JAMA Surgery | Review

Academic Surgery in the Era of Large Language Models
A Review

Timothy A. Rengers, BS; Cornelius A. Thiels, DO, MBA; Hojjat Salehinejad, PhD

IMPORTANCE This review aims to assess the benefits and risks of implementing large language model (LLM) solutions in an academic surgical setting.

OBSERVATIONS The integration of LLMs and artificial intelligence (AI) into surgical practice has generated international attention with the emergence of OpenAI's ChatGPT and Google's Bard. From an administrative standpoint, LLMs have the potential to revolutionize academic practices by reducing administrative burdens and improving efficiency. LLMs have the potential to facilitate surgical research by increasing writing efficiency, building predictive models, and aiding in large dataset analysis. From a clinical standpoint, LLMs can enhance efficiency by triaging patient concerns and generating automated responses. However, challenges exist, such as the need for improved LLM generalization performance, validating content, and addressing ethical concerns. In addition, patient privacy, potential bias in training, and legal responsibility are important considerations that require attention. Research and precautionary measures are necessary to ensure safe and unbiased use of LLMs in surgery.

CONCLUSIONS AND RELEVANCE Although limitations exist, LLMs hold promise for enhancing surgical efficiency while still prioritizing patient care. The authors recommend that the academic surgical community further investigate the potential applications of LLMs while being cautious about potential harms.

JAMA Surg. doi:10.1001/jamasurg.2023.6496
Published online February 14, 2024.

CME at jamacmelookup.com

Author Affiliations: Mayo Clinic Alix School of Medicine, Mayo Clinic, Rochester, Minnesota (Rengers); Division of Hepatobiliary and Pancreas Surgery, Mayo Clinic, Rochester, Minnesota (Thiels); Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, Minnesota (Salehinejad).

Corresponding Author: Cornelius A. Thiels, PhD, Division of Hepatobiliary & Pancreas Surgery, Mayo Clinic, 200 First Street SW, Rochester, MN 55905 (thiels.cornelius@mayo.edu).

After passing all 3 steps of the US Medical Licensing Examination (USMLE),[1] achieving authorship on indexed research articles,[2] and possessing the capabilities to generate a compelling personal statement, ChatGPT is poised to prepare a competitive application for surgical residency. ChatGPT, a large language model (LLM), represents a form of artificial intelligence (AI) built to comprehend and produce humanlike text. Created by OpenAI, ChatGPT has become among the most popular AI resources, amassing over 100 million users in January 2023, only 2 months after its launch in November 2022.[3]

AI is the intersection of science and engineering aiming to build and implement mathematical models capable of reasoning and achieving tasks akin to those of humans. One of the main subfields of AI is machine learning, which uses data, generally at a large scale, without predefined explicit rules, to learn and extract patterns within the data (**Figure 1**). Deep learning is a subfield of machine learning that uses layers of connected artificial neurons inspired by biological neurons to mimic human decision-making process. In comparison with conventional machine learning models, training deep learning models generally requires very large training datasets. This is indeed 1 of the potential barriers of deploying deep learning solutions in many surgical applications, particularly in predictive tasks with respect to rare diseases or outcomes.

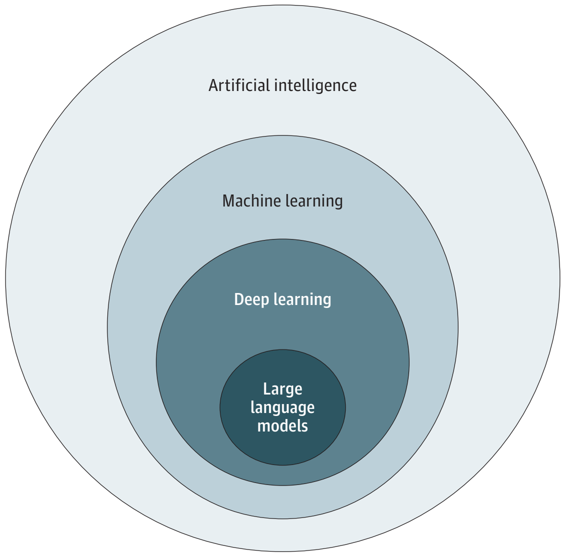
Generative AI refers to the use of deep learning models to generate content in response to a submitted prompt, such as text or image. Recent advances in general generative AI models are mainly due to the availability of data at scale, advances in building and training deep learning models, and high-speed processors. Generative AI models trained on very large textual datasets for generating text or completing a sentence are called LLMs. Most recent LLMs are based on a special type of deep learning architecture called transformer networks. An example is ChatGPT, which is based on the generative pretrained transformer (GPT)[4] architecture.

Hospitals possess large volumes of electronic health records (EHRs), which can be used to fine-tune LLMs, trained on public datasets, for various medical applications. Although LLMs have already impacted the field of medicine, literature pertaining to its use in general surgical practice is sparse. This review seeks to assess the benefits and risks of implementing LLM software in an academic surgical setting.

AI in Medicine

ChatGPT's mainstream debut has generated international conversation about AI, but AI is not a new concept in medicine (**Figure 2**).[5-13] The medical field has been engaged with AI since the 1970s with the formation of the American Association for Artificial Intelligence in 1979 (known as the Association for the Advancement of Artificial Intelligence since 2007).[5,14] Early models, like Mycin, a computer software developed to aid in the diagnosis and treatment of bacterial infections, and Internist-I, a diagnostic program used for consulta-

Figure 1. Associations Between Different Subfields of Artificial Intelligence



tion and educational purposes, exhibited limited performance when dealing with sequential data.[6,7,15] However, it was not until the recent advances in AI that the field started to progress.

The introduction of transformer architecture marked a turning point.[8] Transformer architecture enabled the fusion of language models with computer vision for multimodal tasks. Contrastive language-image pretraining (CLIP), for example, combines the transformer architecture with visual components, allowing it to be trained on text and image data.[16] The multimodal capacity of contemporary LLMs has led to advancements in AI and the exploration of new techniques for medicine. For instance, LLMs have been evaluated in radiology with impressive performance in both identifying important radiographic findings and augmenting the resident training experience.[17,18]

Technical Overview of LLMs

An LLM operates by training on large amounts of text data. This training phase exposes the model to information and patterns of linguistics. A transformer-based model relies on a self-attention mechanism, which allows the model to focus on different parts of an input sequence. Generally, a transformer comprises an encoder and a decoder. The encoder receives input and generates hidden representations, known as embeddings, while the decoder interprets hidden representations to generate output. A multihead attention enables the transformer to place priority on certain words based on relevance. This design can handle long-term relationships between words and improves the performance on various language tasks.[19,20] Depending on the design and application, either the encoder, decoder, or both can be used. For example, Google's bidirectional encoder representations from transformers (BERT) uses only the encoder blocks from the transformer, while OpenAI's GPT relies primarily on the decoder blocks.

The pretraining phase is generally where the LLM learns to predict upcoming words based on context using a large, typically unlabeled textual dataset. During this phase, the model evaluates patterns of linguistics and builds internal representations of the relationship between the words and their meanings. Once pretrained, the LLM can be fine-tuned on a specialized dataset for specific tasks. LLMs can be pretrained on general linguistical datasets and then be finetuned on medical datasets for specialized applications. LLMs continue to advance with medical pretraining and fine-tuning. However, caution must still be taken until further investigation allows for safe integration into clinical practice.

Applications in Academic Surgery

Reduce Administrative Burden

One area of interest is for increased efficiency of administrative tasks. A 2019 study revealed that surgeons derive the most enjoyment from the core tasks of their career while perceiving administrative duties as hindrances to these core tasks.[21] One of the functions that an LLM can perform is called text-to-task where the LLM is trained to perform tasks or actions based on input.[22] Text-to-task and talk-to-text functions may be the next step in reducing a surgeons' administrative burden.

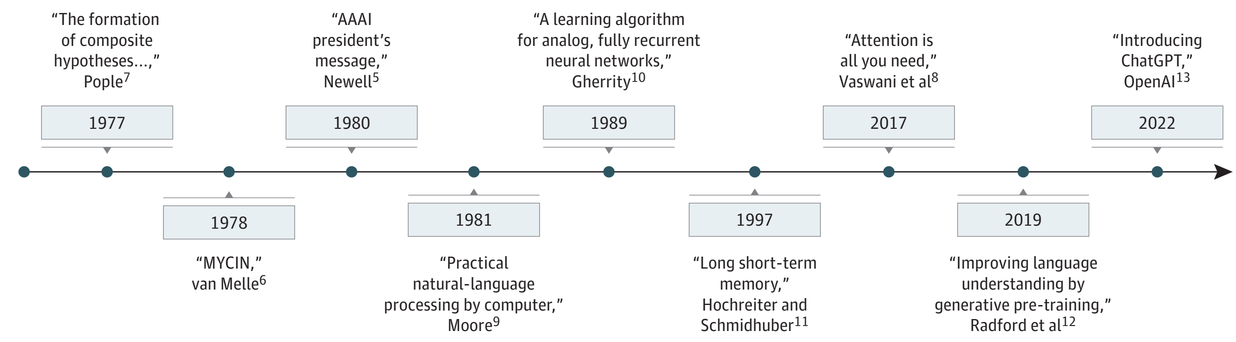
Eloundou et al[23] are developing a rubric to assess the impact on various occupations with LLM exposure. Their preliminary findings suggest that approximately 80% of the US workforce could have at least 10% of their work tasks affected by LLMs. For the academic surgeon, this could be tasks such as an email to address department or team specific issues that can be drafted by an LLM simply by providing agenda, context, and timing. Alternatively, if an email is already written, an LLM can shift tone from passive to active voice or check for grammatical errors. If connected to a calendar system shared by all users, LLMs can be used to schedule meetings, coordinate deadlines, and even send out reminders to team members about upcoming events. Additionally, mentorship is an important element of an academic surgeon's success and letters of recommendation for an aspiring mentee could be another early, practical application of LLM integration.

Early integration of LLM technology has started to appear in health care systems across the US. For instance, Notable, a health care technology company, has released their intelligent automation platform to help reduce administrative burden. They have partnered with North Kansas City Hospital, North Kansas City, Missouri; Medical University of South Carolina, Charleston; and Intermountain Healthcare Salt Lake City, Utah.[24] These health systems have started using the platform for scheduling, registration, intake, referrals, and authorizations. These practices have not yet been widely accepted by the medical field, possibly due to a lack of computational resources, insufficient access to training data at an appropriate scale, and a dearth of trained AI personnel in surgical research labs. Recent advancements have made LLM use within health care more accessible but ongoing efforts and fine-tuning are required before broader implementation.

Medical Education

The landscape of medical education has evolved from the times when students only went to lectures and studied from textbooks. Even before the COVID-19 pandemic, technologic improvements created a shift in the delivery of medical education.[25,26] Now, many students have traded textbooks for third-party resources and spaced-repetition flashcards.[27,28] With the rising popularity of LLMs, medical students and residents now have a new tool to add to their study

Figure 2. Timeline of Advancements in Artificial Intelligence Technology



arsenal as some learners have been asking chatbots to simplify complex concepts for them.[29] Surgeons involved in medical education, whether working with medical students or residents, can use LLMs to assist in generating learning objectives and cases for didactic learning purposes.

Some physicians have been using LLMs to develop an empathetic bedside manner. For example, a department chair recently requested physicians to provide a script demonstrating compassionate delivery of difficult news but received only a few or insufficient responses over several months. However, with the help of ChatGPT, an acceptable script was generated within moments. The reviewers then fine-tuned the response, which is now widely used across the service.[30] Another hospital system developed a clinical scenario for which residents could role play with ChatGPT and have the LLM assess their responses using the Setting up, Perception, Invitation, Knowledge, Emotions With Empathy, and Strategy or Summary (SPIKES) framework for sharing bad news.[31]

Research

The integration of LLMs may also have a considerable impact on surgical research. LLMs can assist in various stages, such as identifying and refining research questions, conducting literature reviews, and designing studies.[32] LLMs are also used in the writing process for tasks like drafting outlines, condensing papers to meet word or character limits, and seeking clarity through chatbot assessments. Researchers have proposed the use of LLMs to generate text for abstracts, manuscripts, or grants. For example, Gao et al[33] explored the use of ChatGPT in generating research abstracts, and reviewers encountered surprising difficulty in distinguishing between abstracts produced by ChatGPT and those from high-impact factor medical journals, solely based on the title and journal.

The use of LLMs for text generation can enhance the efficiency of the writing process. However, their use poses challenges to academic integrity and introduces potential legal issues regarding intellectual property. Authors must be cautious to avoid inadvertently plagiarizing text generated by an LLM. Moreover, there is a growing trend among meetings and journals to broaden their policies regarding the use of LLM-generated text, with AI detection software used on manuscript submissions. Some journals explicitly require authorization for LLM-generated content, and a few request a declaration of generative AI usage during the submission process.

The use of LLMs may become an integral part of analyzing large datasets as well. Lin et al[34] showcase the ability of LLMs to predict detailed atomic-level protein structures from primary sequence. These models may lead to a considerable acceleration in high-resolution structure prediction and have enabled the construction of the ESM Metagenomic Atlas, which provides insights into the extensive range and variety of natural proteins.[35]

Additionally, pharmaceutical company Absci has used generative AI to develop de novo design antibodies using large libraries of data.[36] These strategies could be used to accelerate pharmacologic development and multiomics or to analyze data from the EHR to make health record review and other labor-intensive processes less burdensome. For instance, in a cohort of patients recruited from 2009 to 2010, Jamthikar and colleagues[37] compared statistically derived and machine learning–based calculators of cardiovascular disease and stroke risk. The machine learning–based calculator had higher predictive ability to identify 10-year risk of cardiovascular disease and stroke compared to 13 statistically derived calculators.

Although chatbots are constantly improving, the current range of publicly available LLMs is limited in their training scope, primarily focusing on broad and general information. However, AI developers are actively working on addressing this limitation in medicine by creating LLMs specifically tailored for biomedical applications. Notable LLM platform examples include BioBERT, PubMedBERT, and BioGPT, all of which have undergone extensive training using large-scale biomedical literature. These models have exhibited enhanced performance across various medically related tasks.[38-40] A recent collaboration between the Generative AI platform MosaicML and the Stanford Center for Research on Foundation Models has resulted in the development of a medicine-specific, purpose-built LLM. This model, trained on 2.7 billion parameters, closely approaching the general parameters of ChatGPT (which stands at 3 billion), draws its training data exclusively from PubMed abstracts and full-text articles.[41] By staying updated on these advancements in LLM technology and how other practitioners are using LLMs (**Box**), the academic surgeon can leverage these specialized models to enhance research endeavors and drive progress in the field of medicine.

Clinical Applications

Academic surgeons may benefit from limited use of LLM assistance with patient care as well. Online networking service Doximity

Box. How Academic Surgeons Are Using Large Language Models

Reduce Administrative Burden
- Summarizing large datasets from patient charts for clinical use cases and coding.
- Large language model assisting in drafting work-related emails.
- Shifting tone and checking grammatical errors in emails.
- Scheduling meetings, coordinating deadlines, and sending reminders.
- Assisting with patient registration and prior authorization.
- Drafting letters of recommendations in mentorship.
- Faxing preauthorization and appeal letters to insurers.
- Writing code that can integrate health apps with the electronic health record.
- Summarizing meeting notes and generating meeting minutes.
- Drafting portal messages to patients.

Medical Education
- Simplifying complex concepts to increase learning efficacy.
- Creating learning objectives for learners.
- Generating practice questions and scenarios for didactics.
- Developing script to improve bedside manner when delivering difficult news.
- Role playing with trainees using the Setting up, Perception, Invitation, Knowledge, Emotions With Empathy, and Strategy or Summary (SPIKES) framework to assess bedside manner.
- Detecting missed diagnoses, comorbidities, and other gaps in care.

Research
- Brainstorming to identify research questions.
- Conducting literature review.
- Editing text for clarity or condensing word count to meet character limits.
- Generating outlines, titles, abstracts, and cover letters.
- Processing large datasets efficiently and accurately.
- Creating predictive model from large datasets.
- Connecting patients to clinical trials.
- Checking for grammatical errors or shift tone from passive to active voice.

has already launched the use of ChatGPT with a HIPAA-compliant writing assistant called DocsGPT.[42] Templates for progress and procedure notes, letters to insurance, patient education material, or letters for medical leave of absence can be quickly generated in the Health Insurance Portability and Accountability Act secured interface and copied into the EHR. Digital health company Wefight compared the responses of a chatbot to physician responses pertaining to breast cancer care.[43] The results found noninferiority with success rates of 69% in the chatbot group vs 64% in the physician group. Another study assessed the ability of ChatGPT to produce clinical letters. Ali et al[44] developed 38 hypothetical scenarios pertaining to skin cancer findings and found that ChatGPT could generate clinical letters with high levels of correctness and humanness. However, a recent study[45] investigating ChatGPT's responses to frequently asked questions about breast augmentation found the model generated grammatically accurate, comprehensive responses but with inappropriate or outdated references.

Ethical Concern and Limitations

The implementation of LLMs into academic surgery has limits and requires consideration of several ethical concerns. Potential issues with LLMs include privacy implications, informed consent, patient safety issues, risk of algorithmic bias, and academic dishonesty.

Data Privacy

Providing LLM-based tools access to an EHR comes with concerns about patient privacy and data use. A systematic review of information breach within health care facilities was conducted by Khanijahani et al.[46] One of their findings was that both higher rates of EHR use and earlier phases of EHR adoption were associated with higher rates and levels of information breach. Early transition into LLM-based applications in surgery may put patients at risk of information breach for similar reasons. Use of LLMs and AI must include safety precautions and novel technologies to ensure proper data handling and processing.

Another risk associated with the use of LLMs for academic surgery is breach of surgeon privacy. Current LLMs house data for storage that can be accessed by platform employees with authorization. ChatGPT's frequently asked questions state that "conversations may be reviewed by our AI trainers to improve our systems."[47] Surgeons who intend to use public platforms, such as Bard and ChatGPT, should be informed about how their data are being housed and accessed. The storage format of these LLMs means that research ideas or trade secrets pertaining to their career are vulnerable to breach as well.

Informed Consent

Informed consent is the process of explaining the details of treatment in patient care or data and tissue use in research. In clinical settings, health care professionals educate patients about the risks, benefits, and alternative options related to the intervention being considered. When obtaining consent for research purposes, a representative from the laboratory or a health care professional must disclose project details and how the patient's data will be used. This process ensures patients are able to make informed and educated decisions regarding their participation in treatment or research.

When an academic surgeon intends to use patient data through AI, it is crucial to provide proper informed consent specifically regarding the use of AI devices. However, unlike the European Union's General Data Protection Regulation, the US lacks a comprehensive federal data privacy law. The General Data Protection Regulation recognizes a patient's right to withdraw informed consent from a study, but integrating this concept with AI algorithms can present challenges.[48] Additionally, without an equivalent to the General Data Protection Regulation in the US, there are no established regulations safeguarding individual privacy rights in terms of how hospitals acquire, store, and use patient data. While legislative bills have been proposed, further action is necessary before LLM integration can expand access to patient data.

Patient Safety

LLM-assisted patient education holds the potential to enhance access to suitable health care for patients in remote locations or with limited financial resources who are considering the need for in-person evaluations. An internet search may generate misinformation and lead patients into support forums with peer-to-peer medical advice instead of information generated by health care professionals. For instance, a qualitative descriptive analysis[49] was performed to examine the accuracy of responses pertaining to heart failure in online health forums. The forum provided an excellent source of support between users, but of the approximately 300 diagnostic responses that were queried, only 5% were guideline and evidence based.

The current state of chatbots comes with a set of risks as well. The US Food and Drug Administration (FDA) has yet to review and approve any LLM (including ChatGPT and Bard). The absence of FDA reviews for LLMs raises concerns regarding their use in the medical field. If these systems are intended for medical purposes and do not fall under FDA exceptions, they could trigger regulatory oversight. While ChatGPT generates differential diagnoses and advises users to consult medical professionals, ethical questions arise when medical professionals use the system themselves. OpenAI's terms caution users about the potential for incorrect, biased, or hallucinated information from ChatGPT, posing liability risks even for health care experts. Liability may depend on assessment of use and whether a physician's reliance on AI-generated data breaches guideline-established standard of care. The interpretation of this standard in cases involving LLMs will determine potential physician liability, with increasing acceptance and FDA authorization potentially affecting the outcome.[50]

Tung et al[51] recently conducted an assessment of ChatGPT's perioperative patient communication. Although the chatbot was able to respond with empathy, reassurance, and accuracy in discerning minor vs major postoperative complications, some of the information presented was incorrect but presented with the authority of factual truth. Despite the chatbot's statements about consulting a licensed professional for further inquiries, concerns regarding patient safety arise when using LLMs for educational purposes. Additionally, when the chatbot was prompted with a nonspecific concern, ChatGPT was unable to generate follow-up questions to help narrow the differential. Until medically specific LLMs are investigated further and become more widely available, the academic surgeon should continue to caution patients on the limitations of turning to LLMs for medical information.

Algorithmic Bias

Another concern around integration of LLMs into surgery is the risk of bias in training the models. An example can come from medical education and the resources used to study and evaluate learners. A 2022 study analyzed the resources used to prepare for the *USMLE Step 2 Clinical Knowledge* examination to see if and how diseases may be racialized.[52] The analysis demonstrated patterns of race-based disease associations with potential for bias, promoting false associations, and upholding cultural conventions as normative. Additionally, LLMs run the risk of generating AI hallucinations, which are similar to neuropsychiatric confabulations, where the output text is delivered with confidence but may be an inaccurate synthesis leading to false information. Training of LLMs must include deep layers in the artificial neural network to ensure consideration of risk-factors pertaining to race, gender, ethnicity, and immigration status to include nuance and the ability to dissociate risk from a biological basis of disease.

Academic Dishonesty

Although LLMs provide valuable resources for students and researchers, they also bring about ethical concerns that are inherent to these new technologies. One of the major risks is academic dishonesty, which poses a threat to the integrity of education and research as a whole. It is most important to note that the National Institutes of Health (NIH) has announced the prohibition of generative AI to analyze and formulate critiques of grant proposals.[53] The NIH has expressed concern about the security of intellectual property as many AI tools have no guarantees regarding where data are being stored and whether they are vulnerable to confidentiality breech. Academic surgeons had best use caution when using LLM's when submitting for grants to the NIH or any other funding body, both from an academic dishonesty and data privacy perspective. With the introduction of LLMs, there is an increased potential for plagiarism in grants or assignments, cheating on examinations, and an overall power imbalance that disproportionately benefits individuals in privileged positions who have access to the most up to date LLMs. Brynjolfsson[54] warns of the Turing trap, wherein humans excessively rely on LLMs, leading to job displacement and diminished decision-making capacity. If the focus is on substituting rather than augmenting human experience, we may witness the replacement of human labor by AI, concentrating power and wealth in the hands of those who possess the most advanced technology, thus further marginalizing already disadvantaged groups. The academic surgical community must actively explore LLM applications and implement them thoughtfully to enhance existing practices, ensuring they prioritize patients' best interests while minimizing harm.

Conclusions

The public has already begun to accept and integrate LLMs into their lives. Despite current limitations, the rapidly evolving nature of LLMs show promise for integration into surgical practice for improved efficiency and may afford surgeons more time in the operating room for greater job satisfaction. However, as LLMs and other AI technologies advance, the field must approach integration with caution and constant critique. The future has arrived and instead of an anchored resistance to change, the academic surgical community would be best served by investigating the potential applications and slowly implementing the safest elements that augment surgical practice while restricting use that creates harm.

ARTICLE INFORMATION

Accepted for Publication: October 12, 2023.

Published Online: February 14, 2024. doi:10.1001/jamasurg.2023.6496

Author Contributions: Dr Salehinejad had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: All authors.
Acquisition, analysis, or interpretation of data: Rengers, Thiels.
Drafting of the manuscript: Rengers, Thiels.
Critical review of the manuscript for important intellectual content: All authors.
Administrative, technical, or material support: All authors.
Supervision: Thiels, Salehinejad.

Conflict of Interest Disclosures: None reported.

REFERENCES

1. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2(2):e0000198. doi:10.1371/journal.pdig.0000198

2. Stokel-Walker C. ChatGPT listed as author on research papers: many scientists disapprove. *Nature*. 2023;613(7945):620-621. doi:10.1038/d41586-023-00107-z

3. Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT

and a call for papers. *JMIR Med Educ*. 2023;9:e46885. doi:10.2196/46885

4. Open AI. Accessed May 13, 2023. https://openai.com/

5. Newell A. AAAI president's message. *ACM SIGART Bull*. 1980;(73):3-7. doi:10.1145/1056768.1056769

6. van Melle W. MYCIN: a knowledge-based consultation program for infectious disease diagnosis. *Int J Man Mach Stud*. 1978;10(3):313-322. doi:10.1016/S0020-7373(78)80049-2

7. Pople HE. The formation of composite hypotheses in diagnostic problem solving an exercise in synthetic reasoning. In: *Proceedings of the 5th International Joint Conference on Artificial Intelligence*. 1977;2:1030-1037.

8. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *arXiv*. Posted online December 5, 2017. doi:10.48550/arXiv.1706.03762

9. Moore RC. Practical natural language processing by computer: Sri Project 1605. Sri International; 1981.

10. Gherrity M. A learning algorithm for analog, fully recurrent neural networks. In: *Proceedings of the International 1989 Joint Conference on Neural Networks*. 1989;1(1):643-644. doi:10.1109/IJCNN.1989.118645

11. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735-1780. doi:10.1162/neco.1997.9.8.1735

12. Radford A, Narasimham K, Salimans T, Sutskever I, OpenAI. Improving language understanding by generative pre-training. Accessed January 2, 2024. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

13. OpenAI. Introducing ChatGPT. November 30, 2022. Accessed June 29, 2023. https://openai.com/blog/chatgpt

14. AAAI. About AAAI. Accessed May 13, 2023. https://aaai.org/about-aaai/

15. Hasman A. My journey through the field of medical informatics. *Stud Health Technol Inform*. 2022;300:38-52. doi:10.3233/SHTI220940

16. Radford A, Kim JW, Hallacy C, et al. Learning transferable visual models from natural language supervision. *arXiv*. Posted online February 26, 2021. doi:10.48550/arXiv.2103.00020

17. Feng Y, Sim Zheng Ting J, Xu X, et al. Deep neural network augments performance of junior residents in diagnosing covid-19 pneumonia on chest radiographs. *Diagnostics (Basel)*. 2023;13(8):1397. doi:10.3390/diagnostics13081397

18. Bressem KK, Adams LC, Gaudin RA, et al. Highly accurate classification of chest radiographic reports using a deep learning natural language model pre-trained on 3.8 million text reports. *Bioinformatics*. 2021;36(21):5255-5261. doi:10.1093/bioinformatics/btaa668

19. Elhage N, Nanda N, Olsson C, et al. A Mathematical Framework for Transformer Circuits. Transformer Circuits Thread. December 21, 2021. Accessed January 11, 2024. https://transformer-circuits.pub/2021/framework/index.html

20. Cao Y, Li S, Liu Y, et al. A comprehensive survey of AI-generated content (AIGC): a history of generative AI from GAN to ChatGPT. Posted online March 7, 2023. *arXiv*. https://arxiv.org/abs/2303.04226

21. Holzer E, Tschan F, Kottwitz MU, Beldi G, Businger AP, Semmer NK. The workday of hospital surgeons: what they do, what makes them satisfied, and the role of core tasks and administrative tasks; a diary study. *BMC Surg*. 2019;19(1):112. doi:10.1186/s12893-019-0570-0

22. Hampf K. Development of a task management application. Thesis. Åland University of Applied Sciences; 2021. Accessed on May 21, 2023. https://www.theseus.fi/bitstream/handle/10024/356037/Hampf_Kare.pdf

23. Eloundou T, Manning S, Mishkin P, Rock D. GPTs are GPTs: an early look at the labor market impact potential of large language models. Posted online March 23, 2023. *arXiv*. https://arxiv.org/abs/2303.10130

24. PR Newswire. Notable launches Patient AI, the world's first use of large language models and GPT to bring personalization at scale to healthcare. Accessed September 5, 2023. https://www.prnewswire.com/news-releases/notable-launches-patient-ai-the-worlds-first-use-of-large-language-models-and-gpt-to-bring-personalization-at-scale-to-healthcare-301796373.html

25. White C, Bradley E, Martindale J, et al. Why are medical students 'checking out' of active learning in a new curriculum? *Med Educ*. 2014;48(3):315-324. doi:10.1111/medu.12356

26. White C, McCollum M, Bradley E, et al. Challenges to engaging medical students in a flipped classroom model. *Med Sci Educ*. 2015;25(3):219-222. doi:10.1007/s40670-015-0125-7

27. Buchanan BB, Allen GB, Pumphrey CM, Swaiti AR, Harris HM, Boyer PJ. Engaging medical students in the foundational curriculum using third-party resources. *Med Teach*. Published online July 26, 2022. doi:10.1080/0142159X.2022.2102472

28. Wothe JK, Wanberg LJ, Hohle RD, et al. Academic and wellness outcomes associated with use of Anki spaced repetition software in medical school. *J Med Educ Curric Dev*. 2023;10:23821205231173289. doi:10.1177/23821205231173289

29. Sinha RK, Deb Roy A, Kumar N, Mondal H. Applicability of ChatGPT in assisting to solve higher order problems in pathology. *Cureus*. 2023;15(2):e35237. doi:10.7759/cureus.35237

30. Doctors are using ChatGPT to improve how they talk to patients. *The New York Times*. Accessed June 19, 2023. https://www.nytimes.com/2023/06/12/health/doctors-chatgpt-artificial-intelligence.html

31. Webb JJ. Proof of concept: using ChatGPT to teach emergency physicians how to break bad news. *Cureus*. 2023;15(5):e38755. doi:10.7759/cureus.38755

32. Sanmarchi F, Bucci A, Nuzzolese AG, et al. A step-by-step researcher's guide to the use of an AI-based transformer in epidemiology: an exploratory analysis of ChatGPT using the STROBE checklist for observational studies. *Z Gesundheitswissenschaften J Public Health*. Published online May 26, 2023:1-36. doi:10.1007/s10389-023-01936-y

33. Gao CA, Howard FM, Markov NS, et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *NPJ Digit Med*. 2023;6(1):75. doi:10.1038/s41746-023-00819-6

34. Lin Z, Akin H, Rao R, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*. 2023;379(6637):1123-1130. doi:10.1126/science.ade2574

35. Philippidis A. AI-driven pharma tech firm expands its discovery platform into biologics. *Genet Eng Biotechnol News*. 2023;43(1):10-11. doi:10.1089/gen.43.01.02

36. Shanehsazzadeh A, Bachas S, McPartlon M, et al. Unlocking de novo antibody design with generative artificial intelligence. *bioRxiv*. Posted online January 29, 2023. doi:10.1101/2023.01.08.523187

37. Jamthikar A, Gupta D, Saba L, et al. Cardiovascular/stroke risk predictive calculators: a comparison between statistical and machine learning models. *Cardiovasc Diagn Ther*. 2020;10(4):919-938. doi:10.21037/cdt.2020.01.07

38. Lee J, Yoon W, Kim S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. 2020;36(4):1234-1240. doi:10.1093/bioinformatics/btz682

39. Han Q, Tian S, Zhang J. A PubMedBERT-based classifier with data augmentation strategy for detecting medication mentions in tweets. *arXiv*. Posted online November 3, 2021. doi:10.48550/arXiv.2112.02998

40. Luo R, Sun L, Xia Y, et al. BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Brief Bioinform*. 2022;23(6):bbac409. doi:10.1093/bib/bbac409

41. BioMedLM: a domain-specific large language model for biomedical text. Mosaic. Accessed June 28, 2023. https://www.mosaicml.com/blog/introducing-pubmed-gpt

42. Doximity rolls out beta version of ChatGPT tool for docs. Fierce Healthcare. Accessed September 4, 2023. https://www.fiercehealthcare.com/health-tech/doximity-rolls-out-beta-version-chatgpt-tool-docs-aiming-streamline-administrative

43. Bibault JE, Chaix B, Guillemassé A, et al. A chatbot versus physicians to provide information for patients with breast cancer: blind, randomized controlled noninferiority trial. *J Med Internet Res*. 2019;21(11):e15787. doi:10.2196/15787

44. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. *Lancet Digit Health*. 2023;5(4):e179-e181. doi:10.1016/S2589-7500(23)00048-1

45. Seth I, Cox A, Xie Y, et al. Evaluating chatbot efficacy for answering frequently asked questions in plastic surgery: a ChatGPT case study focused on breast augmentation. *Aesthet Surg J*. 2023;43(10):1126-1135. doi:10.1093/asj/sjad140

46. Khanijahani A, Iezadi S, Agoglia S, Barber S, Cox C, Olivo N. Factors associated with information breach in healthcare facilities: a systematic literature review. *J Med Syst*. 2022;46(12):90. doi:10.1007/s10916-022-01877-1

47. OpenAI. What is ChatGPT? Accessed June 28, 2023. https://help.openai.com/en/articles/6783457-what-is-chatgpt

48. Tozzi AE, Cinelli G. Informed consent and artificial intelligence applied to RCT and Covid-19.

BioLaw J. 2021;(2S):97-108. doi:10.15168/2284-4503-849

49. Farnood A, Johnston B, Mair FS. An analysis of the diagnostic accuracy and peer-to-peer health information provided on online health forums for heart failure. *J Adv Nurs*. 2022;78(1):187-200. doi:10.1111/jan.15009

50. Duffourc M, Gerke S. Generative AI in health care and liability risks for physicians and safety concerns for patients. *JAMA*. 2023;330(4):313-314. doi:10.1001/jama.2023.9630

51. Tung JY, Lim DY, Sng GG. Potential safety concerns in use of the artificial intelligence chatbot 'ChatGPT' for perioperative patient communication. *BJU Int*. 2023;132(2):157-159. doi:10.1111/bju.16042

52. Cerdeña JP, Jaswaney R, Plaisime MV. Race, ethnicity, and immigration status in a medical licensing educational resource: a systematic, mixed-methods analysis. *J Gen Intern Med*. 2022;37(5):1045-1051. doi:10.1007/s11606-021-06843-0

53. NIH puts the kibosh on generative AI | SSTI. Accessed September 4, 2023. https://ssti.org/blog/nih-puts-kibosh-generative-ai

54. Brynjolfsson E. The Turing trap: the promise & peril of human-like artificial intelligence. *Daedalus*. 2022;151(2):272-287. doi:10.1162/daed_a_01915