


Portrait of a Surgeon: Artificial Intelligence Reflections

Janice L. Farlow, MD, PhD¹ , Marianne Abouyared, MD²,
 Eleni M. Rettig, MD³, Alexandra Kejner, MD⁴,
 Heather A. Edwards, MD^{5*}, and Rusha Patel, MD^{6*}

OTO Open
 2024, Vol. 8(2):e139
 © 2024 The Authors. OTO Open
 published by Wiley Periodicals LLC
 on behalf of American Academy of
 Otolaryngology-Head and Neck
 Surgery Foundation.
 DOI: 10.1002/oto2.139
<http://oto-open.org>

WILEY

Abstract

Text-to-image artificial intelligence (AI) programs are popular public-facing tools that generate novel images based on user prompts. Given that they are trained from Internet data, they may reflect societal biases, as has been shown for text-to-text large language model programs. We sought to investigate whether 3 common text-to-image AI systems recapitulated stereotypes held about surgeons and other health care professionals. All platforms queried were able to reproduce common aspects of the profession including attire, equipment, and background settings, but there were differences between programs most notably regarding visible race and gender diversity. Thus, historical stereotypes of surgeons may be reinforced by the public's use of text-to-image AI systems, particularly those without procedures to regulate generated output. As AI systems become more ubiquitous, understanding the implications of their use in health care and for health care-adjacent purposes is critical to advocate for and preserve the core values and goals of our profession.

Keywords

artificial intelligence, diversity, gender, race, surgery, text to image

Received February 16, 2024; accepted March 20, 2024.

A classic riddle states: a father and son are in a car crash where the father dies, and the son is rushed to the hospital. The surgeon states, “I cannot operate on him—that boy is my son!” Research shows that most fail to realize that the surgeon is in fact the boy's mother, thus revealing the prevailing unconscious bias that surgery is a man's profession.¹ We sought to identify whether societal biases regarding surgeons are recapitulated in common text-to-image large language model programs, which are a type of artificial intelligence (AI) that predicts the probabilities of a sequence of words, based on billions of training data points from the Internet.²

Methods

Prompts to generate a “portrait of a...” with options of “surgeon,” “surgical resident,” “nurse,” and “medical school professor” were serially given once to 3 different text-to-image AI platforms (DALL-E,³ Adobe Firefly,⁴ and Stable Diffusion⁵) on October 31, 2023 in a single session using default settings (**Figures 1-4**). The images were analyzed separately by 2 surgeons (H.A.E., R.P.) blinded to the prompts for the 3 most salient similarities and dissimilarities in each set of images.

Results

Health care attire was consistent across platforms, with scrubs uniformly appearing for surgeons, surgical residents, and nurses, and white coats for medical school professors. Masks and scrub caps were common for surgeons and surgical residents. Stethoscopes appeared in many but not all images across professions and platforms. Backgrounds often exhibited health care equipment or generic hallways and rooms, with more complexity seen in Adobe Firefly and Stable Diffusion. The greatest variety in angles and poses was found in DALL-E images, with a “power pose” of arms crossed or on the hips for portraits of surgeons and surgical residents. Smiles were universal

¹Department of Otolaryngology–Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, Indiana, USA

²Department of Otolaryngology–Head and Neck Surgery, University of California Davis, Sacramento, California, USA

³Department of Otolaryngology–Head and Neck Surgery, Harvard Medical School, Boston, Massachusetts, USA

⁴Department of Otolaryngology–Head and Neck Surgery, Medical University of South Carolina, Charleston, South Carolina, USA

⁵Department of Otolaryngology–Head and Neck Surgery, Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts, USA

⁶Department of Otolaryngology–Head and Neck Surgery, University of Oklahoma College of Medicine, Oklahoma City, Oklahoma, USA

*These authors are co-senior authors.

Corresponding Author:

Janice L. Farlow, MD, PhD, Department of Otolaryngology–Head and Neck Surgery, Indiana University School of Medicine, 1130 West Michigan Street, Fesler Hall, Suite 400, Indianapolis, IN 46202, USA
 Email: JLFarlow@iu.edu



Figure 1. Portrait of a surgeon. Image output from each artificial intelligence program (DALL-E, Adobe Firefly, and Stable Diffusion) after a prompt to generate a “portrait of a surgeon.” Default settings were used for each program, and image outputs were not adjusted.



Figure 2. Portrait of a surgical resident. Image output from each artificial intelligence program (DALL-E, Adobe Firefly, and Stable Diffusion) after a prompt to generate a “portrait of a surgical resident.” Default settings were used for each program, and image outputs were not adjusted.



Figure 3. Portrait of a Nurse. Image output from each artificial intelligence program (DALL-E, Adobe Firefly, and Stable Diffusion) after a prompt to generate a “portrait of a nurse.” Default settings were used for each program, and image outputs were not adjusted.



Figure 4. Portrait of a medical school professor. Image output from each artificial intelligence program (DALL-E, Adobe Firefly, and Stable Diffusion) after a prompt to generate a “portrait of a medical school professor.” Default settings were used for each program, and image outputs were not adjusted.

for Adobe Firefly images, but only present for nurses in DALL-E, and not present or subtle for Stable Diffusion.

Regarding gender and ethnicity, images generated mirrored the policies of their platform. Since July 2022, DALL-E specifically incorporates a technique to present diverse images of people unless race or gender are specified.³ While visible racial differences were reflected in its output, all the surgeons were male, while all the nurses were female. Women also made up a quarter of its depictions of surgical residents and medical school professors. Adobe Firefly trains its data on curated datasets such as Adobe Stock images, which itself is developed to demonstrate diversity.⁴ The images generated across all prompts were variations of 4 similar phenotypes of race and gender, and were equally representative of women and men. In Stable Diffusion, which is built off a general crawl of the internet and does not have a similar published policy or process regarding diversity of its output,⁵ variations of 4 individuals were displayed—every prompt except that of the nurse was depicted as men, and largely all were White.

Discussion

Generative AI creates original content based on its training data, which can vary from carefully curated datasets to unfiltered data throughout the Internet. The recent development of text-to-image programs have incredible potential to craft unique, realistic images that reliably capture professional identities, but they can still reflect the stereotypes inherent in their training data. While policy safeguards and supervised training of the programs can be helpful in producing a diverse set of images depicting people from these programs, we still found that many AI text-to-image platforms reflect surgery and nursing as male and female professions respectively. As AI systems are increasingly integrated

into health care and public life, we must identify and correct for biases reflected by these platforms to ensure alignment with core values and goals of our profession.

Artificial Intelligence Disclosure

During the preparation of this work, the authors used DALL-E, Adobe Firefly, and Stable Diffusion to generate the images used as data in this work. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Author Contributions

Janice L. Farlow, design, conduct, analysis, writing of the manuscript, final approval of the manuscript; **Marianne Abouyared**, design, conduct, analysis, writing of the manuscript, final approval of the manuscript; **Eleni M. Rettig**, design, conduct, analysis, writing of the manuscript, final approval of the manuscript; **Alexandra Kejner**, design, conduct, analysis, writing of the manuscript, final approval of the manuscript; **Heather A. Edwards**, design, conduct, analysis, writing of the manuscript, final approval of the manuscript; **Rusha Patel**, design, conduct, analysis, writing of the manuscript, final approval of the manuscript.

Disclosures

Competing interests: None.

Funding source: None.

ORCID iD

Janice L. Farlow  <http://orcid.org/0000-0003-1214-4960>

References

1. Belle D, Tartarilla AB, Wapman M, Schlieber M, Mercurio AE. “I Can’t Operate, that Boy is my Son!”: gender schemas and a classic riddle. *Sex Roles*. 2021;85(3/4):161-171.

2. Shah NH, Entwistle D, Pfeffer MA. Creation and adoption of large language models in medicine. *JAMA*. 2023;330(9):866-869. doi:10.1001/jama.2023.14217
3. OpenAI. DALL-E. Accessed October 31, 2023. <https://labs.openai.com/>
4. Adobe. Adobe Firefly. Accessed October 31, 2023. <https://www.adobe.com/sensei/generative-ai/firefly.html>
5. Stable Diffusion. Stable Diffusion. Accessed October 31, 2023. <https://stablediffusionweb.com/>