# Auto-Generative Text-bots in Graduate Medical Education: Resident and Faculty Uses and Opinions

Case H Keltner, Joshua H.M. Sakai, Zachary T Colburn, Luke E Mease

Submitted to: JMIR Formative Research
on: November 19, 2024

# *Table of Contents*

# Auto-Generative Text-bots in Graduate Medical Education: Resident and Faculty Uses and Opinions

Case H Keltner[1] MS, MPH, MD; Joshua H.M. Sakai[2] MS; Zachary T Colburn[2] PhD, MS, MBA; Luke E Mease[2] MD, MPH

[1]Kimbrough Ambulatory Care Center Fort George G Meade US
[2]Madigan Army Medical Center Joint Base Lewis McChord US

**Corresponding Author:**
Case H Keltner MS, MPH, MD
Kimbrough Ambulatory Care Center
2480 Llewellyn Ave, Fort George G Meade, MD 20755
Fort George G Meade
US

## *Abstract*

**Background:** The rise of text-bots such as Chat Generative Pre-Trained Transformer prompts scrutiny of artificial intelligence (AI) in graduate medical education (GME), where data on provider and trainee views on AI's appropriateness and governance remain sparse.

**Objective:** To identify common perceptions and concerns about text-bots among GME trainees and faculty and characterize differences in opinions.

**Methods:** In 2023, we conducted a cross-sectional Qualtrics™-based survey of medical residents and faculty at Madigan Army Medical Center. Multiple choice and free response questions focused on text-bot uses and appropriateness of use in various settings. We completed descriptive analyses and obtained odds ratios (ORs) in Qualtrics™ and R. We compiled common themes from free responses.

**Results:** 43 trainees and 42 faculty responded to at least one question. Relative to faculty, trainees are 69% less likely to report text-bot use for evaluation preparation as appropriate (OR = 0.31, 95% CI: 0.10-0.88), and 2.5-times more likely to consider text-bot use for clinical note writing as appropriate (OR = 3.47, 95% CI: 1.18-10.86). Most trainees and residents agree text-bot utilization is appropriate for administrative tasks, education, learning, and research, with no differences in opinion between the two groups ($p < .05$).

**Conclusions:** Trainees and faculty possess divergent opinions regarding the AI text-bot use for clinical notes and evaluation preparation. Concerns over privacy and oversight countered optimism surrounding efficiency and completion of tedious patient care tasks like letter generation. Additional studies should explore these views, and provider beliefs about medical application(s) of text-bots should inform GME AI policies.

(JMIR Preprints 19/11/2024:69009)
DOI: https://doi.org/10.2196/preprints.69009

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

# Original Manuscript

*Research Letter*

**Title:** Auto-Generative Text-bots in Graduate Medical Education: Resident and Faculty Uses and Opinions

Case H. Keltner, MD, MPH, MS[1,2] (PGY-3 Resident, Preventive Medicine Residency), Zachary T. Colburn, PhD, MBA, MS[2] (Research Scientist), Joshua H. Sakai, MS[2] (Biostatistician), Luke E. Mease, MD, MPH[2] (Program Director, Preventive Medicine Residency)

***Affiliations:***

[1]Department of Public Health, Kimbrough Ambulatory Care Center, Maryland, USA.

[2]Department of Public Health, Madigan Army Medical Center, Washington, USA.

[3]Department of Clinical Investigation, Madigan Army Medical Center, Washington, USA.

**Prior or Related Publications:** None

**Acknowledgements:** We would like to thank Dr. James Chang for providing feedback on the survey before its implementation.

**Disclaimer:** The views expressed in this material are those of the authors, and do not reflect the official policy or position of the U.S. Government, the Department of Defense, the Defense Health Agency, the Department of the Army, or Madigan Army Medical Center. This project was approved by the MAMC Human Research Protections Office, which concluded this study is exempt from the regulatory requirements of 32 CFR 219. Furthermore, a required review by the Defense Health Agency Survey Program Office approved the survey.

**Corresponding Author:** Dr. Luke Mease, Program Director, Public Health Residency, Madigan Army Medical Center, Washington, USA.

Email: luke.e.mease.mil@health.mil

Address: Brigadier General Guthrie L. Turner, Jr. Preventive Medicine Clinical Services Building 9025 5th Street Tacoma, WA 98431

Phone: 253-320-4127

**Keywords:** Artificial intelligence (AI), text-bots, large language models (LLMs), ChatGPT, graduate medical education (GME)

**Conflicts of Interest:** The authors have no conflicts of interest to disclose.

**Data availability:** Data may be available upon request.

**Word Count (Abstract):** 250

**Word Count (Manuscript):** 749 words

**Number of Tables:** 0

**Number of Figures:** 1

**Background:** The rise of text-bots such as Chat Generative Pre-Trained Transformer prompts scrutiny of artificial intelligence (AI) in graduate medical education (GME), where data on provider and trainee views on AI's appropriateness and governance remain sparse.

**Objective:** To identify common perceptions and concerns about text-bots among GME trainees and faculty and characterize differences in opinions.

**Methods:** In 2023, we conducted a cross-sectional Qualtrics™-based survey of medical residents and faculty at Madigan Army Medical Center. Multiple choice and free response questions focused on text-bot uses and appropriateness of use in various settings. We completed descriptive analyses and obtained odds ratios (ORs) in Qualtrics™ and R. We compiled common themes from free responses.

**Results:** 43 trainees and 42 faculty responded to at least one question. Relative to faculty, trainees are 69% less likely to report text-bot use for evaluation preparation as appropriate (OR = 0.31, 95% CI: 0.10-0.88), and 2.5-times more likely to consider text-bot use for clinical note writing as appropriate (OR = 3.47, 95% CI: 1.18-10.86). Most trainees and residents agree text-bot utilization is appropriate for administrative tasks, education, learning, and research, with no differences in opinion between the two groups ($p < .05$).

**Conclusion:** Trainees and faculty possess divergent opinions regarding the AI text-bot use for clinical notes and evaluation preparation. Concerns over privacy and oversight countered optimism surrounding efficiency and completion of tedious patient care tasks like letter generation. Additional studies should explore these views, and provider beliefs about medical application(s) of text-bots should inform GME AI policies.

## Introduction

Integration of artificial intelligence (AI) in medicine continues at breakneck pace, particularly with large language models (LLMs) like ChatGPT. LLMs analyze large datasets and perform complex language tasks, showing promise in fields including medicine for improving efficiency and generating high-quality responses.[1-6]

Despite universal availability to internet users, LLMs remain unregulated in medicine. Nonetheless, LLMs demonstrate superior abilities on medical tasks. Notably, ChatGPT responses to patient queries outperform written physician responses in terms of quality and empathy.[7] Given their ubiquity and utility, LLMs will play a role in GME. However, validating use cases is an ongoing effort.

This work identifies current uses and perceptions of text-bots at a military medical center. Using these data, physicians, GME program directors, and hospital administrators can recognize patterns of concern related to the use, oversight, and application of LLMs in the workspace.

## Methods

A cross-sectional study was conducted at Madigan Army Medical Center (MAMC), which houses 17 Accreditation Council for GME-recognized programs. An interdisciplinary team designed a survey distributed from July 17 to December 13, 2023, via email and QR codes. The Qualtrics™ survey included demographic data, past and current text-bot use, output editing practices, perceptions of appropriate text-bot use, and opinions on AI oversight and ethics. Both multiple-choice and free-response questions were included.

Descriptive statistics were generated within Qualtrics™ and R. Odds ratios compared trainee and faculty opinions on appropriate text-bot use. Free-text responses were analyzed to identify shared perceptions. Responses were anonymized to maintain confidentiality.

## Results

The trainee response rate was 15%. The faculty response rate could not be established; the number of eligible faculty was unknown. 43 trainees (50.6%) and 42 faculty members (49.4%)

responded to at least one survey question. 32.4% (12/37) of trainees believe text-bots would be appropriately used for evaluation preparation in medical settings, compared to 61.1% (22/36) of faculty. Conversely, 54.0% (20/37) of trainees believe clinical note creation using text-bots would be appropriate, versus 25.0% (9/36) of faculty. Most (>50%) trainees and faculty deem text-bot utilization for administrative tasks, education, learning, and research as appropriate.

Relative to faculty, residents are 69% less likely to report text-bot use for evaluation preparation as appropriate (OR = 0.31, 95% CI: 0.10-0.88), and 2.5-times more likely to consider text-bot use for clinical note writing as appropriate (OR = 3.47, 95% CI: 1.18-10.86). With respect to administrative tasks, education, learning, and research, there were no statistically significant differences ($p < .05$) (**Figure 1)**.

Analysis of free responses underscored several themes. Many respondents believe text-bots could improve efficiency, simplify completion of mundane clinical tasks, and facilitate prompt-based clinical note writing. Concerns about privacy and inaccuracy are evident, and some providers feel that only physicians should be allowed to use text-bots in the medical setting. Respondents emphasize the need for double-checking information generated by text-bots, and absence of human oversight could be problematic.

**Discussion**

The development of ChatGPT and other LLMs raises concerns about text-bot use in clinical care, research, and education, but few systematic evaluations of resident and clinical staff sentiments toward text-bots in the medical setting exist. Interestingly, we discovered two key areas of resident-staff disagreement: evaluation preparation and clinical note writing. While over half of residents view clinical note writing as a reasonable text-bot use, only one-quarter of faculty agree. On the other hand, while over three out of five faculty members believe evaluation preparation represents appropriate text-bot use, less than one in three residents agree. Most residents and faculty believe text-bots could be appropriately used for administrative purposes, education (learner and educator

tasks), and research.

Various high-profile institutions have issued AI guidelines for educators and trainees.[8,9] Yet few, if any, have incorporated systematic analysis of educator and trainee views of AI into their policies. Providers and trainees have strong—and divergent—opinions regarding AI in certain capacities. Devising AI policies without educator and trainee input could alienate faculty and/or residents. Further assessment of these views is imperative and should be incorporated into AI best-use guidelines at GME institutions.

**Conclusion**

GME faculty and trainees have mostly favorable and convergent views of AI in the medical setting. While trainees have favorable views toward using text-bots for note writing, faculty do not. The opposite trend holds for evaluation preparation, with faculty, but not trainees, viewing text-bot utilization for this purpose as appropriate. Additional studies at other civilian and military GME locations should investigate these findings further. Hospitals affiliated with GME programs should ensure that faculty and resident sentiments toward text-bots in the medical setting are interrogated prior to publishing AI use policies.

## References


1. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023;29(8):1930-1940. doi:10.1038/s41591-023-02448-8
2. Brown, T. et al. Language models are few-shot learners. In Advances in Neural Information Processing Systems Vol. 33 1877–1901 (Curran Associates, 2020).
3. Shoeybi, M. et al. Megatron-LM: training multi-billion parameter language models using model parallelism. Preprint at arXiv https://doi.org/10.48550/arXiv.1909.08053 (2020).
4. Thoppilan, R. et al. LaMDA: language models for dialog applications. Preprint at arXiv https://doi.org/10.48550/arXiv.2201.08239 (2022).
5. Zeng, A. et al. GLM-130B: an open bilingual pre-trained model. Preprint at arXiv https://doi.org/10.48550/arXiv.2210.02414 (2022).
6. Amatriain, X. Transformer models: an introduction and catalog. Preprint at arXiv https://doi.org/10.48550/arXiv.2302.07730 (2023).
7. Ayers, J. W. et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern. Med. 183, 589–596 (2023).
8. Harvard University. Guidelines for Using ChatGPT and Other Generative AI Tools at Harvard. Available from: https://provost.harvard.edu/guidelines-using-chatgpt-and-other-generative-ai-tools-harvard. Accessed February 25, 2024.
9. Yale University. Yale University AI Guidelines for Staff. Available from: https://yaledata.yale.edu/yale-university-ai-guidelines-staff. Accessed February 25, 2024.

# Supplementary Files

# Figures

Trainees and faculty responded to the question “If you think that text-bots should be allowed in medical settings, which settings would be appropriate? (may select more than one).” Odds ratios represent the odds for trainees relative to faculty.

**Figure 1**. Trainee versus Faculty Views of Text-bot Uses in Medical Settings