Al-Noor Journal for Digital Media Studies
April, 2025; v.1, n.1, en.2 (19-33)
ISSN: 3007-9853, https://doi.org/10.69513/jnfdms.v1.i1.en2

Research Article

Media literacy and mechanisms for verifying AI-generated images

Haneen Saad Salman[a*]

Hanin.saad@tu.edu.iq

ORCID:0000-0002-5913-6266

Farid Saleh Fayyad[b]

faridsf2000@tu.edu.iq

ORCID:0000-0001-6095-9019

[ab]Tikrit University, College of arts, Republic of Iraq

Received: 20/11/2024 Accepted:19/12/2024 Published:28/ 04/2025

Abstract:

In today's world, we are inundated with fake content, particularly with the proliferation of social media, which has contributed to the spread of various forms of deception. Among these, AI-generated images have become increasingly prevalent. Therefore, it has become imperative to activate the role of media literacy to counter the risks posed by AI-generated images.

* Corresponding author
E-mail address: hanin.saad@tu.edu.iq

Media literacy enhances the audience's ability to think critically when consuming media content, enabling them to analyze images more consciously and thus determine their authenticity.

Media literacy also encourages individuals to take responsibility when sharing images and content online, thereby reducing the spread of misinformation. It empowers individuals to distinguish between genuine and fake content circulating on social media. Media literacy focuses on training the users to use tools and techniques to verify images in general and AI-generated images in particular. It raises awareness about the signs that may indicate an image is AI-generated, such as distortions in the image, illogical details… etc.

The study primarily focuses on the risks associated with AI-generated images, such as the spread of misinformation, the creation of ideal beauty standards, biases, and violations of intellectual property rights, among others. The study also emphasizes the mechanisms for verifying these images, employing various methods to determine the origin of the image, such as visual verification, search engine verification, and watermark tracking.

Keywords: fact checking, artificial intelligence, deep fake, news

Media Literacy and Its Importance

Media literacy is defined as the process of empowering individuals in society to criticize and refine fake media products and how to deal with them to achieve knowledge, understanding, awareness, and analysis of media messages, enabling them to critique, make decisions, and solve problems (Feud, 2020).

The importance of media literacy lies in several aspects, most notably protecting the audience from the negative influences of various media and shielding them from it (Eldeeb, 2023). Today, audiences receive a vast amount of data and information, especially through social media platforms, such as texts, videos, and images. Therefore, it has become necessary to have specialized media literacy programs. This diverse content carries the values and orientations of the sender who prepared the message and transmitted it to the audience to achieve specific goals. The risk here does not lie in the quantity or diversity of this content but in its accessibility to all ages, eliminating privacy. Thus, activating media literacy has become essential to raise awareness among children and adolescents, extending to other age groups (Bin Zarrouk & Qarandi, 2021).

Media literacy aims to increase media awareness among the target audience, especially students. Awareness is crucial for students, whose acquisition extends throughout their lives as it is not merely a subject studied for exams and forgotten afterward (Feud, 2020). It is worth noting that media literacy does not require students to reach a professional level in its various skills but rather asks for a minimum level of knowledge to interpret the media content they receive from various media, especially digital ones, which have become an integral part of their daily lives (Abdel Ghani, 2022)

In light of the spread of harmful digital content on social media, it has become necessary to activate the role of media literacy. This type of content may aim to mislead, harm, or deceive an individual or group of users, taking various forms, such as hate speech and misleading content. It has increased in recent years,

especially with the growing number of social media users (Shahi & Tsoplefack, 2022). Other forms of harmful content include electronic games that have contributed to the suicide of a significant number of students, as well as fraud, electronic piracy, viruses, cyberbullying, the spread of pornography, intellectual property crimes, and the potential recruitment of students by extremist groups through websites, especially social media (Elnaghi & Mustafa, 2018).

The Researcher's Perspective on Fake Content and AI-Generated Images

The researcher posits that fake content is at the forefront of harmful digital content proliferating on social media platforms. These platforms have facilitated the dissemination of such content in various forms, including images and fake news crafted to appear as genuine news content, thereby deceiving and misleading users (Salman, 2022).

The Media Literacy Council has identified six types of fake content or news (Sedrati, 2021) as follows:

1. Satirical Content: This content uses humor and sometimes reaches the level of satire, aiming to deceive and influence the audience, although some recipients do not take this type of news seriously.
2. Misleading Content: This type of content includes some facts but interprets them in a way that is far from reality, aligning with the goals and perspectives of the content creators.
3. False Connection: In this type of content, the headline is different from the presented content, aiming to increase audience numbers and, consequently, profits.

4. False Context: The goal of this content is to deceive the audience.
5. Manipulated Content: This content involves altering and modifying details.
6. Fabricated Content: This is entirely false content, which can be extremely dangerous if believed by the audience.

Due to the risks associated with the circulation and dissemination of fake content, many countries have sought to impose laws to limit it. Such content can incite unrest and destabilize internal security. However, imposing laws is not an effective option given the vast amount of fake content published on social media, whether by individuals or through AI algorithms capable of disseminating a massive amount of fake news via fake accounts. Therefore, several researchers have adopted theoretical frameworks based on media literacy, focusing on raising awareness among individuals, fostering critical thinking, and enhancing their mediation role. These theoretical frameworks are based on the assumption that media literacy is the optimal solution to combat fake content (Mekawey, Moayad , & Othman , 2021).

AI-Generated Images and Their Risks

AI-generated images fall under Generative Artificial Intelligence (GAI), defined as a new technology capable of automatically producing new content using input data, including texts, images, videos, audio and code, transitioning from supervised learning to self-supervised learning without human classification (Thabit, 2024).

The risks associated with this type of image are as follows:

Al-Noor Journal for Digital Media Studies
April, 2025; v.1, n.1, en.2 (19-33)
ISSN: 3007-9853, https://doi.org/10.69513/jnfdms.v1.i1.en2

1. Spread of Misinformation:

AI-generated images have contributed to the spread of fake news on the internet. Investigative editor, Conley Adebagbo, suggests that image generators should not be freely available to everyone due to their significant potential to spread misinformation. He believes that making these tools paid would make users more cautious, knowing that money is being spent for this purpose. However, he also acknowledges that many users who deliberately spread misinformation would not mind spending money to access this service. He emphasizes the need for these tools, including Midjourney, to establish specific frameworks to ensure this possibility is minimized (Tijani, 2023).

2. Misleading Consumers About Product Effectiveness:

Many companies have promoted the results of using their products on social media by using AI-generated images. This type of promotion is considered misleading to the consumer audience. One of the most notable examples identified by Misbar, a website for verifying fake news, is campaigns by a page using such images to promote hair oil, claiming that using this oil for 35 days would result in thick and long hair for the girl (Alayobi, 2024).

3. Creating Ideal Beauty Standards

The impact of artificial intelligence on beauty standards has become evident. The Bulimia Project conducted a study on this topic, asking popular image generators, such as DALL-E 2, Midjourney, and Stable Diffusion, to create images depicting beauty and the ideal body for both women and men, as promoted on social media. The study revealed that the standards promoted by AI algorithms are unattainable and their proliferation among social

media users fosters unrealistic expectations. These standards have negative effects on mental health, contributing to issues such as eating disorders and low self-esteem. (Aljazeera net, 2024)

4. Violation of Intellectual Property Rights

AI image generators, including Midjourney, face numerous lawsuits, due to a class action lawsuit filed by a group of artists. They claimed that the company used copyrighted images without the owners' permission to train its generative AI models. This lawsuit highlights the ongoing discussions about intellectual property rights in the age of AI, raising questions about the ethics of using copyrighted materials for learning and training purposes (Ramadan, 2024).

5. Bias and Reinforcement of Stereotypes

AI-generated images continuously face challenges in achieving accurate representation and diversity for all societal groups. Most generated images reflect culturally and geographically specific stereotypes. Reports indicate that these tools depict Africans as "primitive" and show prisoners as black when asked to create such images. This bias stems from the data used to train these tools, which is often sourced from the internet, where much content is biased and toxic. This data includes pornographic content, misogynistic material, and violent and extremist content, directly affecting the results produced by these tools (transparency.news, 2023).

6. Lack of Representation for All Components

One of the challenges facing AI-generated images is the inadequate representation of certain regions of the world. Christoph Schuhmann, a co-founder of the LAION organization,

noted that non-profit organizations providing data do not focus on some areas, such as India and China, leading to distorted perceptions of these cultures (transparency.news, 2023).

Mechanisms for Verifying AI-Generated Images

Verification is defined as the process of attempting to confirm or refute the assertions made in speech, printed media, or online content (Lahmir, 2020). Regarding the mechanisms for verifying AI-generated images, the verifier should follow a series of steps to determine the origin of the image, which are as follows:

Visual Verification

For an information verifier, especially one verifying AI-generated images, it is crucial to conduct a visual inspection by focusing on the following aspects:

A. Inspection of Hands and Feet: In many AI-generated images, hands often appear strange. Sometimes, a person may appear without a thumb, or with six fingers instead of five, and occasionally, fingers may blend together. Although image generation models have advanced to make hands appear more realistic, hands may still look distorted when there are multiple people in the image. Similarly, legs may appear distorted in AI-generated images, with individuals sometimes appearing with one leg or an additional number of legs (Aljazeera net, 2024).

B. Inspection of Hair: Hair remains a challenge for AI-generated images. Human hair typically consists of strands flowing from the head downwards, whereas AI-generated strands often have unclear beginnings and ends and may appear drawn when viewed closely (Aljazeera net, 2024).

Al-Noor Journal for Digital Media Studies
April, 2025; v.1, n.1, en.2 (19-33)
ISSN: 3007-9853, https://doi.org/10.69513/jnfdms.v1.i1.en2

C. Inspection of Ears: Image generation models often fail to accurately depict ears. Therefore, the verifier should examine both ears; if one ear is longer than the other, it may indicate AI generation. Additionally, the shape of the earlobe should be checked, as it may appear distorted on one side (Sky News Arabia, 2024).

D. Inspection of Eyes: The verifier should pay attention to the color of the eyes, as AI-generated images sometimes depict individuals with each eye a different color. The size of the iris should also be checked, as it may appear in different sizes (Sky News Arabia, 2024).

E. Inspection of the Background: A completely blurry background can be suspicious, as experts consider this a sign of inauthenticity. However, if the background is clear, the verifier should examine it in detail. AI-generated images often have distorted backgrounds, and in group photos, the verifier should pay attention to the people in the background, as there may be distortions in their legs or hands (Aljazeera net, 2024).

F. Inspection of Texts: The verifier should examine the texts in images containing street signs or shop fronts. Image generation tools use data from the images they were trained on to produce similar images but are not trained to generate texts. Therefore, the texts in these images often do not make sense and should be scrutinized (Aljazeera net, 2024).

G. Inspection of Details: One of the major shortcomings of image generation models, as of the completion of this study, is their inability to accurately depict details of closely positioned elements. These elements often appear merged, raising suspicion. An

example is the widely circulated image of the Pentagon explosion, which had a clear distortion where a lamp merged with the fence at the bottom, as well as the building merging with the fence (Hamdan, 2023). Another detail that image generators often fail to accurately depict is earrings worn by women, which sometimes appear as a single earring or as two mismatched earrings (Sky News Arabia, 2024).

H.Image Gloss: A significant proportion of AI-generated images have a high degree of gloss, making them appear unrealistically perfect. Individuals in these images often have smooth, flawless skin and shiny, perfect teeth, making many images look almost drawn (Hamdan, 2023).

Second: Verification via Search Engines

Reverse image search is a crucial step for information verifiers. This process helps determine whether the content has been previously published, as well as the time and place of its publication. It is essential to try multiple search engines, as each offers different features that can aid in obtaining the required information, particularly the advanced search options available on most engines (AbdelHaq, Abu Attia, & Al-Masry, 2021).

Third: Verification via AI Tools

Several tech companies have developed specialized tools for verifying AI-generated images. These tools allow users to upload images to determine whether they are real or AI-generated. Examples include AI or Not, JPEGSnoop, and Hugging Face AI Image Detect (Arab Verification Community, 2024). Despite their importance, the researcher found that these tools do not always provide reliable results and require further development by programmers.

Fourth: Monitoring User Comments

Information verifiers should monitor user comments on social media regarding the circulated image. Often, users' comments reveal the truth about the image, sometimes pointing to the creator of the work. Occasionally, the creators themselves may indicate the authenticity of the images without being directly referenced (Hamdan, 2023).

Fifth: Following AI Image Generators

Information verifiers should follow social media groups dedicated to AI-generated images to keep track of the produced works. This practice helps them reach out to artists and enables them to inquire about the authenticity of circulated images. It also keeps them updated on the latest works, some of which may later be associated with false and misleading claims. Notable groups on Facebook include Midjourney Official, AI Art Universe, and AI Generated Art, among others (Hamdan, 2023).

Sixth: Tracking Watermarks on Images

Watermarks on images can help identify their creator. This method assisted the platform "Misbar," which specializes in verifying fake news, in uncovering the truth about an image claimed to be of Julian Assange, the founder of WikiLeaks. By tracking the watermark, the platform identified the image's creator, who confirmed to news sites that he had created the image using Midjourney (Hamdan, 2023).

Conclusion

In light of current technological advancements and the increasing spread of fake news, it has become essential to activate the role of media education to raise awareness about verifying such

news in its various forms, especially AI-generated images. From the above discussion, the study concludes that activating the role of media literacy in society will contribute to achieving the following:

Increasing Public Awareness of AI Image Risks: Media literacy raises awareness of the potential risks associated with AI-generated images, such as their use in spreading misinformation or manipulating public opinion.

Enhancing Critical Analysis Skills: Media literacy helps develop critical analysis skills, enabling the public to evaluate images more accurately and consciously.

Providing Knowledge and Tools: Media literacy equips the public with the knowledge and tools necessary to distinguish between real and fake images. It also teaches the public how to use media responsibly, including verifying the authenticity of images before sharing or relying on them. Additionally, it can play a vital role in supporting the community to become capable of facing these challenges, thereby improving the quality of public discussions and contributing to building an informed society.

Understanding Multiple Verification Methods: Media literacy makes the public aware that verifying AI-generated images relies on multiple methods. These methods include visual inspection of image composition, reverse image search to determine the first publication date, monitoring user comments on the published image for potentially useful details, and utilizing AI image verification tools that may provide valuable results for this process.

References:

1. Abdel Ghani, S. (2022). The Relationship of University Students' exposure to digital media and their knowledge level of the media education concept. *The Egyptian Journal of Media Research*, pp. 1775-1813. doi:https://doi.org/10.21608/ejsc.2022.268258 .

2. Alayobi, J. (2024). *Misleading advertising of goods using AI-generated images*. Available at: https://bit.ly/3X1yJ2U (Accessed 10 September 2024).

3. Aljazeera net. (2024). *How to detect AI generated images?* Available at: https://bit.ly/3X5L8UG (Accessed 11 September 2024).

4. Arab Verification Community. (2024). *Methodology and tools for detecting inauthentic electronic campaigns*. Available at: https://bit.ly/3SQMX5g (Accessed 15 September 2024).

5. Bin Zarrouk, J., & Qarandi, S. (2021). The Role Of Media Education Guide - Lebanon And Tunisia As An Example -. *AL MIAAR, 2*, pp. 444-459. Available at: https://www.asjp.cerist.dz/en/article/144733 (Accessed 15 September 2024).

6. Eldeeb, N. (2023). The Effectiveness of an The Media Education Course to Develop The Cognitive Achievement and The Attitude Towards the Course For Faculty of Specific Education Students Tanta University/Quasi-experimental study. *Journal of Research in the Fields of Specific Education, 45*, pp. 1525-1568. doi: https://doi.org/10.21608/jedu.2023.188898.1825 .

7. Elnaghi, W. M., & Mustafa, H. (2018). The effectiveness of a program to develop digital citizenship values among middle school students in light of media education. *Journal of Media Research*, pp. 601-642. doi: https://doi.org/10.21608/jsb.2018.66977.

8. Feud, I. A. (2020). The impact of the teaching of the media education curriculum on the development of some of the higher thinking skills of a sample of media students in the Academia of Sunrise. *Arab studies in education and psychology, 117*, pp. 191-218. doi: https://doi.org/10.21608/saep.2020.233075 .

9. Hamdan, B. (2023). *How do we detect AI generated images?* Available at: https://bit.ly/3M9Lpzz (Accessed 13 September 2024).

10. Lahmir, N. (2020). Fake news on social media and its effects on public opinion trends: A study of the concept, relationship and objectives. *EL BAHITH FOR ACADEMIC STUDIES, 2*, pp. 579-595. Available at: https://www.asjp.cerist.dz/en/article/124418 (Accessed 10 September 2024).

11. Mekawey, M. A., Moayad , H., & Othman , E. (2021). Mechanisms for Arab youth to circulate fake digital content through social media A proposed model as part of the digital media education approach. *Journal of Mass Communication Research "JMCR", 2*, pp. 527-584. doi: https://doi.org/10.21608/jsb.2021.143195 .

12. Aljazeera.net. (2024). *Will Artificial Intelligence Change the Future of Beauty Standards?* Available at: https://bit.ly/3X52jnT (Accessed 16 September 2024).

13. Ramadan, N. (2024). *AI-powered image generator Midjourney becomes a comprehensive web interface*. Retrieved from Middle East Newspaper. Available at: https://bit.ly/4dAlD3v (Accessed 20 September 2024).

14. Salman, H. S. (2022). The Role of the Lebanese Newspaper An-Nahar in Checking the Fake News: An Analytical Study. *Journal of Tikrit University for Humanities, 6*, pp. 343–362. doi: https://doi.org/10.25130/jtuh.29.6.2022.17 .

15. Sedrati, A. (2021). *Media misinformation through social media during the Covid-19 pandemic: An analytical study on a sample of Facebook pages*. Algeria: University of Larbi Ben M'hidi - Oum El Bouaghi.

16. Shahi, G. K., & Tsoplefack, W. (2022). Mitigating Harmful Content on Social Media Using An Interactive User Interface. *In International Conference on Social Informatics* (pp. 490-505). Cham: Springer International Publishing. doi: https://doi.org/10.1007/978-3-031-19097-1_34.

17. Sky News Arabia. (2024, 6 29). *How to detect fake people's photos and videos with artificial intelligence?* Retrieved from Sky News Arabia: https://2u.pw/lvumOM2Q .

18. Thabit, M. A. (2024). Generative Artificial Intelligence (GAI) and the Future of Scientific Writing: A Survey of Literature Review Tools. *college of Arts Journal, 72*, pp. 73-118. doi: https://doi.org/10.21608/jfabsu.2024.257031.1304.

19. Tijani, A. (2023). *How AI-Generated Images Complicate Efforts to Combat Misinformation*. Available at: https://bit.ly/3Xo3JeR (Accessed 10 September 2024).

20. Transparency.news. (2023). *Analysis: Challenges of Bias in AI Image Generation Tools*. Available at: https://www.transparency.news/news/320706 (Accessed 4 September 2024).