

# Biomedical AI in the Age of ChatGPT

## Advancements, Limitations, and the Enduring Role of Traditional Machine Learning

David C. Anastasiu<sup>1,✉</sup>

<sup>1</sup>Computer Science and Engineering Department, Santa Clara University, Santa Clara, CA, USA

The intersection of artificial intelligence (AI) and biomedicine is rapidly evolving, presenting unprecedented opportunities to transform healthcare and research. Large language models (LLMs), exemplified by the widespread public interest in models like ChatGPT, stand at the forefront of this revolution, demonstrating remarkable capabilities in processing and generating human-like text. These advancements have ushered in a new era of possibilities within the biomedical field, ranging from enhancing diagnostic accuracy to accelerating drug discovery. However, despite their impressive progress, LLMs are not without limitations, particularly in scenarios involving small datasets where traditional machine learning methods may still offer significant advantages. This report will explore the recent advancements of LLMs in biomedicine, delve into their inherent limitations, especially concerning data scarcity, highlight the strengths of traditional machine learning approaches in specific biomedical contexts, and provide comparative examples to illustrate these points.

biomedicine | AI | large language models | ChatGPT  
Correspondence: [danastasiu@scu.edu](mailto:danastasiu@scu.edu)

### Introduction

The intersection of artificial intelligence (AI) and biomedicine is rapidly evolving, presenting unprecedented opportunities to transform healthcare and research. Large language models (LLMs), exemplified by the widespread public interest in models like ChatGPT, stand at the forefront of this revolution, demonstrating remarkable capabilities in processing and generating human-like text [1]. These advancements have ushered in a new era of possibilities within the biomedical field, ranging from enhancing diagnostic accuracy to accelerating drug discovery [1–3]. However, despite their impressive progress, LLMs are not without limitations, particularly in scenarios involving small datasets where traditional machine learning methods may still offer significant advantages. This report will explore the recent advancements of LLMs in biomedicine, delve into their inherent limitations, especially concerning data scarcity, highlight the strengths of traditional machine learning approaches in specific biomedical contexts, and provide comparative examples to illustrate these points.

### Recent Advancements of LLMs in Biomedicine

LLMs are sophisticated deep learning models that have achieved remarkable proficiency in natural language processing (NLP). Their core functionality lies in the ability to understand, generate, and interact through language by predicting the next word in a sequence based on statistical pat-

terns learned from vast amounts of text data [1, 4]. This capability forms the bedrock for their expanding applications across various domains, including the text-intensive field of biomedicine.

At the heart of many LLMs lies the Transformer model [5], a novel architecture that has revolutionized how AI processes sequential data like text. Unlike earlier models that processed text word by word, the Transformer can look at all the words in a sentence simultaneously, allowing it to understand the context and relationships between words more effectively. A key component of the Transformer is the “self-attention mechanism,” which enables the model to weigh the importance of different words in the input when processing each word. This allows the model to understand, for example, that the word “protein” in one part of a sentence might be related to “structure” mentioned later in the same sentence, regardless of the distance between them [6]. This ability to grasp long-range dependencies in text is crucial for biomedical applications where understanding complex relationships within scientific literature or patient records is essential.

One of the most promising areas for Transformers is **clinical decision support** [7], where LLMs can assist healthcare professionals in many ways. They possess the potential to diagnose diseases with improved accuracy and speed by processing and interpreting large volumes of patient data, such as electronic health records and imaging results. Moreover, LLMs can aid in treatment planning by suggesting potential options based on the latest medical guidelines and patient-specific information. Their ability to swiftly navigate extensive medical literature also provides clinicians with essential research, guidelines, and information, saving valuable time and ensuring that medical treatments are grounded in current knowledge. Table 1 showcases a few recent LLMs and their key features and primary applications.

In **medical research**, LLMs serve as powerful tools for accelerating discovery. They can efficiently retrieve and organize crucial information from the vast landscape of biomedical literature, enabling researchers to access and synthesize the latest findings on specific drugs, diseases, or genes [4]. Initiatives like BioChatter [18], an open-source Python framework, are making LLMs more accessible to biomedical researchers, streamlining workflows for tasks such as data extraction and analysis.

**Drug discovery and development** is also being significantly impacted by LLMs. These models can predict the activity, toxicity, and pharmacokinetic properties of new compounds, facilitating early-stage drug screening and potentially shortening the drug development timeline [19, 20]. LLM-based AI agents, capable of orchestrating complex research work-

**Table 1.** Examples of Biomedical LLMs

Model Name	Key Features	Primary Applications	Source
MediSwift	Sparse pre-training on biomedical text, computationally efficient	Information retrieval, question answering, text generation in biomedicine	Thangarasa et al. (2024) [8]
BioMedLM	2.7B parameter GPT model trained on PubMed abstracts and full articles	Biomedical question answering	Bolton et al. (2024) [9, 10]
BioMistral	Open-source LLM tailored for the biomedical domain	Various biomedical NLP tasks	Labrak et al. (2024) [11]
CEHR-GPT	Clinical text generation using pre-training on medical text data	Automated report generation, clinical documentation, discharge summaries	Pang et al. (2024) [12, 13]
AlphaFold (AF2)	Deep learning architecture for protein structure prediction from amino acid sequences	Predicting protein structures, accelerating drug discovery and functional genomics	Jumper et al. (2021) [14–17]

flows and generating molecular design ideas, are emerging as valuable assistants for researchers.

In **genomics and proteomics**, LLMs are proving instrumental in analyzing the intricate language of biological sequences. Models like DNABERT [21] and AlphaFold [14] leverage the transformer architecture to predict functional regions in DNA, understand RNA splicing, and predict protein structures with unprecedented accuracy. This capability enhances our understanding of gene function, disease mechanisms, and facilitates the development of novel therapeutics [15, 16, 22–24].

The analysis of **electronic health records (EHRs)** represents another significant application area for LLMs, which can very quickly process and summarize vast amounts of information contained within EHRs, potentially improving care continuity and decision-making [25–28]. Custom LLMs tailored for healthcare language nuances can enhance the precision of clinical documentation and streamline the extraction of pertinent information from voluminous datasets [27].

Furthermore, the advent of **multimodal LLMs (MLLMs)** has expanded the possibilities by enabling the integration of different data modalities, such as text and medical images [29–32]. This capability mirrors the way clinicians often process information, potentially leading to more accurate diagnoses by analyzing both a patient’s written symptoms and corresponding radiological images.

## Limitations of Transformers and LLMs in Biomedicine

Despite remarkable recent advancements, the application of LLMs in biomedicine is accompanied by several inherent limitations that must be carefully considered. Table 2 lists these limitations and potential mitigation strategies that may allow LLMs to be effective in Biomedicine.

One significant challenge arises when dealing with **small datasets**. LLMs are data-intensive models, requiring vast amounts of information to learn robust representations and generalize effectively [33, 34]. Additionally, in biomedical domains where data is scarce, such as in the study of rare diseases or specialized clinical trials, LLMs may not perform

optimally. Overall, the limited scale and often lower quality of available annotated biomedical datasets can hinder their full potential.

Another critical concern is the phenomenon of **hallucinations**, where LLMs generate incorrect or fabricated information [35–37]. This tendency to produce non-factual content poses a serious risk in medical applications where accuracy is paramount for patient safety and treatment decisions. In high-stakes areas like radiology, where precise interpretation of medical language is crucial, hallucinations can be particularly detrimental [31].

**Bias and fairness** represent another significant hurdle. LLMs are trained on large datasets that may contain inherent biases related to race, gender, socioeconomic status, and other factors [38–41]. These biases can be reflected and even amplified by the models, potentially leading to disparities and inaccuracies in diagnoses and treatment recommendations for certain patient populations [39, 40].

The **interpretability and explainability** of LLMs also present challenges. Often described as "black-box" models, it can be difficult to understand the reasoning behind their predictions [39]. This lack of transparency can hinder trust and adoption in clinical settings where healthcare professionals require a clear understanding of the rationale behind suggested diagnoses or treatment options.

The **computational resources and infrastructure** required to train and deploy large LLMs can be substantial [4]. This can be a barrier to their widespread implementation, particularly in resource-constrained healthcare settings where access to high-end hardware like GPUs may be limited [38].

Finally, **privacy and security concerns** are paramount when dealing with sensitive patient data in healthcare [39]. Ensuring the confidentiality and security of this information when using LLMs, especially those hosted on external servers, requires robust security protocols and strict adherence to privacy regulations like HIPAA.

**Table 2.** Limitations of LLMs in Biomedicine

Limitation	Description	Potential Mitigation Strategies	Relevant Sources
Small Datasets	LLMs require large amounts of data to train effectively and may struggle with limited biomedical datasets.	Fine-tuning on domain-specific data, data augmentation techniques, transfer learning from larger models, leveraging traditional ML methods for data-scarce scenarios.	[33, 34]
Hallucinations	LLMs can generate incorrect or fabricated information, posing risks in medical applications.	Retrieval-Augmented Generation (RAG) to ground responses in reliable knowledge sources, prompt engineering, fine-tuning with factual data, developing methods to assess the reliability of LLM outputs.	[35–37]
Bias and Fairness	Biases in training data can lead to unfair or inaccurate predictions for certain patient populations.	Careful curation of training data to ensure diversity and representation, bias detection and correction techniques, fairness-aware model training, continuous monitoring of model outputs for bias.	[38–41]
Lack of Interpretability	The "black-box" nature of LLMs makes it difficult to understand the reasoning behind their predictions.	Development of Explainable AI (XAI) methods for LLMs, using traditional ML models for interpretable insights, focusing on applications where interpretability is less critical.	[39]
Computational Resource Demands	Training and running large LLMs requires significant computational power and specialized hardware.	Development of more efficient LLM architectures, using cloud-based computing resources, focusing on smaller, task-specific models where appropriate, leveraging traditional ML for computationally lighter solutions.	[4, 38]
Privacy and Security Concerns	Handling sensitive patient data with LLMs raises concerns about data breaches and compliance with privacy regulations.	Implementing robust security protocols, data anonymization techniques, using privacy-preserving methods like federated learning, ensuring compliance with regulations like HIPAA.	[39]

## Advantages of Traditional Machine Learning in Biomedical Contexts

While LLMs have garnered significant attention, traditional machine learning methods continue to hold several advantages, particularly in specific biomedical scenarios.

One key strength lies in their **effectiveness with small datasets**. Unlike LLMs that thrive on massive amounts of data, traditional ML algorithms such as decision trees, support vector machines, and logistic regression can often achieve good performance with much smaller, well-structured datasets. This makes them particularly suitable for applications where data is limited, such as in rare disease research or specialized clinical studies with a limited number of participants.

**Interpretability and explainability** are another significant advantage. Many traditional ML models, like decision trees and linear regression, offer a clear understanding of how predictions are made. This transparency is crucial in biomedical contexts where clinicians and researchers need to understand the factors driving a particular prediction to validate its reliability and build trust.

Traditional ML models generally have **lower computational requirements** compared to LLMs. They can often be trained

and run efficiently on standard CPUs, making them more accessible and cost-effective for a wider range of institutions and applications, especially in resource-limited settings.

The field of traditional machine learning boasts **established methodologies and a rich ecosystem of tools and libraries**, such as scikit-learn. Researchers and practitioners have a wealth of experience and resources available for developing, validating, and deploying these models in biomedical applications.

Furthermore, traditional ML methods are often **well-suited for specific types of tasks** commonly encountered in biomedicine. While LLMs excel at generative tasks and natural language understanding, traditional ML algorithms are highly effective for prediction and classification tasks, such as disease risk assessment, diagnostic prediction, and identifying patterns in structured data.

## Comparative Examples

Several studies have demonstrated scenarios where traditional machine learning methods outperform LLMs in biomedical applications, particularly in clinical prediction tasks. Research using electronic health record (EHR) data from Vanderbilt University Medical Center and MIMIC-

IV has shown that traditional ML models, such as gradient boosting trees, achieved substantially higher accuracy (AUROC) and better calibration (Brier Score) in predicting patient discharge compared to both GPT-3.5 and GPT-4, even when using in-context learning [42, 43]. This suggests that for certain well-defined prediction tasks with structured data, traditional ML remains a more effective approach than general-purpose LLMs.

Moreover, traditional ML has been shown to be more **robust to the generalization of demographic information** for privacy protection [42, 43]. This indicates that when sensitive demographic features need to be obscured or removed, traditional ML models may maintain better predictive performance compared to LLMs.

LLMs, primarily trained on unstructured text data, have also been observed to lag behind traditional ML methods in **predicting outcomes from structured tabular data**, which is a common format in biomedical datasets containing numerical measurements and clinical parameters [44]. Traditional ML algorithms are specifically designed to learn patterns and relationships within this type of data, giving them a potential advantage.

Beyond clinical prediction, traditional ML methods like XGBoost [45, 46] and classical deep learning methods like convolutional neural networks (CNNs) [47, 48], recurrent neural networks (RNNs) [49, 50] have a long and successful history in various other biomedical applications, including **genomics**, where they are used for tasks such as predicting the type of different peptides, classifying cell types from single-cell RNA sequencing data and identifying genetic variants. While LLMs are increasingly being explored in genomics, traditional ML remains a workhorse for many specific analytical tasks in this field.

## Conclusion

Large language models have undoubtedly ushered in a transformative era for biomedical AI, demonstrating remarkable capabilities across a wide spectrum of applications, from clinical decision support to drug discovery and genomic analysis. Their ability to process and generate human-like text has opened up new avenues for understanding and interacting with complex biomedical data. However, inherent limitations such as challenges with small datasets, the risk of hallucinations and bias, and issues with interpretability necessitate a nuanced perspective on their role. Traditional machine learning methods continue to offer significant advantages, particularly in scenarios involving limited data, the need for transparent decision-making, and specific prediction or classification tasks. The evidence suggests that for certain clinical prediction problems and the analysis of structured tabular data, traditional ML can still outperform even advanced LLMs. The future of biomedical AI likely lies in a synergistic approach, where the strengths of both LLMs and traditional ML are leveraged to create more robust, reliable, and ethically sound solutions that ultimately benefit human health. Ongoing research efforts aimed at mitigating the limitations of LLMs in biomedicine hold the promise of

further expanding their utility, but for the foreseeable future, traditional machine learning will continue to play a vital and complementary role in advancing the field.

## Bibliography

1. Kuo Zhang, Xiangbin Meng, Xiangyu Yan, Jiaming Ji, Jingqian Liu, Hua Xu, Heng Zhang, Da Liu, Jingjia Wang, Xuliang Wang, Jun Gao, Yuan-geng-shuo Wang, Chunli Shao, Wenyao Wang, Jiarong Li, Ming-Qi Zheng, Yaodong Yang, and Yi-Da Tang. Revolutionizing health care: The transformative impact of large language models in medicine. *J Med Internet Res*, 27:e59069, Jan 2025. ISSN 1438-8871. doi: 10.2196/59069.
2. Xiao-Huan Liu, Zhen-Hua Lu, Tao Wang, and Fei Liu. Large language models facilitating modern molecular biology and novel drug development. *Front Pharmacol*, 15:1458739, December 2024.
3. Milad Alucozai, Will Fondrie, and Megan Sperry. From data to drugs: The role of artificial intelligence in drug discovery. <https://wyss.harvard.edu/news/from-data-to-drugs-the-role-of-artificial-intelligence-in-drug-discovery/>, Jan 2025. Accessed on: 2025-04-18.
4. Chong Wang, Mengyao Li, Junjun He, Zhongruo Wang, Erfan Darzi, Zan Chen, Jin Ye, Tianbin Li, Yanzhou Su, Jing Ke, Kaili Qu, Shuxin Li, Yi Yu, Pietro Liò, Tianyun Wang, Yu Guang Wang, and Yiqing Shen. A survey for large language models in biomedicine. <https://arxiv.org/abs/2409.00133>, 2024.
5. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
6. Ha Na Cho, Tae Joon Jun, Young-Hak Kim, Heejun Kang, Imjin Ahn, Hansle Gwon, Yunha Kim, Jiahn Seo, Heejung Choi, Minkyung Kim, Jiye Han, Gaewon Kee, Seohyun Park, and Soyoung Ko. Task-specific transformer-based language models in health care: Scoping review. *JMIR Med Inform*, 12:e49724, Nov 2024. ISSN 2291-9694. doi: 10.2196/49724.
7. Xiao-Huan Liu, Zhen-Hua Lu, Tao Wang, and Fei Liu. Large language models facilitating modern molecular biology and novel drug development. *Front Pharmacol*, 15:1458739, December 2024.
8. Vithursan Thangarasa, Mahmoud Salem, Shreyas Saxena, Chen-Yu Leong, Joel Hestness, and Sean Lie. MediSwift: Efficient sparse pre-trained biomedical language models. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Findings of the Association for Computational Linguistics: ACL 2024*, pages 214–230, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.14.
9. Elliot Bolton, Abhinav Venigalla, Michihiro Yasunaga, David Hall, Betty Xiong, Tony Lee, Roxana Daneshjou, Jonathan Frankle, Percy Liang, Michael Carbin, and Christopher D. Manning. Biomedlm: A 2.7b parameter language model trained on biomedical text. <https://arxiv.org/abs/2403.18421>, 2024.
10. Elliot Bolton, Betty Xiong, Vijaytha Muralidharan, Joel Schamroth, Vivek Muralidharan, Christopher D. Manning, and Roxana Daneshjou. Assessing the potential of mid-sized language models for clinical qa. <https://arxiv.org/abs/2404.15894>, 2024.
11. Yanis Labrak, Adrien Bazoge, Emmanuel Morin, Pierre-Antoine Gourraud, Mickael Rouvier, and Richard Dufour. BioMistral: A collection of open-source pretrained large language models for medical domains. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Findings of the Association for Computational Linguistics: ACL 2024*, pages 5848–5864, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.348.
12. Chao Pang, Xinzhuo Jiang, Krishna S. Kalluri, Matthew Spotnitz, RuiJun Chen, Adler Perotte, and Karthik Natarajan. Cehr-bert: Incorporating temporal information from structured ehr data to improve prediction tasks. In Subhrajit Roy, Stephen Plohl, Emma Rocheteau, Girmaw Abebe Tadesse, Luis Oala, Fabian Falk, Yuyin Zhou, Liyue Shen, Ghada Zamzmi, Purity Mugambi, Ayah Zirikly, Matthew B. A. McDermott, and Emily Alsentzer, editors, *Proceedings of Machine Learning for Health*, volume 158 of *Proceedings of Machine Learning Research*, pages 239–260. PMLR, 04 Dec 2021.
13. Chao Pang, Xinzhuo Jiang, Nishanth Parameshwar Pavinkurve, Krishna S. Kalluri, Elise L. Minto, Jason Patterson, Linying Zhang, George Hripcsak, Gamze Gürsoy, Noémie Elhadad, and Karthik Natarajan. Cehr-gpt: Generating electronic health records with chronological patient timelines. <https://arxiv.org/abs/2402.04400>, 2024.
14. John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Židek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A A Kohl, Andrew J Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislaw Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zieliński, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021.
15. Leticia M. F. Bertoline, Angélica N. Lima, Jose E. Krieger, and Samantha K. Teixeira. Before and after alphafold2: An overview of protein structure prediction. *Frontiers in Bioinformatics*, Volume 3 - 2023, 2023. ISSN 2673-7647. doi: 10.3389/fbinf.2023.1120370.
16. Nazim Bouattata, Peter Sorger, and Mohammed AlQuraishi. Protein structure prediction by AlphaFold2: are attention and symmetries all you need? *Acta Crystallographica Section D*, 77(8):982–991, Aug 2021. doi: 10.1107/S2059798321007531.
17. Thomas C Terwilliger, Dorothee Liebschner, Tristan I Croll, Christopher J Williams, Airlie J McCoy, Billy K Poon, Pavel V Afonine, Robert D Oeffner, Jane S Richardson, Randy J Read, and Paul D Adams. AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nat Methods*, 21(1):110–116, November 2023.
18. Sebastian Lobentanzer, Shaohong Feng, Noah Bruderer, Andreas Maier, Adrián G Díaz, Amy Strange, Anis Ismail, Anton Kulaga, Aurelien Gourgoud, Barbara Zdravil, Bastien Chas-



- sagnol, Cyril Pommier, Daniele Lucarelli, Ellen M McDonagh, Emma Verkinderen, Fernando M Delgado-Chaves, Georg Fuellen, Hannah Sonntag, Jonatan Menger, Lionel Christiaen, Ludwig Geistlinger, Luna Zacharias Zetsche, Marlis Engelke, Megan McNutt, Melissa Harrison, Melissa Hizli, Nikolai Usanov, Patrick Baracho, Sebastian Beier, Stefan Boeing, Taru A Muranen, Trang T Le, Valeria Dragan, Xiao-Ran Zhou, Yasmin Nielsen-Tehranchian, Yuyao Song, Cankun Wang, Jan Baumbach, Jorge Abreu-Vicente, Nils Krehl, Qin Ma, Thomas Lemberger, Julio Saez-Rodriguez, and The BioChatter Consortium. A platform for the biomedical application of large language models. *Nature Biotechnology*, 43(2):166–169, February 2025.
19. Bowen Gao, Yanwen Huang, Yiqiao Liu, Wenxuan Xie, Wei-Ying Ma, Ya-Qin Zhang, and Yanyan Lan. Pharmagents: Building a virtual pharma with large language model agents. <https://arxiv.org/abs/2503.22164>, 2025.
  20. Xiao-Huan Liu, Zhen-Hua Lu, Tao Wang, and Fei Liu. Large language models facilitating modern molecular biology and novel drug development. *Front Pharmacol*, 15:1458739, December 2024.
  21. Yanrong Ji, Zhihan Zhou, Han Liu, and Ramana V Davuluri. Dnabert: pre-trained bidirectional encoder representations from transformers model for dna-language in genome. *Bioinformatics*, 37(15):2112–2120, 02 2021. ISSN 1367-4803. doi: 10.1093/bioinformatics/btab083.
  22. Wei Ruan, Yanjun Lyu, Jing Zhang, Jiazhang Cai, Peng Shu, Yang Ge, Yao Lu, Shang Gao, Yue Wang, Peilong Wang, Lin Zhao, Tao Wang, Yufang Liu, Luyang Fang, Ziyu Liu, Zhengliang Liu, Yiwei Li, Zihao Wu, Junhao Chen, Hanqi Jiang, Yi Pan, Zhenyuan Yang, Jingyuan Chen, Shizhe Liang, Wei Zhang, Terry Ma, Yuan Dou, Jianli Zhang, Xinyu Gong, Qi Gan, Yusong Zou, Zebang Chen, Yuanxin Qian, Shuo Yu, Jin Lu, Kenan Song, Xian-qiao Wang, Andrea Sikora, Gang Li, Xiang Li, Quanzheng Li, Yingfeng Wang, Lu Zhang, Yohannes Abate, Lifang He, Wenxuan Zhong, Rongjie Liu, Chao Huang, Wei Liu, Ye Shen, Ping Ma, Hongtu Zhu, Yajun Yan, Dajiang Zhu, and Tianming Liu. Large language models for bioinformatics. <https://arxiv.org/abs/2501.06271>, 2025.
  23. Binghao Yan, Yunbi Nam, Lingyao Li, Rebecca A. Deek, Hongzhe Li, and Siyuan Ma. Recent advances in deep learning and language models for studying the microbiome. *Frontiers in Genetics*, Volume 15 - 2024, 2025. ISSN 1664-8021. doi: 10.3389/fgene.2024.1494474.
  24. Oluwafemi A Sarumi and Dominik Heider. Large language models and their applications in bioinformatics. *Comput Struct Biotechnol J*, 23:3498–3505, October 2024.
  25. Maya Kruse, Shiyue Hu, Nicholas Derby, Yifu Wu, Samantha Stonbraker, Bingsheng Yao, Dakuo Wang, Elizabeth Goldberg, and Yanjun Gao. Zero-shot large language models for long clinical text summarization with temporal reasoning. <https://arxiv.org/abs/2501.18724>, 2025.
  26. Zhenbang Wu, Anant Dadu, Mike Nalls, Faraz Faghri, and Jimeng Sun. Instruction tuning large language models to understand electronic health records. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 54772–54786. Curran Associates, Inc., 2024.
  27. Sophia Adams. Natural language processing in electronic health records: Custom llm approaches. <https://gaper.io/electronic-health-records-custom-llm/>, Feb 2024.
  28. Liyan Tang, Zhaoyi Sun, Betina Iday, Jordan G Nestor, Ali Soroush, Pierre A Elias, Ziyang Xu, Ying Ding, Greg Durrett, Justin F Rousseau, Chunhua Weng, and Yifan Peng. Evaluating large language models on medical evidence summarization. *npj Digital Medicine*, 6(1): 158, August 2023.
  29. Wan Hang Keith Chiu, Wei Sum Koel Ko, William Chi Shing Cho, Sin Yu Joanne Hui, Wing Chi Lawrence Chan, and Michael D Kuo. Evaluating the diagnostic performance of large language models on complex multimodal medical cases. *J Med Internet Res*, 26:e53724, May 2024. ISSN 1438-8871. doi: 10.2196/53724.
  30. Tianyu Han, Lisa C. Adams, Keno K. Bresse, Felix Busch, Sven Nebelung, and Daniel Truhn. Comparative analysis of multimodal large language model performance on clinical vignette questions. *JAMA*, 331(15):1320–1321, 04 2024. ISSN 0098-7484. doi: 10.1001/jama.2023.27861.
  31. Rajesh Bhayana. Chatbots and large language models in radiology: A practical primer for clinical and research applications. *Radiology*, 310(1):e232756, 2024. doi: 10.1148/radiol.232756. PMID: 38226883.
  32. Marina Hayashida, Masataka Suzuki, Yosuke Nakata, Hiroko Kakita, and Hiroshi Ezawa. Cardiac tamponade and duodenal hemorrhage caused by inappropriate use of dabigatran in a patient with End-Stage renal failure: A case report. *Cureus*, 16(1):e52521, January 2024.
  33. Xunxin Cai, Chengrui Wang, Qingqing Long, Yuanchun Zhou, and Meng Xiao. Knowledge hierarchy guided biological-medical dataset distillation for domain llm training. <https://arxiv.org/abs/2501.15108>, 2025.
  34. Deshiwei Zhang, Xiaojuan Xue, Peng Gao, Zhijuan Jin, Menghan Hu, Yue Wu, and Xiayang Ying. A survey of datasets in medicine for large language models. *Intelligence & Robotics*, 4(4), 2024. ISSN 2770-3541. doi: 10.20517/ir.2024.27.
  35. Dimitri Roustan and François Bastardot. The clinicians’ guide to large language models: A general perspective with a focus on hallucinations. *Interact J Med Res*, 14:e59823, January 2025.
  36. Sumera Anjum, Hanzhi Zhang, Wenjun Zhou, Eun Jin Paek, Xiaopeng Zhao, and Yunhe Feng. Halo: Hallucination analysis and learning optimization to empower llms with retrieval-augmented context for guided clinical decision making. <https://arxiv.org/abs/2409.10011>, 2024.
  37. Yubin Kim, Hyewon Jeong, Shen Chen, Shuyue Stella Li, Mingyu Lu, Kumail Alhamoud, Jimin Mun, Cristina Grau, Minseok Jung, Rodrigo R Gameiro, et al. Medical hallucination in foundation models and their impact on healthcare. *medRxiv*, pages 2025–02, 2025.
  38. Kerstin Denecke, Richard May, and Octavio Rivera-Romero. Transformer models in healthcare: A survey and thematic analysis of potentials, shortcomings and risks. *J Med Syst*, 48(1):23, February 2024.
  39. Ehsan Ullah, Anil Parwani, Mirza Mansoor Baig, and Rajendra Singh. Challenges and barriers of using large language models (LLM) such as ChatGPT for diagnostic medicine with a focus on digital pathology - a recent scoping review. *Diagn Pathol*, 19(1):43, February 2024.
  40. John J Hanna, Abdi D Wakene, Andrew O Johnson, Christoph U Lehmann, and Richard J Medford. Assessing racial and ethnic bias in text generation by large language models for health care-related tasks: Cross-sectional study. *J Med Internet Res*, 27:e57257, Mar 2025. ISSN 1438-8871. doi: 10.2196/57257.
  41. Samuel Schmidgall, Carl Harris, Ime Essien, Daniel Olshvang, Tawsifur Rahman, Ji Woong Kim, Rojin Ziaei, Jason Eshraghian, Peter Abadir, and Rama Chellappa. Evaluation and mitigation of cognitive biases in medical language models. *npj Digital Medicine*, 7(1):295, October 2024.
  42. Katherine E Brown, Chao Yan, Zhuohang Li, Ximmeng Zhang, Benjamin X Collins, You Chen, Ellen Wright Clayton, Murat Kantarcioglu, Yevgeniy Vorobeychik, and Bradley A Malin. Large language models are less effective at clinical prediction tasks than locally trained machine learning models. *Journal of the American Medical Informatics Association*, page ocaf038, 03 2025. ISSN 1527-974X. doi: 10.1093/jamia/ocaf038.
  43. Katherine E. Brown, Chao Yan, Zhuohang Li, Ximmeng Zhang, Benjamin X. Collins, You Chen, Ellen Wright Clayton, Murat Kantarcioglu, Yevgeniy Vorobeychik, and Bradley A. Malin. Not the models you are looking for: Traditional ml outperforms llms in clinical prediction tasks. *medRxiv*, 2024. doi: 10.1101/2024.12.03.24318400.
  44. Jiahuan Yan, Jintai Chen, Chaowen Hu, Bo Zheng, Yaojun Hu, Jimeng Sun, and Jian Wu. Small models are llm knowledge triggers on medical tabular prediction, 2025.
  45. Bipasa Bose, Taylor Downey, Anand K. Ramasubramanian, and David C. Anastasiu. Identification of distinct characteristics of antibiofilm peptides and prospection of diverse sources for efficacious sequences. *Frontiers in Microbiology*, 12, 2022. ISSN 1664-302X.
  46. Nivedha Balakrishnan, Rahul Katkar, Peter V. Pham, Taylor Downey, Prarthna Kashyap, David C. Anastasiu, and Anand K. Ramasubramanian. Prospection of peptide inhibitors of thrombin from diverse origins using a machine learning pipeline. *Bioengineering*, 10(11), 2023. ISSN 2306-5354.
  47. Alex Whelan, Soham Phadke, and David C. Anastasiu. On-device prediction for chronic kidney disease. In *2022 IEEE Global Humanitarian Technology Conference (GHTC)*, GHTC 2022, pages 325–332, 2022.
  48. Alex Whelan, Ragwa Elsayed, Alessandro Bellofiore, and David C. Anastasiu. Selective partitioned regression for accurate kidney health monitoring. *Annals of Biomedical Engineering*, Feb 2024. ISSN 1573-9686.
  49. Wardah S Alharbi and Mamoon Rashid. A review of deep learning applications in human genomics using next-generation sequencing data. *Human Genomics*, 16(1SN - 1479-7364): 26, July 2022.
  50. Nikesht Gyawali, Yangfan Hao, Guifang Lin, Jun Huang, Ravi Bika, Lidia Calderon Daza, Huakun Zheng, Giovana Cruppe, Doina Caragea, David Cook, Barbara Valent, and Sanzhen Liu. Using recurrent neural networks to detect supernumerary chromosomes in fungal strains causing blast diseases. *NAR Genomics and Bioinformatics*, 6(3):lqae108, 08 2024. ISSN 2631-9268. doi: 10.1093/nargab/lqae108.