

Large language model

- [Protein Electrostatic Properties are Fine-Tuned Through Evolution](#)
- [More than a chatbot: a practical framework to harness artificial intelligence across key components to boost digital therapeutics quality](#)
- [Artificial Intelligence in Vascular Neurology: Applications, Challenges, and a Review of AI Tools for Stroke Imaging, Clinical Decision Making, and Outcome Prediction Models](#)
- [ProBASS-a language model with sequence and structural features for predicting the effect of mutations on binding affinity](#)
- [Integrating Fine-Tuned LLM with Acoustic Features for Enhanced Detection of Alzheimer's Disease](#)
- [An Intervention to Increase Advance Care Planning Among Older Adults With Advanced Cancer: A Randomized Clinical Trial](#)
- [ND-AMD: A Web-Based Database for Animal Models of Neurological Disease With Analysis Tools](#)
- [Derivational morphology reveals analogical generalization in large language models](#)

see [Large language models for neurosurgery](#).

Large Language Models (LLMs) have emerged as a groundbreaking technology with immense potential to revolutionize various aspects of healthcare. These models, such as [GPT-3](#), [GPT-4](#), and [Med-PaLM 2](#) have demonstrated remarkable capabilities in understanding and generating human-like text, making them valuable tools for tackling complex medical tasks and improving patient care. They have notably shown promise in various medical applications, such as medical question-answering (QA), dialogue systems, and text generation. Moreover, with the exponential growth of [electronic health records](#) (EHRs), [medical literature](#), and patient-generated data, LLMs could help healthcare professionals extract valuable insights and make informed decisions.

As of now, several **large language models (LLMs)** are widely employed across industries and applications. These include models developed by major research organizations and tech companies. Here are some of the most employed LLMs:

OpenAI 1. GPT-4

1. Developer: OpenAI
2. Applications: ChatGPT (this model), Microsoft products like Word and Excel (via Copilot), customer service, coding assistance.
3. Notable Features: State-of-the-art performance in natural language understanding and generation.

2. Codex

1. Developer: OpenAI
2. Applications: Coding assistance (e.g., GitHub Copilot), automating programming tasks.

Google DeepMind 1. Bard (LaMDA series)

1. Developer: Google
2. Applications: Conversational AI, Google Search enhancements, customer support.
3. Notable Features: Integration with Google's ecosystem, tuned for real-world dialogue.

2. PaLM (Pathways Language Model)

1. Developer: Google
2. Applications: Multimodal tasks, large-scale data processing, research.
3. Notable Features: Multilingual capabilities, advanced reasoning.

Meta (formerly Facebook) 1. LLaMA (Large Language Model Meta AI)

1. Applications: Research, open access for experimentation.
2. Notable Features: Lightweight and efficient compared to other models.

Anthropic 1. Claude

1. Developer: Anthropic
2. Applications: Conversational AI, safety-focused use cases.
3. Notable Features: Emphasis on alignment and ethical considerations.

Microsoft 1. Orca and Turing-NLG

1. Applications: Enterprise solutions, customer interaction, integration with Azure.
2. Notable Features: Optimized for cloud-based applications.

Hugging Face 1. Bloom

1. Developer: BigScience/Hugging Face
2. Applications: Open-access research and multilingual tasks.
3. Notable Features: Open-source, designed for global languages.

Alibaba 1. Tongyi Qianwen

1. Applications: E-commerce support, enterprise solutions in China.
2. Notable Features: Multilingual capabilities, business-centric use.

Other Noteworthy LLMs 1. Grok (X.AI by Elon Musk)

1. Applications: Conversational AI, integrated into X (formerly Twitter).
2. Notable Features: Focus on personalized and social interactions.

2. Ernie (Baidu)

1. Applications: Chinese-language tasks, enterprise AI.
2. Notable Features: Specialized for Mandarin and other Chinese dialects.

3. ChatGLM

1. Developer: Tsinghua University and Zhipu AI
2. Applications: Bilingual (English-Chinese) tasks, research.
3. Notable Features: Compact and optimized for specific tasks.

—

These models are widely employed in domains such as **customer service**, **content generation**, **coding**, **research**, **multilingual translation**, and **education**. Their applications are rapidly expanding as AI technology evolves.

These LLMs have had a significant impact in the fields of biomedicine and health care, particularly in the context of [medical education](#).

A large language model (LLM) is a type of [artificial intelligence](#) that uses [deep learning](#) techniques to understand and generate [human language](#). It is typically trained on vast amounts of text data to recognize patterns in language, which allows it to predict, generate, or translate text based on input prompts. These models are capable of performing a wide range of tasks, such as answering questions, summarizing content, writing essays, and engaging in conversation. One of the most well-known LLMs is OpenAI's GPT (Generative Pretrained Transformer) model, which powers systems like [ChatGPT](#).

Individuals are likely turning to Large Language Models (LLMs) to seek health advice, much like searching for diagnoses on [Google](#). Sandmann et al. evaluated the clinical [accuracy](#) of GPT-3.5 and GPT-4 for suggesting initial diagnosis, examination steps, and treatment of 110 medical cases across diverse clinical disciplines. Moreover, two model configurations of the Llama 2 open-source LLMs are assessed in a sub-study. For benchmarking the diagnostic task, they conduct a naïve Google search for comparison. Overall, GPT-4 performed best with superior performances over GPT-3.5 considering diagnosis and examination and superior performance over Google for diagnosis. Except for treatment, better performance on frequent vs rare diseases is evident for all three approaches. The sub-study indicates slightly lower performances for Llama models. In conclusion, the commercial LLMs show growing potential for medical question answering in two successive major releases. However, some weaknesses underscore the need for robust and regulated AI models in health care. Open-source LLMs can be a viable option to address specific needs regarding data privacy and transparency of training ¹⁾.

Large [language models](#) (LLMs) have broad potential applications in medicine, such as aiding with education, providing reassurance to patients, and supporting clinical decision-making. However, there

is a notable gap in understanding their applicability and performance in the surgical domain and how their performance varies across specialties. This paper aims to evaluate the performance of LLMs in answering surgical questions relevant to clinical practice and to assess how this performance varies across different surgical specialties. We used the MedMCQA dataset, a large-scale multi-choice question-answer (MCQA) dataset consisting of clinical questions across all areas of medicine. We extracted the relevant 23,035 surgical questions and submitted them to the popular LLMs Generative Pre-trained Transformers (GPT)-3.5 and GPT-4 (OpenAI OpCo, LLC, San Francisco, CA). A Generative Pre-trained Transformer is a large language model that can generate human-like text by predicting subsequent words in a sentence based on the context of the words that come before it. It is pre-trained on a diverse range of texts and can perform a variety of tasks, such as answering questions, without needing task-specific training. The question-answering accuracy of GPT was calculated and compared between the two models and across surgical specialties. Both GPT-3.5 and GPT-4 achieved accuracies of 53.3% and 64.4%, respectively, on surgical questions, showing a statistically significant difference in performance. When compared to their performance on the full MedMCQA dataset, the two models performed differently: GPT-4 performed worse on surgical questions than on the dataset as a whole, while GPT-3.5 showed the opposite pattern. Significant variations in accuracy were also observed across different surgical specialties, with strong performances in anatomy, vascular, and pediatric surgery and worse performances in orthopedics, ENT, and neurosurgery. Large language models exhibit promising capabilities in addressing surgical questions, although the variability in their performance between specialties cannot be ignored. The lower performance of the latest GPT-4 model on surgical questions relative to questions across all medicine highlights the need for targeted improvements and continuous updates to ensure relevance and accuracy in surgical applications. Further research and continuous monitoring of LLM performance in surgical domains are crucial to fully harnessing their potential and mitigating the risks of misinformation ²⁾.

ChatGPT is a recently developed [Large Language Model](#) LLM that was trained on a massive [dataset](#) of [text](#) for [dialogue](#) with users. Although AI-based language models like ChatGPT have demonstrated impressive capabilities, it is uncertain how well they will perform in real-world scenarios, particularly in fields such as [medicine](#) where high-level and complex thinking is necessary. Furthermore, while the use of ChatGPT in writing scientific [articles](#) and other scientific outputs may have potential benefits, important ethical concerns must also be addressed. Consequently, Cascella et al. investigated the feasibility of ChatGPT in clinical and research scenarios: (1) support of the [clinical practice](#), (2) scientific [production](#), (3) [misuse](#) in medicine and research, and (4) reasoning about public health topics. Results indicated that it is important to recognize and promote education on the appropriate use and potential pitfalls of AI-based LLMs in medicine ³⁾.

¹⁾

Sandmann S, Riepenhausen S, Plagwitz L, Varghese J. Systematic analysis of ChatGPT, Google search and Llama 2 for clinical decision support tasks. *Nat Commun.* 2024 Mar 6;15(1):2050. doi: 10.1038/s41467-024-46411-8. PMID: 38448475.

²⁾

Murphy Lonergan R, Curry J, Dhas K, Simmons BI. Stratified Evaluation of GPT's Question Answering in Surgery Reveals Artificial Intelligence (AI) Knowledge Gaps. *Cureus.* 2023 Nov 14;15(11):e48788. doi: 10.7759/cureus.48788. PMID: 38098921; PMCID: PMC10720372.

³⁾

Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *J Med Syst.* 2023 Mar 4;47(1):33. doi: 10.1007/s10916-023-01925-4. PMID: 36869927; PMCID: PMC9985086.

From:

<https://neurosurgerywiki.com/wiki/> - **Neurosurgery Wiki**

Permanent link:

https://neurosurgerywiki.com/wiki/doku.php?id=large_language_model

Last update: **2025/05/08 22:23**

