15. https://doi.org/10.3389/fpsyg.2024.1497469

Velásquez-Salamanca, D., Martín-Pascual, M. Á., & Andreu-Sánchez, C. (2025). Interpretation of AI-Generated vs. Human-Made Images. Journal of Imaging, 11(7), 227. https://doi.org/10.3390/jimaging11070227

Zhang, C., & Xu, S. (2025). Aesthetic Experience and Educational Value in Co-creating Art with Generative AI: Evidence from a Survey of Young Learners (arXiv:2509.10576). arXiv. https://doi.org/10.48550/arXiv.2509.10576