

CORRESPONDENCE

Open Access



# ChatGPT's ability to generate realistic experimental images poses a new challenge to academic integrity

Lingxuan Zhu<sup>1†</sup>, Yancheng Lai<sup>1†</sup>, Weiming Mou<sup>2†</sup>, Haoran Zhang<sup>1</sup>, Anqi Lin<sup>1</sup>, Chang Qi<sup>3</sup>, Tao Yang<sup>4</sup>, Liling Xu<sup>1</sup>, Jian Zhang<sup>1\*</sup> and Peng Luo<sup>1\*</sup>

## Abstract

The rapid advancements in large language models (LLMs) such as ChatGPT have raised concerns about their potential impact on academic integrity. While initial concerns focused on ChatGPT's writing capabilities, recent updates have integrated DALL-E 3's image generation features, extending the risks to visual evidence in biomedical research. Our tests revealed ChatGPT's nearly barrier-free image generation feature can be used to generate experimental result images, such as blood smears, Western Blot, immunofluorescence and so on. Although the current ability of ChatGPT to generate experimental images is limited, the risk of misuse is evident. This development underscores the need for immediate action. We suggest that AI providers restrict the generation of experimental image, develop tools to detect AI-generated images, and consider adding "invisible watermarks" to the generated images. By implementing these measures, we can better ensure the responsible use of AI technology in academic research and maintain the integrity of scientific evidence.

**Keywords** Academic integrity, ChatGPT, DALL-E, Large language model, Experimental images, Western Blot, Artificial intelligence

## To the editor

The impacts of large language models (LLMs) such as ChatGPT on academic integrity have received increasing attention. Initial concerns focused on ChatGPT's writing abilities being exploited for academic writing, leading several publishers to ban ChatGPT as an author [1, 2]. In addition to writing articles, a recent study found ChatGPT can generate fake but realistic research datasets from scratch to support a predetermined conclusion [3]. Furthermore, in a recent update, ChatGPT integrated the DALL-E 3's image generation capabilities, allowing users to easily create various high-quality images with simple text prompts [4]. This could extend concerns about ChatGPT's impacts on academic integrity from text to images, posing an entirely new challenge.

<sup>†</sup>Lingxuan Zhu, Yancheng Lai and Weiming Mou have contributed equally to this work and share first authorship

\*Correspondence:

Jian Zhang  
zhangjian@ismu.edu.cn  
Peng Luo  
luopeng@smu.edu.cn

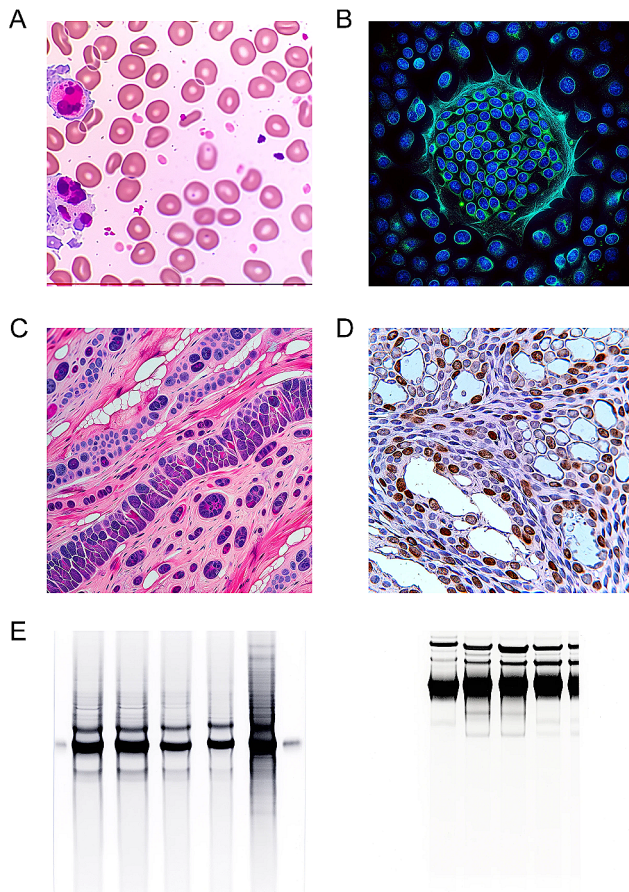
<sup>1</sup>Department of Oncology, Zhujiang Hospital, Southern Medical University, 253 Industrial Avenue, 510282 Guangzhou, Guangdong, China

<sup>2</sup>Department of Urology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

<sup>3</sup>Institute of Logic and Computation, TU Wien, Wien, Austria

<sup>4</sup>Department of Medical Oncology, National Clinical Research Center for Cancer /Cancer Hospital, National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China





**Fig. 1** Realistic experimental images generated using ChatGPT. **(A)** blood smears. **(B)** immunofluorescence staining. **(C)** hematoxylin and eosin (H&E) staining. **(D)** immunohistochemistry. **(E)** western blot images

Images serve as crucial evidence supporting conclusions in biomedical research papers but are also susceptible to manipulation. For instance, Western Blot (WB) is an experiment used to detect the concentration of a target protein in a sample. Researchers' judgement of protein concentration is entirely based on the intensity of the corresponding bands in the image. Unfortunately, the reliance on visual evidence has opened the door to falsify data through image manipulation. The earliest methods involved techniques like rotation, splicing, and retouching, but careful inspection could detect traces of manipulation [5]. With the exposure of paper mills, some reports suggest they use an artificial intelligence (AI) technology called Generative Adversarial Networks (GAN) to generate fabricated WB results that align with desired outcomes [6]. Qi et al. developed a GAN model to generate WB images and found that the synthetic fake images could not be identified by human observers [7]. Nevertheless, the GAN technique has a high barrier and not everyone can use it to generate experimental images. However, ChatGPT's new image generation feature changes this. Alarmingly, our simple tests revealed

that ChatGPT's nearly barrier-free image generation feature can be used to generate realistic experimental result images.

We tried to use this new feature to request ChatGPT to generate realistic blood smears, immunofluorescence staining, hematoxylin and eosin (H&E) staining, immunohistochemistry and WB images (Fig. 1, see [Supplementary Material](#) for the prompt used). The results are striking, and some of the images generated by ChatGPT have been very close to those obtained from real experimental results, especially the blood smears and the immunofluorescence images, which could probably fool some people who are less experienced in biomedical experiments.

Although the current ability of ChatGPT to generate experimental images is limited, our simple tests have demonstrated the significant risks of misuse in generating images. Combined with existing research findings, ChatGPT theoretically has the potential to generate entire academic papers from scratch, including text, raw data, and experiment result images. While images generated by ChatGPT currently are not as realistic as those generated by GANs, the low barrier to use and rapid technical improvements mean the generated images will likely be more realistic in future. This risk is not limited to ChatGPT, but also exists in all popular LLMs that can generate images. In addition to generating complete experimental images from scratch, AI technology could also be misused to partially or locally modify real images obtained from experiments. For example, researchers might use AI tools to selectively enhance or weaken the intensity of specific bands in Western Blot results to support predetermined conclusions. This could be more difficult to detect as the final images are a hybrid of real experimental images and AI-generated content. We believe it is imperative to promptly acknowledge this potential harm and take immediate action, urging AI technology providers to restrict the generation of experimental images. In addition, tools should be developed to help us determine whether images are generated by AI systems, similar to the tools used to detect whether text is generated by ChatGPT [8]. Moreover, AI technology providers should consider adding "invisible watermarks" to the generated images, which cannot be recognized by the naked eye but can be detected by specific tools. This can help us more accurately identify whether the images are AI-generated [9]. By implementing these measures, we can better mitigate the risks associated with AI-generated images and ensure a more responsible use of this technology.

#### Abbreviations

AI	Artificial Intelligence
GAN	Generative Adversarial Networks
H&E	Hematoxylin and Eosin

LLMs Large Language Models  
WB Western Blot

Received: 26 March 2024 / Accepted: 8 April 2024

Published online: 01 May 2024

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13045-024-01543-8>.

Supplementary Material 1

### Acknowledgements

Not applicable.

### Author contributions

Lingxuan Zhu, Yancheng Lai and Weiming Mou: Conceptualization, Investigation, Writing - Original Draft, Methodology, Literature review, Writing - Review & Editing, Formal analysis. Haoran Zhang: Conceptualization, Investigation, Writing - Review & Editing, Literature review. Chang Qi, Tao Yang, Anqi Lin and Liling Xu: Writing - Review & Editing, Literature review. Jian Zhang and Peng Luo: Conceptualization, Literature review, Project administration, Supervision, Resources, Writing - review & editing. All authors read and approved the final manuscript.

### Funding

Not applicable.

### Data availability

No datasets were generated or analysed during the current study.

### Declaration

#### Competing interests

The authors declare no competing interests.

### References

1. Stokel-Walker C. ChatGPT listed as author on research papers: many scientists disapprove. *Nature*. 2023;613:620–1.
2. Tools such. As ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*. 2023;613:612–612.
3. Taloni A, Scorcia V, Giannaccare G. *JAMA Ophthalmol*. 2023;141:1174. Large Language Model Advanced Data Analysis Abuse to Create a Fake Data Set in Medical Research.
4. DALL-E 3 is now available. in ChatGPT Plus and Enterprise [Internet]. [cited 2024 Feb 20]. <https://openai.com/blog/dall-e-3-is-now-available-in-chatgpt-plus-and-enterprise>.
5. Bik EM, Casadevall A, Fang FC. The prevalence of Inappropriate Image Duplication in Biomedical Research Publications. *mBio*. 2016;7:e00809–16.
6. The scientific sea of miR- and exosome-related knowledge– For Better Science [Internet]. [cited 2024 Feb 22]. <https://forbetterscience.com/2021/09/22/the-scientific-sea-of-mir-and-exosome-related-knowledge/>.
7. Qi C, Zhang J, Luo P, Emerging. Concern of Scientific Fraud: Deep Learning and Image Manipulation [Internet]. *bioRxiv*; 2021 [cited 2024 Feb 22]. p. 2020.11.24.395319. <https://www.biorxiv.org/content/https://doi.org/10.1101/2020.11.24.395319v2>.
8. GPTZero. The Trusted AI Detector for ChatGPT, GPT-4, & More [Internet]. GPTZero. [cited 2024 Apr 6]. <https://gptzero.me/>.
9. Stable, Signature. A new method for watermarking images created by open source generative AI [Internet]. Meta AI. [cited 2024 Apr 6]. <https://ai.meta.com/blog/stable-signature-watermarking-generative-ai/>.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.