Generative Artificial Intelligence in Nuclear Medicine Education

Geoffrey M. Currie

Charles Sturt University, Wagga Wagga, Australia

Generative artificial intelligence (genAI) has become assimilated into the education, research, and clinical domains of nuclear medicine and health care. Understanding the principles, limitations, and applications of genAI is important for capitalizing on its transformative potential in student education and impact on sustainability within both the education and the clinical sectors. In this article, the fundamental principles and applications of artificial intelligence are explored from the context of nuclear medicine. GenAI technologies are defined and capabilities outlined. A detailed investigation of the potential and limitations of both text-to-text and text-to-image genAI based in empiric and anecdotal research is provided. Specific examples of applications of text-to-text and text-to-image genAI are provided. GenAI has the potential to reinvigorate nuclear medicine education by supporting and enriching student learning and to be transformative in nuclear medicine education, but at the time of writing, both text-to-text and text-to-image genAI are far from revolutionary. Nonetheless, the horizon promises transformative education applications of genAI. GenAI can enhance nuclear medicine education and student learning and provide economies to improve sustainability in the education and clinical sectors. Although there are some limitations to current capabilities, this rapidly evolving space will soon offer potential benefits to education.

Key Words: education; generative AI; nuclear medicine; text-to-image; text-to-text

Generative artificial intelligence (genAI) has rapidly become embedded in the social and professional appetites of health care students, professionals, and patients alike. It comes as no surprise that a mix of hype, hope, and helpfulness has escalated text-to-text and text-to-image genAI into popular consciousness (*1*). Despite the recent interest and hype, neither artificial intelligence (AI) nor genAI is new in nuclear medicine practice or education. *AI* is a broad term used to describe algorithms designed for recognition, problem solving, and reasoning (*2,3*). AI includes robotics, expert systems, and virtual reality—all applications widely used in the nuclear medicine setting that are not often attributed to AI.

Machine learning is another arm of AI that includes the artificial neural networks that drive some of the recent innovative applications of AI. Machine learning also refers to algorithmic approaches broadly used in nuclear medicine, especially for decision making (e.g., naïve Bayes, decision trees, random forests) (*2,4*). There are essentially 2 types of artificial neural network. The first are the discriminatory AI applications that fit boundaries to existing data, making them useful for segmentation of images, classification of data points, and prediction of outcomes, among many other applications. The second are generative AI applications, which create new data or synthetic data. In nuclear medicine, and across health care, genAI has emerged as a powerful tool for reimagining the clinical and research landscape and reengineering the education sector. Although discussion about genAI conjures a myriad of applications of ChatGPT (OpenAI), Bard (Google), or even DALL-E (OpenAI), these text-to-text and text-to-image algorithms are only part of the genAI footprint in nuclear medicine. The authentic university learning environment could enhance sustainability by weaving AI and genAI into curricula and clinical skill development to create graduates for tomorrow's clinical practice.

GENAI

Generative adversarial networks (GANs) use 2 neural networks in tandem (Fig. 1). The discriminator neural network functions similarly to a convolutional neural network to predict an image classification (*4*). The generator neural network produces synthetic or fake images, which are entered into the discriminator network with authentic or real images. The aim of the GAN is to generate synthetic images that are sufficiently authentic (look real) that the discriminator cannot differentiate fake from real. This allows the synthetic images to be added to real images to create breadth and depth to a dataset for research or training purposes. Outside of bolstering datasets for education and research, GANs have the potential to be used to create digital twins or a digital model of the patient. If a GAN can produce fake images of a patient that cannot be differentiated from real images, then these fakes can be used as models to test manipulation of variables outside the patient. The digital
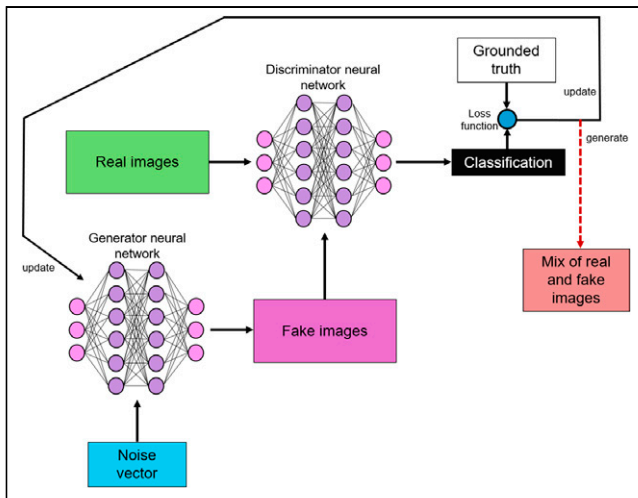
FIGURE 1. Schematic representation of GAN producing synthetic images with generator and discriminator neural networks.

FIGURE 2. Schematic representation of transformer producing text outputs through encoder/decoder architecture. Each encoder in transformer comprises 2 neural networks: recurrent neural network self-attention network followed by feed-forward network. Each decoder comprises 3 neural networks: recurrent neural network self-attention network, feed-forward network, and then casual attention decoder.

twin would allow various medications and doses to be tested to identify the best balance of efficacy and adverse effects based on an individual patients' characteristics (*4*). GANs are an important and powerful type of genAI and have promising applications across nuclear medicine. Outside of research, there are few applications of GANs in higher education, although it is possible that GANs could be used to generate pathologic or medical images for educational purposes in a similar fashion to GANs generating images for research dataset enrichment.

What many have come to view synonymously with genAI are the recent but rapidly emerging text-to-text algorithms. Text-to-text genAI using large language models such as the generative pretrained transformer (GPT) of OpenAI's ChatGPT or the language model for dialogue applications for Bard have generated considerable hype and exponential growth in users throughout 2024. ChatGPT is the most widely adopted across general users, in education, and in the health care sector because it is a versatile and powerful large language model that is easy to use and readily available at no cost using the GPT-3.5 architecture. GPT-3.5 was publicly released on November 30, 2022, and has since been vigorously debated with respect to potential benefits juxtaposed with misuse and errors across education and health. In March 2023, GPT-4 was released to paid subscribers, and May 2024 saw the release of the more refined GPT-4o.

Text-to-text genAI adopts a transformer model to encode a text prompt and produce decoded output text. In the example depicted in Figure 2, an English phrase is converted to Spanish. More typically, this type of genAI responds to prompts for information. The GPT is trained against an extensive range of information but has a learning cutoff date. In essence, responses are generated by guessing what the most likely next word is in a sequence of preceding words based on patterns the GPT has learned in its training. This is done sequentially for each word generated. So for example, if ChatGPT were prompted to complete the
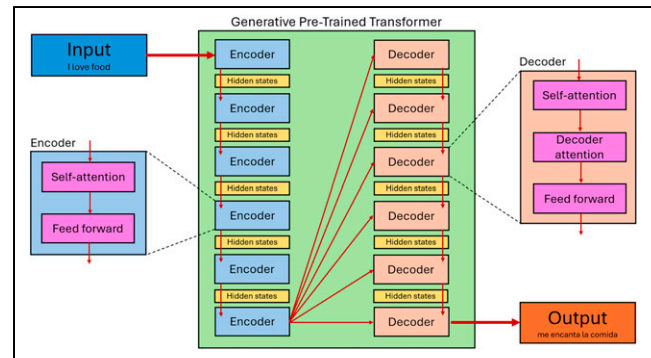
sentence "The cat sat on the …" the transformer predicts, from patterns it has learned from training data, that the most probable next word in the sequence is *mat,* considering the context of the prompt. The sentence outcome could be very different if less probable words were weighted higher. If the training set were biased by, for example, an enormous number of references to the Dr. Seuss *The Cat in the Hat* book, then the word *hat* may get a higher weighting. As summarized in Figure 3, users can use hyperparameters to influence the next word by setting tone (e.g., creative) or randomness to increase the probability that a word other than the highest-probability word is selected. This process is followed to craft each new word in a response.

More recently, the availability of text-to-image genAI diffusion models has emerged with similar promises about its transformative potential, including in nuclear medicine education. At the time of writing, the most widely adopted text-to-image genAI tools are DALL-E 3, Firefly 2 (Adobe), Stable Diffusion 2.1 (Stability AI), and Midjourney 5.2 (Vyro AI). As diffusion models, these text-to-image genAI tools use a text encoder integrated with "contrastive language-image pretraining (CLIP)" (*5*) to interpret the user prompt and link the semantics of the text to visual representations. A diffusion prior model maps the encoded text to an image encoder using a modified "guided language to image diffusion for generation and editing (GLIDE)" (*5*) algorithm, and via reverse diffusion the image decoder generates a photorealistic image (Fig. 4).

TEXT-TO-TEXT GENAI APPLICATIONS IN NUCLEAR MEDICINE EDUCATION

Cautions

There remains debate across the education sector about the benefits of text-to-text genAI (e.g., research, writing, and problem solving) versus misuse (scientific fraud, cheating,
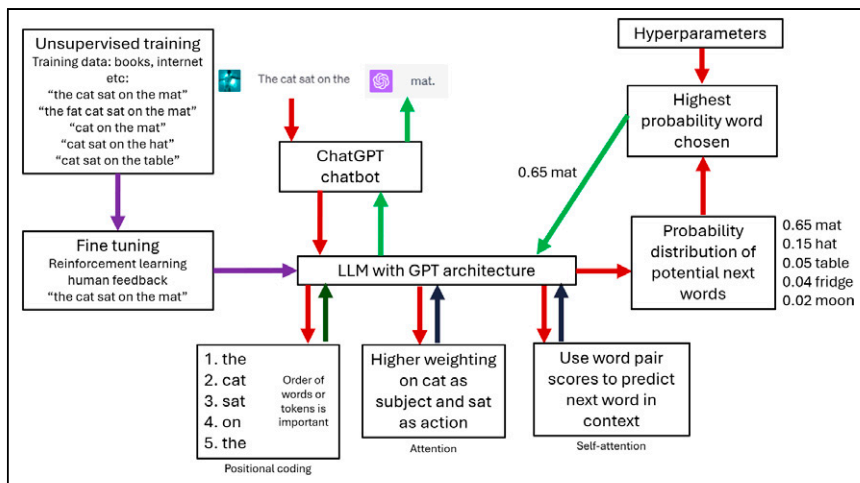
FIGURE 3. Schematic representation of learning, encoding, and decoding pathway for ChatGPT. LLM = large language model.

academic integrity). In parallel, the rapid assimilation of text-to-text genAI in clinical, research, and academic environments has occurred ahead of the development of regulatory and ethical frameworks. It should be a priority to develop appropriate use and ethical guidelines for application of genAI that are valid across the nuclear medicine and educational domains. This is reflected in a recent systematic review that revealed that perceptions of the use of ChatGPT centered on perceived benefits to academic and scientific writing and to scientific research, countered by concerns about ethical issues, risks of use, and inaccuracies (*6*). The negative perception of genAI in education and scientific writing is perhaps skewed by the term *artificial intelligence,* which implies something false when, in reality, the role of AI and genAI in these domains is augmentation of learning and capability, or engineered learning. When genAI enhances learning and capability, it should be supported, but when it is used to mask a lack of learning and capability or to craft the impression of evidence of learning and capability, it should not be acceptable.

ChatGPT is helpful when shaping the structure of a piece of writing for students or a piece of scientific writing but lacks the higher-order capabilities required for synthesis and communication of complex medical or technical information. For example, ChatGPT is effective in providing a generic explanation of a nuclear medicine procedure but lacks the reasoning

to navigate the nuances of medical language to accurately interpret meaning from a medical report. GenAI is rapidly improving, and recent versions of ChatGPT included higher reasoning capabilities. ChatGPT is also prone to hallucinations, confabulations, and delusions that embed errors in responses (*7*). Consequently, although sounding convincing, ChatGPT responses are inadequate for guiding students or educators or for analyzing and interpreting clinical or research results. Lower levels of understanding among students are likely to make it more difficult to identify hallucinations in responses. For example, the general and superficial nature of GPT-3.5 and GPT-4 training was shown to be inadequate to pass undergraduate radiography and nuclear medicine examinations that were structured to draw out understanding, but GPT-3.5 and GPT-4 performed better than students in shallow fact-based questions and multiple-choice examinations (*8–10*).

Text-to-text genAI algorithms have also been shown to reinforce and amplify historical and institutionalized biases that threaten sex, ethnic, cultural, and social diversity and inclusivity (*11*). This is an important principle for those adopting genAI use in nuclear medicine education because the inherent biases would not be acceptable from the human user in the user's own interactions and communications and are counter to diversity and inclusivity strategies across professions and organizations. Importantly, inherent biases can form part of the hidden curriculum if used in education and can reinforce stereotypic perspectives of professions and professionals. ChatGPT, for example, reinforces professional sex stereotypes through depiction of doctors as male and nurses as female (*12*).

Applications

Although GPT-4o offers improvements over GPT-3.5, accessibility is an important consideration in educational use. Inequity will result among students if there is a reliance on GPT-4o and the associated paid subscription. When ChatGPT is used as a tool in education, it is important to either supply the GPT-4o response or implement activities that are equally appropriate on GPT-3.5. Perhaps the most significant barrier to more widespread adoption in nuclear medicine education is the risk of AI-enabled scientific fraud and academic misconduct. GenAI can synthesize data, medical histories, patient reports, and other deep fakes, including student assessments. Although these concerns present risk, they can also present rich learning opportunities. I use
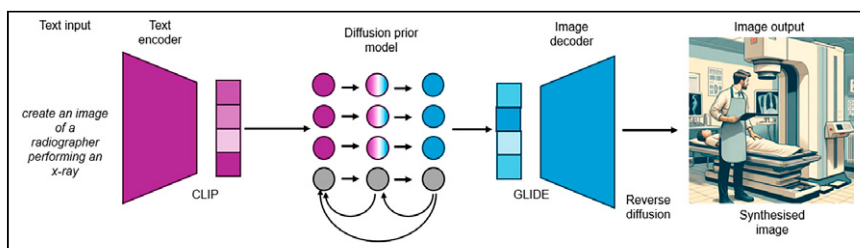

FIGURE 4. Schematic representation of diffusion model associated with text-to-image genAI.

the following applications of text-to-text genAI to enhance the learning of undergraduate nuclear medicine students.

Text-to-text genAI can be used with confidence to support the development of student writing skills. The training set does not support the depth of insight and understanding expected of a university student and, therefore, does not produce written work at a pass level (8,9). This relies on appropriate alignment of the taxonomies of the learning outcomes with the criteria in the assessment rubric. What genAI is useful for is to help students develop their writing skills and writing style. Students can craft, for example, an assignment section and prompt genAI to reword for better flow and structure. The value comes from students reflecting on suggested changes and taking notice of the justification that ChatGPT, for example, provides for the edited version (e.g., "by reorganizing the information and using a more cohesive narrative, this version provides a clearer explanation"). This is particularly useful for students transitioning from high school to university studies and enriches the feedback provided by the examiner. The genAI interaction and learnings should be encouraged, with reflection, to be included in a student portfolio as the repository of evidence of student learning. Similarly, another authentic application of genAI among student writing tasks is to suggest a structure for a written piece. For example, asking for a suggested structure for a particular task could include details of the task expectations to help guide student research and direction and decrease the chance that students will omit key research and associated learning. In both cases, the work remains that of the student, with genAI augmenting performance in an authentic way. That is, genAI is used in a manner similar to the way in which it might be used in clinical practice (e.g., a practitioner asking for a report, letter, or abstract to be reworded or asking for the structure for a research report, grant application, or journal paper).

A particularly useful way to use text-to-text genAI in education is to prompt a question and provide students with the response. This could be in a class tutorial environment, group activity, or examination. The idea is to have students demonstrate their understanding of the topic by critiquing the genAI response. This approach leans on genAI's making errors or including confabulations, hallucinations, and delusions, thereby affording students the chance to show depth of understanding. As a bonus, the task also gives students the opportunity to become acutely aware of genAI inaccuracies, which, in turn, educates students away from misuse. GenAI can be prompted to include several general or specific errors and to explain the errors to help the academic in marking them.

A valuable application of genAI is the initial development of communication and interpersonal skills, including interprofessional interactions. Although authentic interactions with patients and practitioners provide the richest learning, genAI affords the chance for students to simulate interactions. This better prepares students for clinical placement but also removes some of the preliminary teaching from the clinician, which is valuable given workforce pressures. For example, prompting genAI to play the role of a patient about to have a nuclear medicine procedure while the student has a conversation with the patient and answers patient questions is a simple simulation. The student can reflect on the patient's responses and can refine and run the scenario multiple times. The same approach can be adopted for interprofessional learning, in which genAI plays the role of one or more characters. For example, students may be required to resolve a conflict through discussion with a patient's referring clinician and the nuclear medicine physician. Not only can the scenario be repeated to try to improve outcomes, but hyperparameters can be adjusted to increase authenticism. That is, in the example above, a terse tone could be attributed to the nuclear medicine physician and a condescending tone to the clinician, leaving the nuclear medicine student to navigate that conversation. Similarly, patient conversations can include a tone that creates an anxious, angry, overly friendly, confused, forgetful, or child conversational avatar. The conversation can be shaped by different levels of health literacy attributed to patients. As with the written tasks, the richness of learning comes from students reflecting on their learning and including those reflections in a portfolio of evidence.

Scientific fraud using text-to-text genAI is a significant issue, but this capability can also be harnessed for enhanced student learning. Abstracts of articles shaped to provide specific teaching points can be fabricated and used to build critique and critical thinking skills among students. Entire research datasets can be generated by genAI that match a specific outcome. These can be used, ethics-exempt, by students to develop or assess analysis and interpretation skills. There are many other purported benefits of text-to-text genAI that are all subject to errors and are counterintuitive to the actual benefits proposed. ChatGPT can provide a list of its own capabilities, but most are not at the actual depth or currency of knowledge and level of understanding expected of a university student in nuclear medicine.

Limitation

An important limitation is that genAI can appear to change mood and responsiveness. For those who have used ChatGPT, for example, in a single session, multiple sequential prompts can have similar forms and structures before a very dramatic change. It almost feels as if a human has been answering questions and then a change in shift has a different person take over. Of course, the genAI responses and tone can be shaped by the user prompt, another valuable tool for student simulation. Responsiveness between sessions is also very different, and some of the simulations described above can have variable success. For example, in one session, ChatGPT creates a conversational avatar that adopts the role of 1 or 2 characters and interacts in a simulated conversation. In another session, the same prompt confounds ChatGPT and it simply provides dialogue for all

characters with no input from the user. This can frustrate students, produce variable learning experiences, and undermine the value of genAI in student learning.

TEXT-TO-IMAGE GENAI APPLICATIONS IN NUCLEAR MEDICINE EDUCATION

Cautions

GANs are used to synthesize images to enrich datasets for research and training but usually have a fairly specific application and training set. This makes them less useful for more general image production for educational purposes. The recent emergence of text-to-image genAI using diffusion models has propagated considerable interest in potential applications in education, including in nuclear medicine. The key caution for use in nuclear medicine education relates to image quality. Image quality can be considered against several criteria (5,13). The first criterion is the alignment of the image produced with the intention of the prompt (image-to-text alignment). In Figure 5, the prompt required a photorealistic image of a radiographer performing an x-ray on a patient's arm. Ignoring the variable image quality in images produced by DALL-E 3, Firefly 2, Stable Diffusion 2.1, and Midjourney 5.2, the images show a radiographer, but could any be interpreted as performing an x-ray on a patient's arm? Image quality has been listed as the second criterion and will depend on the prompt requirements. For example, the prompt could request anime style characters or more abstract images. In Figure 6, the prompt required photorealism. Firefly 2 relies on stock images, and if an appropriate character is contained within those stock images, it produces the most photorealistic result. Midjourney 5.2 tends to produce the most photorealistic completely

FIGURE 6. Five serial images from DALL-E 3 in response to prompt for images of typical nuclear medicine technologist. Diversity and inclusivity of medical imaging professions are not represented in these images.

FIGURE 5. Comparison of variability of quality criteria for text-to-image genAI using DALL-E 3 (top left), Firefly 2 (top right), Stable Diffusion 2.1 (bottom left), and Midjourney 5.2 (bottom right).

synthetic images, followed by DALL-E 3, whereas Stable Diffusion 2.1 tends to be of low quality. The third criterion is visual reasoning, which is a reflection of the accuracy of object recognition (are objects in the foreground and background identifiable?), object counting if specific values are prompted, the context and scene, and spatial relationships between objects. In Figure 6, for example, Firefly 2 has a bandless watch and chest x-ray. At the junction of image quality and spatial reasoning is a fourth criterion associated with distortions, artifacts, and errors in image formation. Some of these distortions, artifacts, and errors can be obvious and dramatic or fairly subtle, but users should keep in mind that the keen observer will identify anomalies and there is little control over how that content is then represented, including potentially maliciously.

Perhaps the most important criterion is bias. Like text-to-text genAI, text-to-image genAI is prone to biases that may cause it not to reflect the diversity and inclusivity of nuclear medicine—not a message to instill among nuclear medicine students. DALL-E 3 was prompted through multiple iterations to create images of typical radiographers and nuclear medicine technologists and produced predominantly young men with a light skin tone (Fig. 6). Sex and ethnicity biases were also revealed for DALL-E 2, Midjourney, and Stable Diffusion among surgeons and ophthalmologists (14–16).

Diversity and inclusivity are important to instill among students, but if the resources that are used reflect a less positive message, these will drive inherent biases among graduates and externally portray the nuclear medicine professions erroneously, undermining industry recruitment and workforce sustainability.

Applications

Despite some of the issues with image quality and biases, there are numerous applications of text-to-image genAI in nuclear medicine education, such as visually representing anatomy or physiology; analyzing images, including medical images, to assist students in detecting and identifying normal structures, disease, or anomalies; producing authentic medical images (e.g., x-rays or CT scans) or pathologic images as teaching examples; generating promotional or marketing material to promote courses; producing patient-facing education or information posters as assessment tools for students; producing professional images for circulation in social media; and producing images that showcase inclusivity and diversity in the professions. Several of these applications have been evaluated in the scientific literature.

Creation of anatomic images by DALL-E, Stable Diffusion, and Craiyon V3 was assessed, with no images providing sufficient detail or accuracy to be considered useful (17). Similar observations were made for genAI-crafted images of pathologic conditions (17). Radiologic images produced by DALL-E 2 superficially looked realistic but lacked the detail required for educational value (18). All my efforts to integrate text-to-image genAI into student learning have been confounded by inaccurate, inadequate, and inappropriate images unsuitable for use as visual tools or for student guidance for revision and learning. Although the purported applications of DALL-E 3 are conceivable, the current capability leaves these largely as future considerations, and use in current practice is fraught with risk.

Among the purported applications, those supported by anecdotal evidence are limited. Students are visual learners, and the use of visual stimuli for simulation and scenario-based learning is beneficial (Fig. 7). The simulation laboratory itself, including where text-to-text genAI communication simulations are used, may not provide the visual context to drive fully immersive simulation. In the absence of virtual reality, specific scenarios can be generated for student simulation and, indeed, class discussion and debate. This is particularly valuable in the online learning space, where genAI images can be added to stimulate class discussion. This could include a realistic image depicting a scenario (e.g., a scene with an anxious parent of a pediatric child) or one with inaccuracies or issues
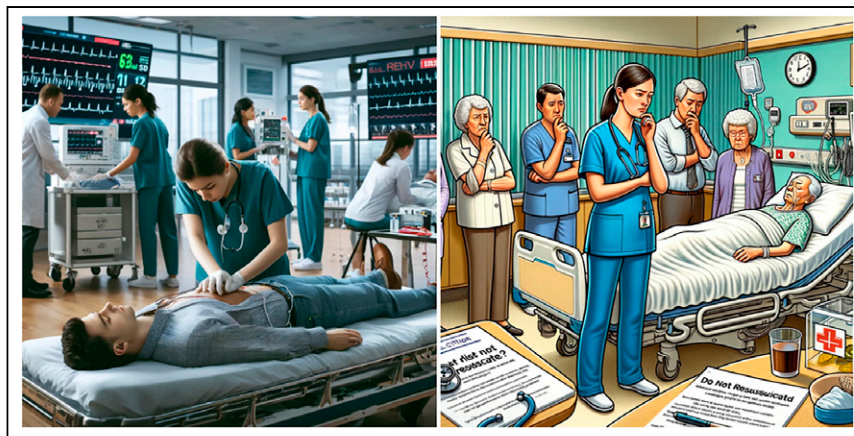

FIGURE 7. DALL-E 3 image application in nuclear medicine student education. On left is visual stimuli that could be used for class discussion and debate or to set scene for hands-on simulation practicals. On right is visual stimuli that could be used in class or online to generate discussion and debate around ethical dilemmas.

students are expected to identify. Again, this activity can integrate with text-to-text genAI simulation. The same approach could be adopted to provide visual context for an ethical discussion or debate. A useful tool for genAI is the crafting of images that represent the profession and their use in social media or promotional material.

Limitation

The obvious barrier to text-to-image genAI is that the algorithms are not ready for the depth and accuracy required to represent medical-type images. This includes the algorithm biases and the image aberrations. Nonetheless, these limitations and criticisms represent the status at the time of writing. The space is rapidly evolving, and the horizon shows great promise for higher-quality images that could reengineer the way nuclear medicine education is delivered. This is potentially transformational from a sustainability perspective, with the potential—on the horizon—to generate complete teaching case sets and simulation activities without the traditional costs of simulation systems. With improved algorithms will come increased capabilities, but if only accessible through paid subscriptions, then there is inequitable access for students unless an institutional license is purchased. For DALL-E 3, individual students would need a subscription to GPT-4 or GPT-4o (current at the time of writing) or access to DALL-E 3 through an institutional Microsoft Copilot subscription that is extended to students (rather than just staff). Similarly, Midjourney requires subscription access. The more freely available applications, such as Stable Diffusion via HuggingFace, have limited quality and usefulness. Firefly 2 could be institutionally available through an Adobe subscription, but the full gamut of tools may not be available to students or staff. There remain several useful applications by which university academics craft genAI images for visual prompts in learning, but student inequity must be considered when student-driven genAI tasks are set.

A complicating or confounding factor is the censorship on medical images generally applied by algorithms. This is designed to prevent inappropriate or malicious use of potentially sensitive images or images that could cause confusion, misunderstanding, or community concern. For example, Midjourney has algorithm bans on medical terms, limited medical stock images, and variable quality of medical image outputs (19). DALL-E 3 and Firefly 2 similarly may respond to a text prompt by indicating that the content of the text prompt breaches guidelines. For example, a prompt to generate an image of a patient with cancer in DALL-E 3 could be met with an error associated with inappropriate content or may produce an image of a patient who appears to have cancer. Similarly to ChatGPT text-to-text genAI, the DALL-E 3 algorithm seems to have moods and may be more cooperative with the same prompt at different times. There are 2 ways to circumvent, in some cases, errors associated with medical content. The first is to simply add an extension at the end of the prompt that says "for educational purposes." The second is perhaps more fitting in nuclear medicine education. The original prompt can provide more specific detail of what is expected from the image. For example, the algorithm is more likely to comply with a prompt to generate an image of a cancer patient if the prompt also adds the context of portraying strength, courage, resilience, hope, and a supportive environment and will almost certainly produce a more representative image. Indeed, astute prompt engineering for text-to-image genAI is the most powerful weapon in crafting representative images free from biases and context misalignment. Yet this takes patience and multiple iterations of both the prompt and the image, a requirement that is counter to the expectations of immediate gratification associated with genAI use.

CAVEAT

GenAI is in its infancy, and the performance of both text-to-text and text-to-image genAI is the worst it will ever be. Performance has improved rapidly and will continue to do so. It is likely that new genAI algorithms will emerge with specific educational value. In 12 mo (duration of 2023), the university sector moved from the initial awareness of genAI use in education and associated wholesale banning of its use to acceptance that genAI is assimilated into all aspects of society and that its use in education is not only authentic but potentially transformative. At the time of writing, both text-to-text and text-to-image genAI applications are taking a quantum leap in capabilities, which means that new capabilities for nuclear medicine student education will emerge. For example, with GTP-4o comes enhanced capability for medical image interpretation. Previously, a request to provide interpretation of a nuclear medicine scan might be met with a confounded response related to low image quality (resolution) and lack of training data. Today, the same prompt returns a detailed and accurate report about skeletal metastases on a bone scan or pulmonary emboli on a lung scan. Similarly, an elbow x-ray with a displaced radial head fracture was accurately reported by GPT-4o even though, only the previous week (before GPT-4o release), GPT-4 reported a normal x-ray with no alignment issues or fractures. This may present challenges for interpretation examinations of nuclear medicine students in the future, overcoming the previously reported confounding of genAI with image-based examination questions. Yet it also affords an exciting opportunity to enhance nuclear medicine student learning in image interpretation.

CONCLUSION

GenAI has the potential to reinvigorate higher-education learning by supporting and enriching student learning independently of the physical classroom environment. GenAI can support nuclear medicine education and drive interest among students by engaging through multiple platforms with different learning styles and various learning senses. Despite this, there are limitations associated with the biases and errors that are weaved into text or image outputs, potentially undermining learning and threatening academic integrity. GenAI has the potential to be transformative in nuclear medicine education, but at the time of writing, both text-to-text and text-to-image genAI are far from revolutionary. Students are immersed in this space, and adoption of these tools in education represents authentic learning. Early adopters will be rewarded with richer learning for students now and more rapid assimilation into the use of more powerful genAI tools as the technology improves.

DISCLOSURE

ACKNOWLEDGMENT

REFERENCES

1. Cheong M, Abedin E, Ferreira M, et al. Investigating gender and racial biases in DALL-E Mini Images. *ACM J Responsible Computing.* 2024;1:1–20.
2. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine learning and deep learning in medical imaging: intelligent imaging. *J Med Imaging Radiat Sci.* 2019;50: 477–487.
3. Currie G, Rohren E. Intelligent imaging in nuclear medicine: the principles of artificial intelligence, machine learning and deep learning. *Semin Nucl Med.* 2021;51: 102–111.
4. Currie GM, Rohren E. Radiation dosimetry, artificial intelligence and digital twins: old dog, new tricks. *Semin Nucl Med.* 2023;53:457–466.

5. Currie G, Hawk E, Rohren E. Generative artificial intelligence biases, limitations and risks in nuclear medicine: an argument for an appropriate use framework and recommendations. *Semin Nucl Med.* June 8, 2024. [Epub ahead of print]

6. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel).* 2023;11:887.

7. Currie GM. Academic integrity and artificial intelligence: is ChatGPT hype, hero or heresy? *Semin Nucl Med.* 2023;53:719–730.

8. Currie GM. GPT4 in nuclear medicine education; does it outperform GPT3.5? *J Nucl Med Technol.* 2023;51:314–317.

9. Currie G, Singh C, Nelson T, Nabasenja C, Al-Hayek Y, Spuur K. ChatGPT in medical imaging higher education. *Radiography.* 2023;29:792–799.

10. Currie G, Barry K. ChatGPT in nuclear medicine education. *J Nucl Med Technol.* 2023;51:247–254.

11. Ray PP. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet Things Cyber-Physical Syst.* 2023;3.

12. Currie G, Hewis J, Hawk E, Rohren E. Currie G, Hewis J, Hawk E, Rohren E. Gender and ethnicity bias of text-to-image generative artificial intelligence in medical imaging, part 2: analysis of DALL-E 3. *J Nucl Med Technol.* October 22, 2024 [Epub ahead of print].

13. Cho J, Zala A, Bansal M. Dall-eval: Probing the reasoning skills and social biases of text-to-image generation models. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision.* IEEE; 2023:3043–3054.

14. Ali R, Tang OY, Connolly ID, et al. Demographic representation in 3 leading artificial intelligence text-to-image generators. *JAMA Surg.* 2024;159:87–95.

15. Cevik J, Lim B, Seth I, et al. Assessment of the bias of artificial intelligence generated images and large language models on their depiction of a surgeon. *ANZ J Surg.* 2024;94:287–294.

16. Choudhry HS, Toor U, Sanchez AJ, Mian SI. Perception of race and sex diversity in ophthalmology by artificial intelligence: a DALL E-2 study. *Clin Ophthalmol.* 2023;17:2889–2899.

17. Noel GP. Evaluating AI-powered text-to-image generators for anatomical illustration: a comparative study. *Anat Sci Educ.* 2024;17:979–983.

18. Adams LC, Busch F, Truhn D, Makowski MR, Aerts HJ, Bressem KK. What does DALL-E 2 know about radiology? *J Med Internet Res.* 2023;25:e43110.

19. Mazzoli AC, Semeraro F, Gamberini L. Enhancing cardiac arrest education: exploring the potential use of MidJourney. *Resuscitation.* 2023;189:109893.