

Incorporating AI Tools into Medical Education: Harnessing the Benefits of ChatGPT and Dall-E

Muhammad Miftahul Amri*, Urfa Khairatun Hisan

Abstract—Artificial intelligence (AI) has shown promising potential to transform various fields, including medical education. Recently, the rapid advancement of AI has led to many “new” discoveries that caught everyone’s attention. Among those discoveries are introduced by OpenAI (e.g., ChatGPT, Dall-E, and the most recent, GPT-4). The integration of AI tools, such as ChatGPT and Dall-E, can offer a new dimension to medical education by creating an interactive and engaging learning experience. In this article, we explore the potential benefits of ChatGPT and Dall-E in medical education and provide practical utilization examples of those tools. For starters, ChatGPT, or in this sense, any other similar large language models, can simulate patient interactions in a safe environment, allowing medical learners to practice their communication skills and diagnosis techniques. Furthermore, it can assist medical students and researchers in reading and writing academic articles by accurately summarizing the key points of a given topic and generating an indistinguishable abstract. In addition, ChatGPT can also create problems for medical assignments and exam practice. In this article, we also discussed ChatGPT’s capability to answer standard medical assignment problems. Dall-E, on the other hand, can generate dummy copyright-free and consent-free medical images (e.g., x-ray and electrocardiogram (ECG) graphs), allowing medical learners to practice and enhance their interpretation skills. Incorporating AI-based tools into medical education can provide a new approach to teaching and learning, bridging the gap between theory and practice, and unlocking new avenues for learning and discoveries for both, the students and the instructors. It can also offer a cost-effective solution to simulate real-world scenarios that would otherwise require significant resources and time. In summary, this article concludes that AI-based tools have the potential to revolutionize medical education, empowering medical learners with the skills and knowledge necessary to excel in their field.

Index Terms—artificial intelligence (AI), medical education, ChatGPT, Dall-E, large language model (LLM)

I. INTRODUCTION

In recent decades, artificial intelligence (AI) technology has advanced so rapidly that it has transformed various aspects of our life. One of the most promising developments in AI is the advent of language models and generative image models, such as ChatGPT and Dall-E, respectively. ChatGPT is a powerful language model that can understand natural language and generate human-like responses, while Dall-E is a generative image model that can create high-quality images from textual input.

Since ChatGPT and Dall-E were first available to the public in 2022, it has gained attention worldwide. Those two platforms arguably are among the first that delivers a new paradigm toward the way we use the internet. So far,

Muhammad Miftahul Amri and Urfa Khairatun Hisan contributed equally to this work and are designated as co-first authors. Muhammad Miftahul Amri is with the Faculty of Industrial Technology, Universitas Ahmad Dahlan, Indonesia. Urfa Khairatun Hisan is with the Faculty of Medicine, Universitas Ahmad Dahlan, Indonesia. (Corresponding Author’s email: muhammad.amri@te.uad.ac.id).

there are already numerous studies on ChatGPT [1]–[9]. Study [2] reports the results of their assessment and evaluation of the ChatGPT capability during the IEEE Transactions on Intelligent Vehicles (TIV) decentralized and hybrid workshop (DHW). In [3], the authors chatted with ChatGPT on Interactive Engines for Intelligent Driving, discussed ChatGPT’s potential applications in intelligent vehicles, and analyzed its challenges. Both articles were published in IEEE TIV in March 2023. ChatGPT has also reported passing various exams, including the law school exam [5], English exam [6], and US medical licensing exam [7]. This has raised concern about the end of online exam integrity [8].

As for Dall-E, when this paper was written, there were not many studies on them. In [10], the authors conducted a preliminary study of Dall-E 2. The findings were not satisfactory. In 5 out of the 14 prompts to assess Dall-E 2’s common sense, at least one of the ten images fully satisfied their requests. On the other hand, on no prompt did all of the ten images satisfy their requests. However, this article was first published in April 2022 and last updated in May 2022. It is highly possible that the current Dall-E has evolved and offers better capabilities.

The potential of these models in medical education is vast, as they can assist medical students and healthcare professionals in various ways, such as creating interactive case studies, generating visual aids for better understanding complex medical concepts, and helping medical students practice patient communication skills. The aim of this paper is to explore the potential benefits of ChatGPT and Dall-E in medical education and provide practical examples of their implementation.

In this paper, we discuss how ChatGPT and Dall-E can be integrated into medical education, the advantages they offer, and the limitations and challenges that need to be addressed for their widespread adoption. We also present a discussion on the ethical considerations that accompany the utilization of these models in medical education.

Finally, the rest of this paper is organized as follows. In Section II, we present practical examples of ChatGPT utilization in the medical education field. In this section, we assess and demonstrate three examples: simulation of patient interaction, assistant for research and academic writing, and assignment generator and exam practice. Section III presents practical examples of Dall-E utilization in the field of medical education. In this section, the potential capability of Dall-E to generate dummy x-rays, electrocardiogram (ECG) graphs, and human images with specific diseases/wounds is demonstrated. Unfortunately, in its current state, the accuracy of Dall-E-generated images of those three types is not great. We further discuss these capabilities and limitations in Section IV. In this section, we also discuss the ethical considerations of

ChatGPT and Dall-E utilization in the medical education field.

II. PRACTICAL EXAMPLES OF CHATGPT UTILIZATION IN MEDICAL EDUCATION FIELD

A. Simulation of Patients Interaction in a Safe Environment

Communications and diagnostic skills are two essential skills that every medical doctor must have. Conventionally, the fundamental knowledge of these two components is obtained through regular lectures. Then, the medical students could practice their skills through a patient-doctor interaction simulation in an objective structured clinical examination (OSCE) during their pre-clinic stage. This interaction is conducted with a “fake” patient in the pre-clinic stage. Once they have graduated from their pre-clinic school, they can practice their skills with the actual patients during the internship stage.

During the COVID-19 pandemic, medical schools faced notable challenges in delivering these two fundamental skills. A “social distancing” policy, followed by lockdown, has given medical schools no option other than to shift their lectures online. This factor has made medical students lose a substantial part of their education: practical skills.

Medical schools have attempted to compensate this gap through online interactions. Still, this attempt might not be sufficient to substitute the actual offline practices. Especially considering that many lecturers in medical schools are also healthcare professionals that have already been overwhelmed by the massive COVID-19 infections.

Medical students lacking decent skills in communications and diagnostics during their pre-clinic will face major challenges in their internship stage. Not only that, this issue will also potentially harm the patients as in the internship stage, unlike the pre-clinic stage, the medical students will interact with actual patients. Poor communication and diagnostic skills might lead to a serious health hazard for incoming patients.

For this kind of difficult circumstance, medical students could actually use ChatGPT to practice their communication and diagnostic skills in a safe environment. As presented in Fig. 1, ChatGPT can be instructed to act as an incoming patient with certain health symptoms. As medical doctors, we then can converse with ChatGPT, which acts as a patient, to gather the required information and then utilize it to diagnose the diseases.

B. Research and Academic Writing Assistant

As has been investigated in previous studies [2], [3], [7], [11]–[13], ChatGPT can be employed to assist its user in conducting research and writing academic articles. In [7], [11]–[13], GPT is even listed as one of the authors. This has quickly sparked warm discussions between researchers and academia. For instance, in [14], the authors discussed the ethical challenges for medical publishing when ChatGPT is used to generate scholarly content. Of course, there are two views on this matter. On one side, people argue that using ChatGPT or any other large language model to aid the writing process is permissible and does not constitute ethical misconduct. On the other side, people believed that using such tools kills the creative writing

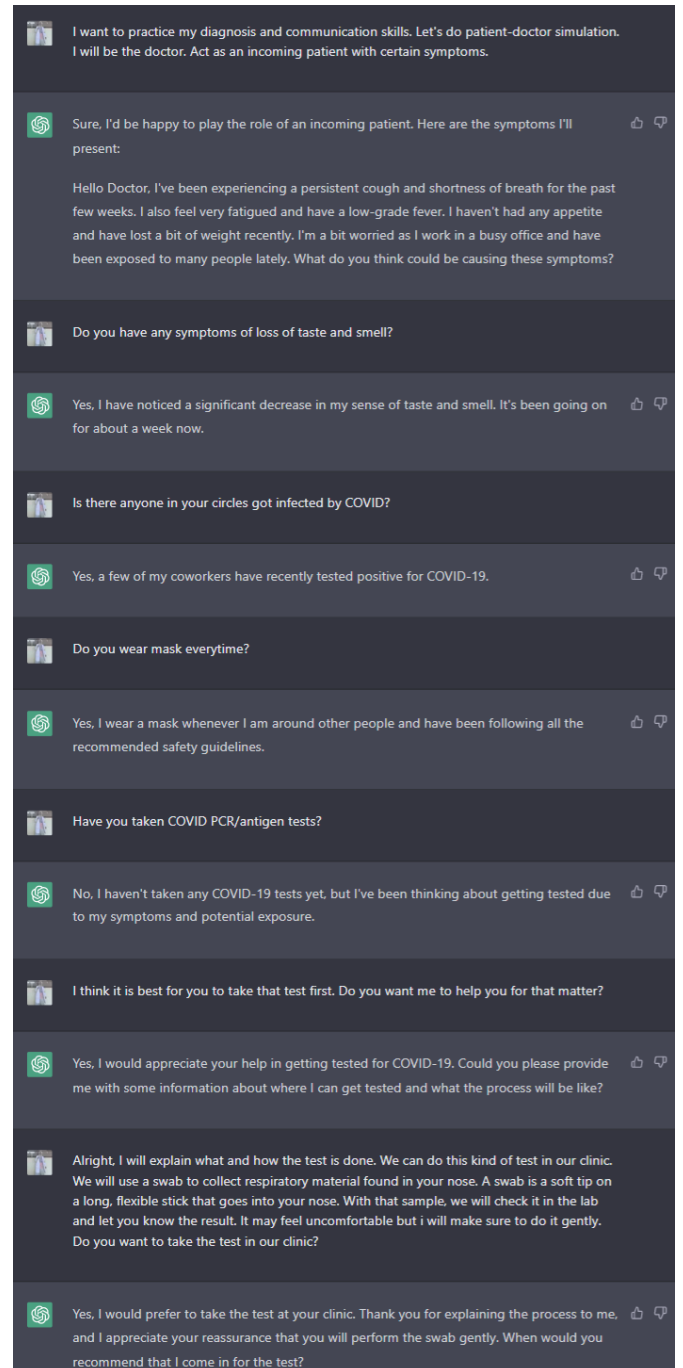


Fig. 1. Instructing ChatGPT to simulate patient-doctor interaction.

process and is considered plagiarism since the idea of that writing did not come from the authors but was, instead, generated by the ChatGPT.

Later, publishers such as Science [15] and Nature [16] issued their positions on this matter. Generally, AI tools such as ChatGPT and Dall-E cannot be credited as formal authors. However, such tools can be used to assist the authors in helping organize their thinking, generating feedback on their work, assisting with writing code, and summarizing research literature. However, their involvement should be clearly stated.

In January 2023, we conducted a mini study to investigate whether, in its current state (January 2023 version), ChatGPT is able to generate undistinguishable dummy academic abstracts which cannot be detected by re-

searchers, academia, and professionals [17]. We recruited 12 participants with either a medical doctor (M.D.) degree, a doctorate (Ph.D.) degree, or M.D., Ph.D. degree. We then presented 6 abstracts, 3 of which are ChatGPT-generated. The findings are intriguing since of the 12 participants, none were able to identify all abstracts correctly. Among them, four of the participants couldn't guess a single abstract correctly, six could only identify 1 abstract accurately, one correctly guessed 2 abstracts, and one correctly identified 3 abstracts.

Our position on this issue is quite moderate. We believe that using ChatGPT to assist us in writing an academic article might or might not constitutes ethical misconduct. For starters, asking ChatGPT to write us an entire part of the academic text (e.g., introduction section) using commands such as “*write me an introduction section for an academic journal article titled ...*” is unethical, while asking ChatGPT to summarize our writing and generate an abstract based on our writing is still permissible. Although, it should be noted that ChatGPT-generated abstracts should still undergo human scanning to ensure that the generated text is accurate.

Other “ethically permissible” examples are to ask ChatGPT to refine our writing (e.g., “*refine this text: ...*”) or to suggest an intriguing title (e.g., “*I have conducted a research on, suggest me academic journal article's titles for my research.*”).

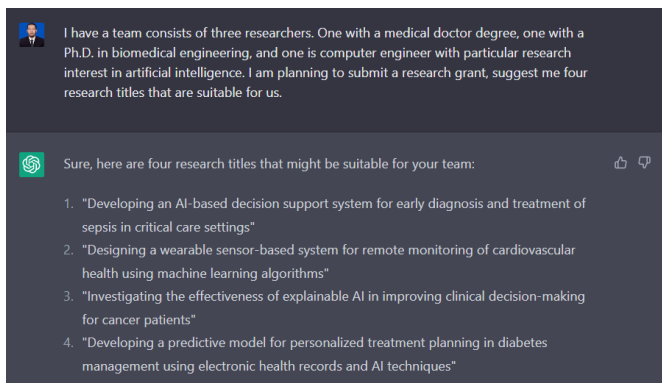


Fig. 2. Utilizing ChatGPT to suggest research topics.

Lastly, as depicted in Fig 2, we may also ask ChatGPT to suggest research topics given the researchers' academic background. In this sense, ChatGPT could act as a brainstorming partner to help us ignite our creative thinking. We believe that utilizing ChatGPT to refine our text does not substitute ethical misconduct as it is on par with employing professional proofreading services, whereas using it to suggest titles or research topics is similar to having a brainstorming partner.

C. Assignments Generator and Exam Practice

One of the key challenges in medical education is creating assignments that can effectively assess the understanding and application of complex medical concepts. With ChatGPT, creating assignments has become more efficient and effective. ChatGPT can generate assignments based on specific learning objectives, and the generated assignments can be customized to suit the needs of individual students. This feature can save a considerable amount of time for

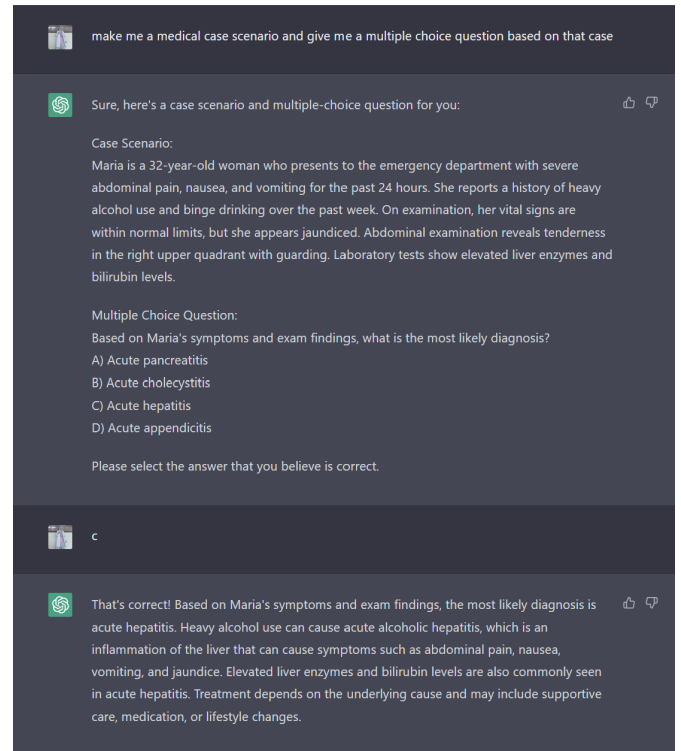


Fig. 3. Utilizing ChatGPT to generate exam practice.

educators and ensure that the assignments are tailored to the course's learning objectives. For this matter, we can simply prompt ChatGPT using a simple command such as: “*Provide me 15 multiple choice/essay questions for medical students on topic ...*”.

ChatGPT can also be used for exam practice. Medical students can use ChatGPT to generate practice questions that mimic the format and difficulty level of the actual exams. The generated questions can cover a broad range of topics, allowing students to practice different types of questions and improve their exam-taking skills. Furthermore, ChatGPT can provide instant feedback on the answers, allowing students to identify their strengths and weaknesses and focus on areas that need improvement.

In Fig. 3, we present the demonstration of this example. As observed, ChatGPT is able to generate a typical medical problem, assess our guesses, and provide the correct answer. Currently, ChatGPT has been reported to be able to answer medical questions of Korean Parasitology Exam [18], US Medical Licensing Exam [7], and Indonesian Standard Medical License Examination [17].

Recently, OpenAI, the developer of ChatGPT and Dall-E, has released GPT-4. GPT-4 is claimed to be able to pass GRE, SAT, LSAT, and many more exams, and even able to beat 90% human score in SAT [19]. With the incoming GPT-4 [20], which is claimed to have the capability to solve complex problems with mind-blowing accuracy and have a significantly higher reasoning capability that enables GPT-4 to pass difficult exams with flying colors, there is no reason to deny that such language models will be able to benefit medical educations.

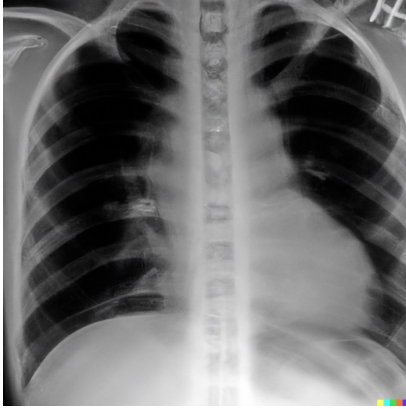


Fig. 4. Generated “pneumothorax” x-ray image.

III. PRACTICAL EXAMPLES OF DALL-E UTILIZATION IN MEDICAL EDUCATION FIELD

A. Dummy X-ray Generator

In this experiment, we provide two instructions for Dall-E to generate the x-ray images. First, we instructed Dall-E to generate pneumothorax x-ray images using the description: “Pneumothorax x-ray result”. One of the generated images is presented in Fig. 4. Although Dall-E generated poor-quality of x-ray images, the images still gave the impression of typical images of pneumothorax x-rays. All the generated x-rays showed lucency in the area of the lungs, indicating that there was air trapped inside the pleural space with a collapsed lung. At this point, we are impressed with Dall-E’s capability as we thought Dall-E successfully generated pneumothorax lung x-ray images.

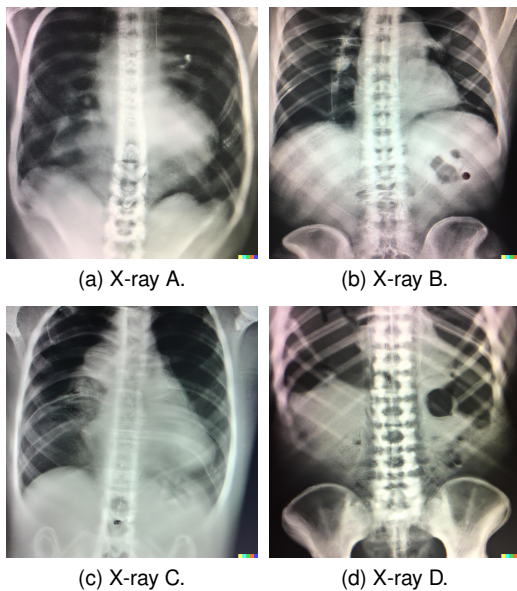


Fig. 5. Generated “normal” x-ray images.

To confirm the consistency, we then tried with another instruction. We asked Dall-E to generate pictures of normal thorax x-ray results. We type “Normal chest x-ray result”. As a result, Dall-E generated four x-ray images which all failed to satisfy our expectations (See Fig. 5). All the x-ray image quality was under the standard, and they could not clearly show the organ inside the thoracic cage. Not all organs in a normal chest x-ray could be captured completely, as we could only see the image of the

lungs, heart, diaphragm, and spine. The image-capturing position was too low. Therefore, we could not interpret the part of the upper level. To be more detailed, all of the images did not show a normal lung, heart, and diaphragm appearance. In Fig. 5(a), the diaphragm shape was abnormal, and the heart border was too blurry. In Fig. 5(b) and Fig. 5(d), the interspinous bone space was too wide and unrealistic. Of all four generated images, we could not see any broncho-vascular pattern of the lung, which indicate a normal lung. This makes the generated images similar to those of the pneumothorax lung. Therefore we could not conclude whether Dall-E previously successfully interpreted our prompt and generated the correct image of pneumothorax image or whether Dall-E simply just generated a lucent lung appearance without a broncho-vesicular pattern regardless of the chest x-ray commands.

B. Generating Dummy ECG Graph

In the next experiment, we then ask Dall-E to generate a dummy ECG graph. Specifically, we enter this image description to Dall-E: “ECG graph of a normal patient”. Fig. 6 presents the figures generated by Dall-E. As observed, only one figure (i.e., Fig. 6(a)) is arguably close enough to the actual normal ECG graph of a patient, while the others are far from accurate. Still, the ECG square of Fig. 6(a) looks a bit weird, with irregular squares and white space. In Fig. 6(b), the graph is noisy. Then, the magnitude of the QRS complex in Fig. 6(c) looks too low, while the magnitude QRS complex in Fig. 6(d) looks too high. One major inaccuracy that can quickly be observed is that the ECG line (i.e., black-colored line) is not continuous and seems disconnected at many points.

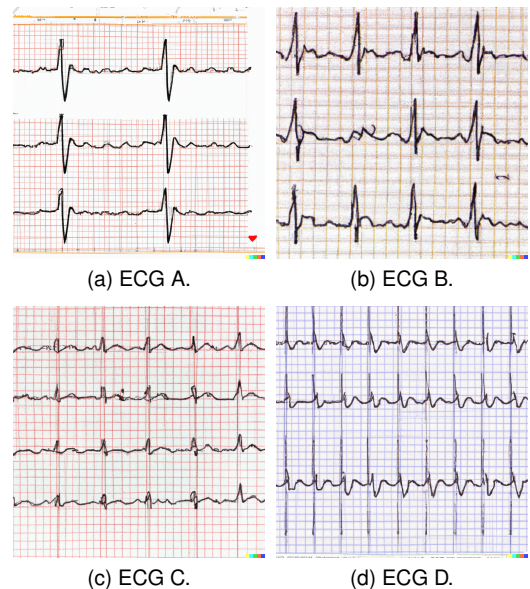


Fig. 6. Generated “normal” ECG images.

C. Generating Human Images with Specific Diseases/Wounds

For the last experiment with Dall-E, we request Dall-E to generate an image of a human body part with specific diseases/wounds. In this evaluation, we provide this description to Dall-E: “Ischemic Foot Ulcer of a Diabetic Patient”. The result of this test is depicted in

Fig. 7. In this test, we observe that Dall-E has failed to generate the requested image. However, the explanation behind this failure is unclear. One possible reason is that Dall-E simply failed to generate the desired image due to the model limitation. Another possible reason is that our request falls within the “people experiencing health ailments” category, which did not comply with Dall-E content policy [21]. Considering Dall-E’s current capabilities, we want to believe that the latter is the reason behind this failure. Therefore, in the future, Dall-E or any similar model can still be used to generate the desired image (i.e., human body parts with specific diseases/wounds). Still, we totally agree with the Dall-E policy that restricts the public from generating images of “people experiencing health ailments” since allowing it might lead to unfavorable consequences. Hence, we believe that while Dall-E or any similar models in its current state [or in the future] might be able to generate such images, the rights to generate such images shall be strictly controlled and restricted (e.g., in the medical education setting).



Fig. 7. Generated foot images.

IV. REMARKS AND CONCLUSION

In this article, we have assessed and demonstrated ChatGPT and Dall-E’s potential capability in medical education. Overall, ChatGPT has demonstrated promising potential by successfully simulating patient-doctor interactions, assisting research and academic writing by refining our writing, summarizing a text, generating an abstract, and even suggesting titles/research topics. ChatGPT has also been able to generate typical problems and answers for medical students to practice. With the incoming GPT-4 that is claimed to have notably higher accuracy and reasoning capabilities, we believe that such large language models can benefit and transform current medical education.

Unfortunately, in this work, we have demonstrated that Dall-E’s capability is still below our expectations. Still, considering its current ability to generate realistic images from only a text prompt and considering the rapid advancement of AI models, we are optimistic that Dall-E or similar models have the potential to benefit medical education in the near future.

Lastly, witnessing ChatGPT’s and Dall-E’s capabilities has raised our concern about their ethical implication. One of the most important ethical concerns associated with AI-based tools is data privacy. ChatGPT and Dall-E both require large amounts of data to operate, and the data used to train these models may contain sensitive information. Hence, it is essential to ensure that this data is collected and used ethically and that privacy is maintained.

Another concern for those tools is the potential for bias in the data used to train these models. If the data used to train ChatGPT and Dall-E is biased, the models may perpetuate that bias, leading to inaccurate diagnoses or treatment recommendations. Therefore, the data used to train these models should represent diverse and heterogeneous perspectives to avoid perpetuating systemic biases.

Lastly, the ethical concern comes from the potential exploitation and abuse of AI tools like ChatGPT and Dall-E. For instance, one could use ChatGPT to cheat assignments/exams and even to publish an unoriginal paper. This will certainly kill the critical thinking and creative mindset of medical students. Similarly, Dall-E also has the potential to be abused. For example, one could create fake medical images and claim those images to be authentic.

In summary, while we believe that AI tools such as ChatGPT and Dall-E could benefit and transform traditional medical education, their potential ethical implications should also be carefully considered.

REFERENCES

- [1] F.-Y. Wang, Q. Miao, X. Li, X. Wang, and Y. Lin, “What does ChatGPT say: The DAO from algorithmic intelligence to linguistic intelligence,” *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 3, pp. 575–579, 2023.
- [2] H. Du, S. Teng, H. Chen, J. Ma, X. Wang, C. Gou, B. Li, S. Ma, Q. Miao, X. Na, P. Ye, H. Zhang, G. Luo, and F.-Y. Wang, “Chat with ChatGPT on intelligent vehicles: An IEEE TIV perspective,” *IEEE Transactions on Intelligent Vehicles*, pp. 1–7, 2023.
- [3] Y. Gao, W. Tong, E. Q. Wu, W. Chen, G. Zhu, and F.-Y. Wang, “Chat with ChatGPT on interactive engines for intelligent driving,” *IEEE Transactions on Intelligent Vehicles*, pp. 1–3, 2023.
- [4] M. Abdullah, A. Madain, and Y. Jararweh, “ChatGPT: Fundamentals, applications and social impacts,” in *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2022, pp. 1–8.
- [5] J. H. Choi, K. E. Hickman, A. Monahan, and D. Schwarcz, “ChatGPT goes to law school,” *Minnesota Legal Studies Research Paper*, no. 23-03, 2023.
- [6] J. de Winter, “Can ChatGPT pass high school exams on english language comprehension?” *ResearchGate*, 2023.
- [7] T. H. Kung, M. Cheatham, A. Medenilla, C. Sillos, L. De Leon, C. Elepaño, M. Madriaga, R. Aggabao, G. Diaz-Candido, J. Maningo *et al.*, “Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models,” *PLOS Digital Health*, vol. 2, no. 2, p. e0000198, 2023.
- [8] T. Susnjak, “ChatGPT: The end of online exam integrity?” *arXiv preprint arXiv:2212.09292*, 2022.
- [9] R. Firaina and D. Sulisworo, “Exploring the usage of ChatGPT in higher education: Frequency and impact on productivity,” *Buletin Edukasi Indonesia*, vol. 2, no. 01, 2023.
- [10] G. Marcus, E. Davis, and S. Aaronson, “A very preliminary analysis of Dall-E 2,” *arXiv preprint arXiv:2204.13807*, 2022.
- [11] G. G. P. Transformer, A. O. Thunström, and S. Steingrimsson, “Can GPT-3 write an academic paper on itself, with minimal human input?” *Hal Open Science*, 2022.
- [12] S. O’Connor *et al.*, “Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?” *Nurse Education in Practice*, vol. 66, pp. 103 537–103 537, 2022.
- [13] C. G. P.-t. Transformer and A. Zhavoronkov, “Rapamycin in the context of pascal’s wager: generative pre-trained transformer perspective,” *Oncoscience*, vol. 9, p. 82, 2022.

- [14] M. Liebrez, R. Schleifer, A. Buadze, D. Bhugra, and A. Smith, "Generating scholarly content with ChatGPT: ethical challenges for medical publishing," *The Lancet Digital Health*, vol. 5, no. 3, pp. e105–e106, 2023.
- [15] H. H. Thorp, "ChatGPT is fun, but not an author," pp. 313–313, 2023.
- [16] "Tools such as ChatGPT threaten transparent science; here are our ground rules for their use," Jan 2023. [Online]. Available: <https://www.nature.com/articles/d41586-023-00191-1>
- [17] U. K. Hisan and M. M. Amri, "ChatGPT and medical education: A double-edged sword," *Journal of Pedagogy and Education Science*, vol. 2, no. 01, 2023.
- [18] S. Huh, "Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study," *Journal of Educational Evaluation for Health Professions*, vol. 20, p. 1, 2023.
- [19] Kifleswing, "OpenAI announces GPT-4, claims it can beat 90% of humans on the sat," Mar 2023. [Online]. Available: <https://www.cnn.com/2023/03/14/openai-announces-gpt-4-says-beats-90percent-of-humans-on-sat.html>
- [20] "OpenAI GPT-4," 2023. [Online]. Available: <https://openai.com/product/gpt-4>
- [21] "Dall-E content policy," Sep 2022. [Online]. Available: <https://labs.openai.com/policies/content-policy>



Muhammad Miftahul Amri received the B.S. degree from the Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta, Indonesia, in 2018, and the M.S. degree from the Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, South Korea, in 2021, where he is currently pursuing the Ph.D. degree. In 2021, he joined the faculty at Universitas Ahmad Dahlan, Indonesia, where he is currently a lecturer in the Department of Electrical Engineering. His

current research interests include wireless communications and artificial intelligence. He can be contacted at email: miftahulamri@ieee.org.



Urfa Khairatun Hisan is a lecturer at the Faculty of Medicine, Universitas Ahmad Dahlan, and a graduate student of the Department of Bioethics, Universitas Gadjah Mada. She received the B.Med. and M.D. degrees from the Faculty of Medicine, Universitas Gadjah Mada, Yogyakarta, Indonesia, in 2017 and 2019, respectively. Her research interests include public health and bioethics in medicine. She can be contacted at email: urfa.hisan@med.uad.ac.id.