

Assessing Students' Ability to Identify AI-Generated Images: A Two-Dimensional Study of Realistic and Artistic Content

Hao WU

School of Education, University of Saint Joseph
Macau

ABSTRACT

With the rapid development of artificial intelligence (AI), especially generative AI, there is increasing concern regarding the risk of misinformation from AI-generated content. This study designed a two-dimensional instrument to assess students' ability to identify AI-generated images in both realistic and artistic contexts. The sample consisted of a group of university students (n=112). The instrument included 8 images (4 realistic and 4 artistic), and students were required to identify whether the image was AI-generated. The instrument demonstrated acceptable reliability and good construct validity. The results suggested that students' abilities to identify AI-generated images are two independent constructs but are positively correlated. Additionally, the findings indicated that it is particularly challenging for students to identify AI-generated images in the artistic context. The results highlight the importance of providing training to students in this area, laying a foundation for future studies.

Keywords: AI Literacy; Generative AI; Educational Assessment; Misinformation

1. INTRODUCTION

Nowadays, artificial intelligence (AI), particularly generative AI, has rapidly become one of the most hotly debated technologies worldwide. Generative AI refers to “computational techniques capable of generating seemingly new, meaningful content such as text, images, or audio from training data” [1]. With the widespread success of ChatGPT and other generative AI platforms (e.g., Midjourney, DALL·E, Copilot, etc.), the use of generative AI has become increasingly common in daily life, work, and education, significantly enhancing efficiency and transforming how people interact with the world.

However, alongside the growing popularity of generative AI, several risks have emerged. Concerns have been raised about its potential threats to data security and privacy, as well as its ability to create biased content [2]. Moreover, generative AI can easily contribute to the spread of misinformation by producing artificial content that is difficult to distinguish from authentic sources [3]. Therefore, in this rapidly changing world, AI literacy has become one of the most important skills to focus on in education, preparing students for future challenges. This includes equipping them to understand and mitigate the risks posed by AI, especially generative AI.

One of the important tasks for educators to do is to help our students to develop their ability to distinguish whether content has been generated by AI, in order to avoid being misled or deceived. Although some software and platforms have been

created to identify if content (e.g., text, images) is generated or edited by AI, these tools are not infallible and can make mistakes. For instance, a study conducted by Liang et al. [4] showed that current GPT detectors are biased against non-native English writers. In their study, the selected GPT detectors incorrectly labeled more than half of the essays written by non-native English writers as “AI-generated”. Additionally, such software and platforms are not always easily accessible in everyday life. Thus, it is still necessary for students to develop the ability to recognize AI-generated content to a certain extent.

To reach this goal, students need to have a good understanding of AI and be able to recognize the characteristics of AI-generated content. Although AI-generated content is increasingly similar to human-created content, there are still some differences that can be used to distinguish between the two. For instance, Georgiou's study [5] identified significant differences between AI-generated text and human-written text in many linguistic features, such as “consonants, word stress, nouns, verbs, pronouns, direct objects, prepositional modifiers, and the use of difficult words, among others”. In regard to images, Mathys et al.'s study [6] pointed out that AI-generated images often contain subtle artifacts, such as inconsistencies in lighting, geometry, or human anatomy, which reveal their synthetic nature. In contrast, real images lack these anomalies and appear more natural and consistent. Common mistakes in AI-generated images include distorted human anatomy (e.g., an incorrect number of fingers or unnatural limb positions), inconsistent lighting and shadows, and unrealistic object shapes or sizes. AI-generated images may also exhibit symmetrical inconsistencies, blurred or unnatural textures, and errors in rendering fine details like hair or text [6].

When distinguishing between AI-generated and real images, the approach may differ between realistic and artistic contexts. For realistic images, individuals might focus on identifying specific technical flaws, such as incorrect anatomy, inconsistent lighting, or unnatural textures, which could indicate the image's synthetic origin. However, in artistic images, the evaluation may be more subjective, as artistic content often employs abstraction and creative distortion. This could make it harder to rely on typical markers of authenticity. Instead, viewers might need to look for things like patterns that are too regular or repeated brushstrokes and textures, which could be the hint that the image was generated by AI. These potential differences in detection strategies are worth exploring and could be tested further in this study.

This study will focus on assessing students' ability to identify AI-generated images. The study will be built on a two-dimensional model consisting of Realistic and Artistic contexts, providing insights into how students can further develop this important skill in the current age.

2. RESEARCH AIMS

The aim of this study is threefold: (1) to assess students' ability to identify AI-generated images from both realistic and artistic contexts, (2) to determine whether students' abilities in these two contexts are independent constructs, and, (3) if so, to explore the relationship between these two constructs.

3. METHODS

Participants

The participants of the study were drawn from a university in China through convenience sampling. The sample consisted of a group (n=112) of second-year Bachelor's students. All students participated voluntarily.

Instrument

The instrument was designed by the author and contained 8 items in total. The first 4 items were based on realistic images, and the remaining 4 items were based on artistic images. Students were asked to view each image and decide whether they thought the picture was generated by AI.

In the assessment presented in this paper, 2 AI-generated images and 2 real images were provided in both the realistic and artistic sections. The AI-generated images were created using Midjourney and DALL·E model embedded within ChatGPT 4o. The images in the assessment were carefully selected to include typical unnatural features found in AI-generated images, such as unnatural fingers, unnatural lighting, repeated brushstrokes, and illogical structures. For instance, Figure 1 shows an image from the artistic section: a watercolor-style cityscape with unnatural lighting and an illogical railway design. These features gave students the opportunity to recognize whether the images were generated by AI.



Figure 1. Sample AI-Generated Image Used in the Assessment

The assessment was conducted using the quiz function of the Learning Management System at the sampled university. Students were given 10 minutes to complete the test. If a student correctly identified whether an image was or wasn't generated by AI, the system automatically awarded 1 point; otherwise, a point of 0.

Data analyses

As the instrument was originally developed by the author and had not been used before, both the reliability and validity needed to be tested. Reliability was assessed using Cronbach's alpha, which was run through SPSS, while construct validity was evaluated using confirmatory factor analysis (CFA), run through Mplus [7]. The Tucker–Lewis index (TLI), comparative fit index (CFI),

root-mean-square error of approximation (RMSEA), and standardized root mean square residual (SRMR) were used as model fit indicators. A TLI and CFI greater than .90, an RMSEA less than .08, and an SRMR less than .08 are commonly considered indicators of acceptable model fit [8].

A descriptive analysis was conducted to examine students' performance in the realistic and artistic sections. For the second and third research aims—to determine whether students' abilities in these two contexts are independent constructs, and to explore the relationship between these constructs—these were investigated using the confirmatory factor analysis model.

4. RESULTS

The reliability of the test was acceptable. The internal consistency for the whole test, evaluated by Cronbach's alpha, was .77. The alpha values for the realistic and artistic sections were .81 and .70, respectively.

To assess the construct validity, a two-dimensional model consisting of the realistic (items 1 to 4) and artistic (items 5 to 8) sections was built using confirmatory factor analysis. The model fit indices (TLI=1.00, CFI=1.00, RMSEA= .00, SRMR= .06) indicated that the data fit the two-dimensional model perfectly. Additionally, all factor loadings were strong and statistically significant (see Table 1). Therefore, the test demonstrated good construct validity, confirming that the realistic and artistic sections are independent constructs and should be analyzed separately.

Dimensions	Items	Factor Loadings
Realistic	Item 1	.76**
	Item 2	.84**
	Item 3	.85**
	Item 4	.98**
Artistic	Item 5	.83**
	Item 6	.67**
	Item 7	.78**
	Item 8	.69**

Table 1. Factor Loadings from the Confirmatory Factor Analysis (**p<.01)

Students' performance in the realistic section (accuracy: M=65.18%, SD=38.17%) was higher than in the artistic section (accuracy: M=54.91%, SD=36.12%). However, although the assessment was not intended to make the artistic section more difficult than the realistic section, the difficulty level of the assessment was not standardized. Therefore, while the results suggest that students may have found the artistic section more challenging, there is still a lack of solid evidence to confirm this finding.

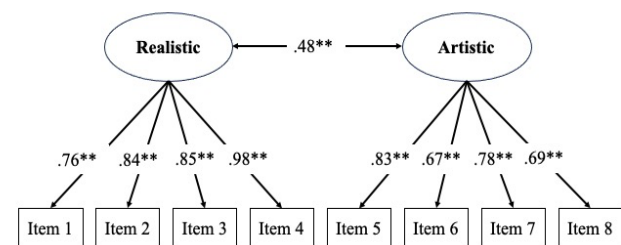


Figure 2. Results of the Confirmatory Factor Analysis Model (**p<.01)

As the realistic and artistic constructs were proven to be independent, it was possible to examine the relationship between these two constructs. We checked the correlation between the latent factors, realistic and artistic, which were created by the confirmatory factor analysis (see Figure 2). The two latent factors showed a moderate yet significant positive correlation ($r=.48$; $p<.01$). In other words, if students can more effectively identify AI-generated images in a realistic context, there is a higher chance they will also be able to identify AI-generated images in an artistic context more effectively, and the reverse is also true.

5. DISCUSSIONS

The presented study used a newly designed instrument to assess students' ability to identify AI-generated images, designed with two dimensions: realistic and artistic, as AI-generated images exhibit different characteristics in these contexts. The newly designed assessment instrument demonstrated acceptable reliability and good validity, allowing us to rely on it to further explore this domain.

Students' performance in the realistic section was higher than in the artistic section. Although all the AI-generated images were deliberately designed to include some typical features, students' accuracy in the artistic section was only slightly higher than what would be expected from random guessing (50%), proving that identifying AI-generated artistic images is indeed a challenging task. The students' average accuracy in the realistic section was higher than in the artistic section. Although, as mentioned, the difficulty level was not standardized, we cannot conclude that students had a better ability to identify AI-generated images in the realistic context. However, it suggests the possibility that students can use their real-world experience to more naturally recognize unnatural features in realistic images. For instance, an abstract artistic human image might not evoke discomfort (e.g., in cartoons or anime), while in a realistic context, it might trigger the 'Uncanny Valley' effect [9]. Although the underlying reasons for this difference in perception are still being explored in cognitive science, this study highlights that the cognitive processes involved in identifying realistic and artistic images are different (as indicated by the CFA results), and there may be differences in the development levels of these abilities in individuals.

Furthermore, the results indicated that the ability to identify AI-generated images in realistic and artistic contexts is positively correlated. This suggests that although the abilities assessed in the two contexts are independent constructs, they still share some common underlying features. This finding implies the potential for educators or researchers to provide explicit training to students, raising their awareness of the risks posed by AI-generated content and helping them learn skills and techniques to effectively identify AI-generated content in both contexts.

6. LIMITATIONS

The limitations of the study include generalizability. The study used convenience sampling, and the sample size is not large enough to make generalized conclusions. Follow-up studies are recommended to further validate these findings. Additionally, the study is based on the current version of generative AI, and due to the rapid advancement of AI technology, the results may be time-sensitive. In other words, the findings might not remain accurate if generative AI undergoes significant improvements or changes in the future.

7. REFERENCES

- [1] Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative ai. *Business & Information Systems Engineering*, 66(1), 111-126.
- [2] Wach, K., Duong, C. D., Ejdys, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., Paliszkievicz, J., & Ziemia, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, 11(2), 7-30.
- [3] Monteith, S., Glenn, T., Geddes, J. R., Whybrow, P. C., Achtyes, E., & Bauer, M. (2024). Artificial intelligence and increasing misinformation. *The British Journal of Psychiatry*, 224(2), 33-35.
- [4] Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, 4(7).
- [5] Georgiou, G. P. (2024). Differentiating between human-written and AI-generated texts using linguistic features automatically extracted from an online computational tool. *arXiv preprint arXiv:2407.03646*.
- [6] Mathys, M., Willi, M., & Meier, R. (2024). Synthetic Photography Detection: A Visual Guidance for Identifying Synthetic Images Created by AI. *arXiv preprint arXiv:2408.06398*.
- [7] Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- [8] Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European journal of developmental psychology*, 9(4), 486-492.
- [9] Hanson, D., Olney, A., Prilliman, S., Mathews, E., Zielke, M., Hammons, D., Fernandez, R., & Stephanou, H. (2005). Upending the uncanny valley. In *AAAI* (Vol. 5, pp. 1728-1729).