INTRODUCING AIDX, THE PHYSICIAN'S AI COPILOT

Alan Alwakeel
Founder, AIDx
alan.alwakeel@aidx.md

ABSTRACT

The dynamic environment of medicine, like the Emergency Department, challenges physicians with an influx of patient data and the need for swift diagnosis and treatment under high-pressure situations. Furthermore, there are numerous AI models developed for medical applications that, despite their potential, rarely see hospital implementation because they fail to integrate into pre-existing clinical workflows. Addressing these challenges, I developed AIDx, a state-of-the-art AI-powered system designed to streamline clinical decision-making, provide accurate disease prediction and treatment planning, offer novel approaches for personalized patient healthcare, and serve as a hub for AI-assisted medical analysis. Unlike standalone systems, AIDx integrates seamlessly with a hospital's pre-existing electronic health records and automatically leverages supplemental tools (like an up-to-date medical knowledge database and capability-enhancing AI models), offering physicians a unified platform for comprehensive AI assistance. I trained AIDx on 400,000 de-identified patient charts and evaluated it on 7,000 medical exam questions, achieving a remarkable accuracy of 83.61% that rivals the performance of industry-leading models like Google's Med-PaLM 2 and OpenAI's GPT-4 with significantly fewer parameters (a model size metric), using only 46.7 billion compared to their 340 billion and 1.7 trillion. This parameter efficiency translates into feasible on-premises deployment, high-speed processing, and adaptability across clinical settings. AIDx also significantly enhances patient-physician communication, simplifying medical terminology and facilitating multilingual translation. Ultimately, AIDx represents a transformative integration of AI in medicine, establishing an unprecedented standard for personalized, reliable, and streamlined patient care, elevating the quality of healthcare, and enhancing the lives of patients and physicians alike.

Keywords Emergency Department AI Integration · Clinical Decision-Making AI · AIDx System · AI-Powered Disease Prediction · Electronic Health Records Integration · AI Model Efficiency in Healthcare · Multilingual Medical Communication · Personalized Healthcare AI · AI-Assisted Medical Analysis · Healthcare Quality Enhancement

1 Introduction

The emergency department, a critical entry-point in healthcare, faces a daily influx of nearly 400,000 patients in the United States alone [5], leading to a series of challenges that significantly affect both patient care and the well-being of healthcare professionals. This repeated influx results in prolonged wait times, creating a bottleneck that impedes essential medical interventions. Such delays can vary from minor inconveniences to critical emergencies, potentially exacerbating medical conditions from controllable to severe. Concurrently, this surge places an enormous strain on healthcare professionals, where the pressure to provide prompt and accurate care in an overstretched environment can contribute to burnout, diminishing the quality of patient interactions and possibly affecting care standards.

The advent of artificial intelligence (AI) in various sectors, including healthcare, holds substantial promise for transformative change; however, the integration of AI within hospital settings has been slow, primarily due to the specialized and narrow task orientation of existing AI models. Despite their sophistication, these models often lack interoperability, complicating their consolidation into comprehensive clinical workflows—a scenario that not only limits their practical application but also incurs significant integration costs. Moreover, the emergence of accessible AI tools, such as ChatGPT [11], has already generated interest among healthcare practitioners. Nevertheless, this interest is accompanied

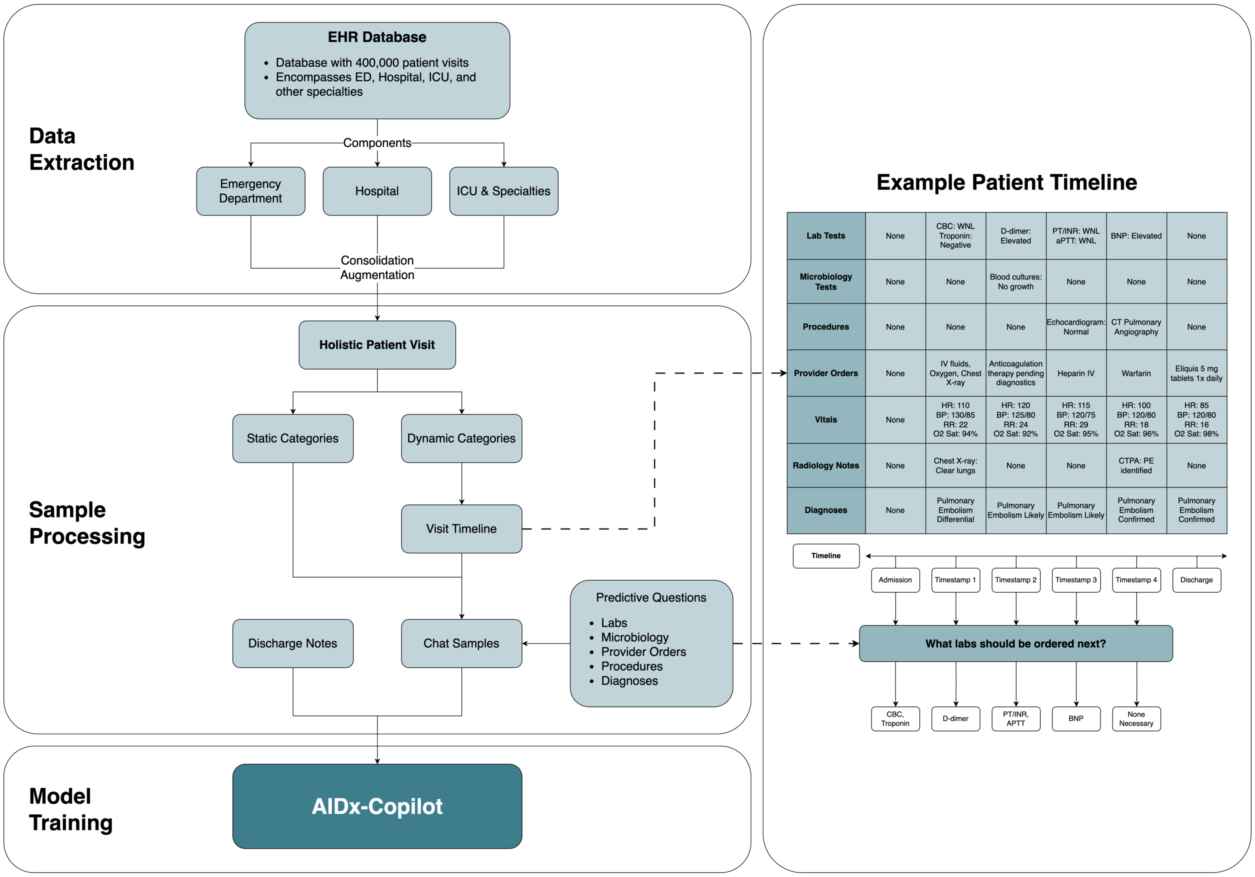


Figure 1: Data processing pipeline for generating training samples for AIDx-Copilot

by concerns regarding the precision of AI-assisted diagnostics and treatments, as well as the safeguarding of patient data in compliance with privacy standards such as HIPAA.

In response to these multifaceted challenges, I developed AIDx, an integrated AI-driven software engineered to surpass these barriers by offering comprehensive support in diagnosis and treatment, enhancing the clinical decision-making process, and promoting personalized patient care. At its core, AIDx features AIDx-Copilot, an advanced, state-of-the-art large language medical model fine-tuned on a comprehensive EHR database encompassing over 400,000 de-identified patient records [8]. AIDx distinguishes itself by employing a nuanced multi-step approach to process physician inquiries, enriching them with relevant medical data, and providing accurate analysis and feedback. This methodology encompasses patient information retrieval, the application of a retrieval-augmented generation (RAG) [10] technique for accessing up-to-date data from a medical knowledge base, and leveraging chain-of-thought (CoT) [19] prompting strategies. These measures ensure the delivery of accurate, dependable outputs by AIDx, thereby setting a new standard for AI-assisted medical care.

2 Software Design

2.1 AIDx-Copilot Training and Implementation

The training of AIDx-Copilot involved an exhaustive preprocessing of patient data from the EHR database to ensure comprehensive coverage of potential use cases and physician-AI interaction scenarios. (Refer to Figure 1)

Consolidation and Augmentation I consolidated comprehensive patient information encompassing various segments of their hospital experience—ranging from their time in the emergency department to stays in the hospital and ICU. This consolidation resulted in a holistic patient chart that separates static (unchanging) categories, such as demographics and medical history, and dynamic (evolving) categories, including lab results, medical orders, and procedural data.

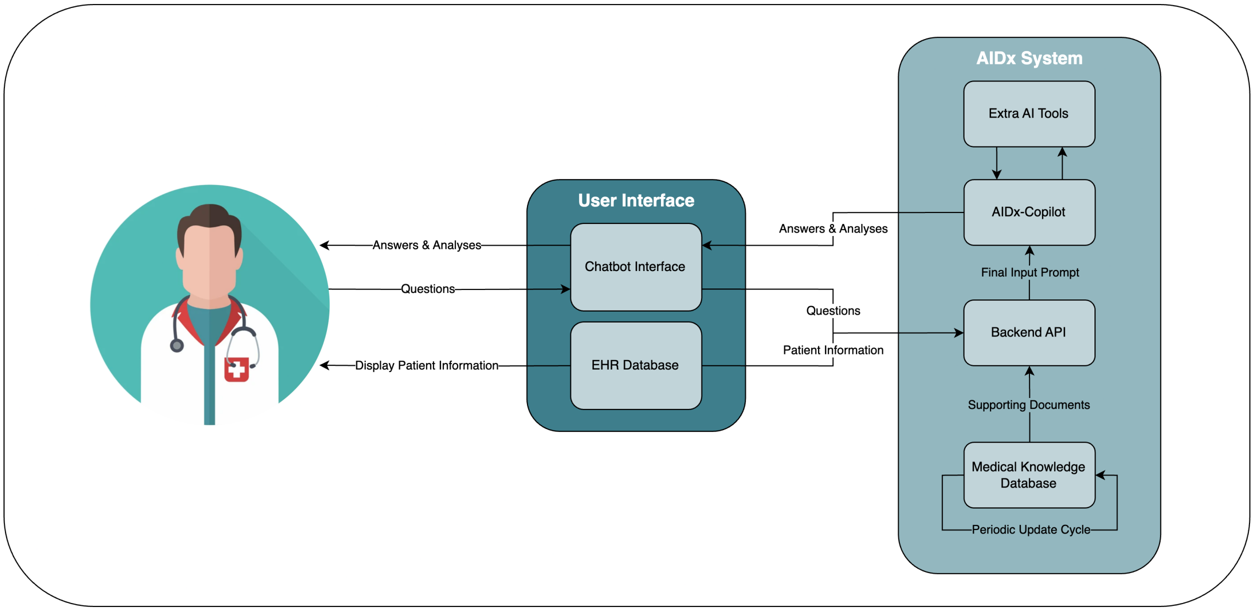


Figure 2: Workflow of AIDx, illustrating its integration atop an existing EHR database to augment physician inquiries with comprehensive patient information and pertinent medical knowledge before processing and generating responses.

Timeline Creation To accurately model the progression of a patient's hospital stay, I developed a timeline using the dynamic elements of the patient chart. Each change in a dynamic category prompted the creation of a new timestamp, encapsulating all patient data up to that point. Static information was consistently included as a header section in this timeline, ensuring a comprehensive patient snapshot at each timestamp.

Additional Patient Data Additional data, including radiology and discharge notes, were integrated following the established timeline methodology, enhancing the depth of the patient charts.

Chat Sample Generation I generated chat samples from these enriched patient charts to simulate potential physician inquiries. Five dynamic categories (labs, microbiology tests, provider orders, procedures, and diagnoses) were selected for generating predictive questions, with the subsequent timestamp's data serving as the answers. This process resulted in nearly 8 million distinct chat samples.

Model Training The development of AIDx-Copilot involved fine-tuning Mixtral-8x7B-Instruct-v0.1, a 46.7 billion parameter mixture of experts model, which compares favorably with industry benchmarks such as OpenAI's GPT-3.5 and Meta's LLaMA 2 in terms of both accuracy and performance [4]. The fine-tuning process, aimed at adapting Mixtral-8x7B for assisting physicians with clinical decision-making, utilized the Low-Rank Adaptation (LoRA) approach [7] in conjunction with the DeepSpeed Zero optimization techniques [14, 13], facilitated by the Huggingface Transformers library [20]. This intensive computational task was executed over three days using eight NVIDIA A100 graphics processing units (GPUs). Following the training phase, I merged the LoRA-adapted AIDx-Copilot model with its foundational Mixtral model. Subsequently, I quantized it using exllamav2 [18] to enhance operational efficiency and enable rapid inference capabilities. The deployment of AIDx-Copilot was achieved through integration with an OpenAI-Compatible API, facilitated by the tabbyAPI framework [17], which is specifically designed for the deployment of exllamav2 models in a production environment.

2.2 Implementation of AIDx (Figure 2)

2.2.1 Backend AIDx System

Retrieval Augmented Generation In the medical domain, where the accuracy and recentness of information can significantly impact patient outcomes, it is imperative for AI models to access and retrieve current information. AIDx incorporates Retrieval Augmented Generation (RAG) [10], a method enabling LLMs to integrate updated information without necessitating model retraining. RAG operates through a dynamic database of knowledge that can be continually updated, ensuring that information pertinent to the query is accessible to the model. For AIDx, this knowledge base comprises a collection of medical textbooks from the LibreTexts Medicine Library [2], integrated via the following steps using LangChain [1]:

1. **Loading**: Utilizing LangChain's PDFLoader, the medical textbooks are processed, extracting textual content and organizing it by pages.
2. **Chunking**: Semantic Chunking is applied to the extracted content, grouping related text segments together for coherence.
3. **Vectorization**: These text segments are transformed into vector embeddings using OpenAI's `text-embedding-3-small` model, facilitating their later retrieval.
4. **Storage**: The embeddings are stored in a Pinecone vector database [12], optimized for rapid and efficient retrieval.

When deploying RAG, the system encodes the physician's query into a vector, retrieves relevant information chunks from the database, and integrates this information into the model's input prompt, ensuring AIDx remains current and factually accurate.

Backend API The backend API serves as the backbone of AIDx, integrating various components—EHR data, RAG, prompting strategies, and auxiliary AI tools—into a streamlined experience for the end-user. The process is initiated with a physician's query and patient identifier, progressing through several steps to formulate a comprehensive input prompt and output response:

1. **Patient Retrieval**: Retrieves and converts the patient's chart into a textual format, prefixed to the query.
2. **RAG**: Enhances the prompt with relevant information fetched through RAG.
3. **Prompting Techniques**: Incorporates system instructions to guide AIDx-Copilot's response generation, leveraging Chain-of-Thought processing for nuanced analyses.
4. **Model Inference**: The enriched prompt is processed through AIDx-Copilot, enabling the model to either respond directly or engage additional AI tools for a more in-depth analysis.
5. **Extra AI Tools**: Facilitates the integration of specialized medical AI models, utilizing their output to augment the system's final response, thereby abstracting complex tool interactions for the physician.

This multi-tiered approach significantly mitigates the risk of generating erroneous or irrelevant content (hallucinations), a big concern to modern LLMs, especially in healthcare [3], ensuring that AIDx's outputs are both accurate and contextually relevant.

2.2.2 Frontend User Interface

The frontend design of AIDx prioritizes simplicity and seamless integration into existing medical workflows, addressing common barriers associated with the adoption of AI technologies in healthcare. By delegating complex tasks to the backend, the frontend maintains a minimalistic design, offering physicians a straightforward chat interface. This design philosophy not only enhances user engagement but also ensures that AIDx can be readily incorporated into various healthcare settings, thereby expanding its utility and accessibility.

3 Evaluation and Results

To rigorously assess the performance of AIDx-Copilot, I conducted a comprehensive evaluation using the MultiMedQA dataset, a benchmark that amalgamates various medical question-answer datasets, including the renowned USMLE, into a bank of approximately 7,000 multiple-choice questions [15]. This dataset is widely recognized in the field as a critical benchmark for evaluating the efficacy of medical language models. The choice of MultiMedQA as a testing ground is pivotal, as it offers a diverse array of questions that simulate real-world medical scenarios, enabling a thorough examination of AIDx-Copilot's diagnostic accuracy, decision-making capabilities, and alignment with current medical knowledge standards.

3.1 Overall Performance

As illustrated in Figure 3, AIDx-Copilot showcases impressive performance, achieving a mean accuracy of 83.61% with a standard deviation of 7.37, closely rivaling Google's Med-PaLM 2, a leader in the field with a mean accuracy of

Table 1: Model Parameter Count (Size Metric)

Model	Number of Parameters (Billions)
AIDx-Copilot	46.7
Med-PaLM 2	340.0
GPT-4	1700.0
GPT-3.5	175.0

86.66%, underscoring AIDx's competitive edge in the realm of medical AI. GPT-4 exhibits comparable performance with a mean accuracy of 83.69% but a higher standard deviation of 8.47, indicating greater variability across different evaluations. GPT-3.5 lags with a mean accuracy of 63.59% and the highest standard deviation of 9.13, reflecting its relatively inconsistent and lower performance on medical benchmarks.

In specific benchmarks such as MMLU Clinical Knowledge, MMLU Medical Genetics, and MMLU Professional Medicine, AIDx-Copilot's scores are particularly strong, approaching or exceeding the 90% threshold. In MMLU Professional Medicine, AIDx-Copilot excels with an impressive 93.4%, showcasing its near-parity with Med-PaLM 2's 95.2%, a testament to its sophisticated understanding of professional medicine.

AIDx-Copilot excels in domains necessitating comprehensive clinical knowledge and a deep understanding of professional medicine, as indicated by its high scores in these segments. Its competencies are further highlighted by its performances in MMLU College Biology (90.5%) and MedQA (USMLE) (84.6%), showcasing its adeptness in spanning a wide array of medical knowledge areas, from foundational biology to intricate clinical scenarios.

The relatively low standard deviation (7.37) across AIDx-Copilot's performance metrics underscores its consistency across various medical knowledge assessments, an attribute that is critical in the high-stakes environment of medical decision-making.

3.2 Overall Efficiency

In assessing the efficacy of AI models within medical question-answering contexts, a discernible pattern emerges regarding the relationship between model size (parameter count) and accuracy. Typically, as a model's size expands, its potential for accuracy increases, albeit with significant trade-offs in inference speed and computational demands. Here, efficiency is conceptualized as the ratio of accuracy to the number of parameters (in billions), offering a metric to gauge a model's proficiency in harnessing computational power for precise outcomes in medical inquiries.

Referencing Table 1 for the models' parameter counts and Figure 4 for their corresponding efficiency scores, AIDx-Copilot stands out as the most efficient model in the cohort ($p < 0.0001$), with a stellar mean efficiency score of 1.79, emphasizing its optimal use of computational resources for superior performance. This indicates that AIDx-Copilot, with its 46.7 billion parameters, excels in optimizing its computational size for heightened accuracy, markedly surpassing its counterparts. In contrast, Med-PaLM 2 and GPT-3.5 exhibit mean efficiencies of 0.25 and 0.36, respectively, suggesting a less optimal parameter utilization. GPT-4, despite its extensive size of 1700.0 billion parameters, registers the lowest efficiency at 0.05, highlighting the challenges large models face in translating vast parameter counts into enhanced task-specific accuracy.

Examining the models across various benchmarks reveals AIDx-Copilot's consistent superiority in efficiency, with notable achievements in MMLU Professional Medicine (2.00) and MMLU College Biology (1.94). These figures underscore AIDx-Copilot's adeptness in deploying its parameters judiciously across diverse medical domains. While Med-PaLM 2 and GPT-3.5 both demonstrate more moderate efficiencies, GPT-3.5 occasionally outperforms Med-PaLM 2, signaling potential resource efficiency in certain scenarios. Conversely, GPT-4's efficiency scores are uniformly low, illustrating the diminishing returns in accuracy for models surpassing a particular size threshold.

The standard deviation in efficiency scores sheds further light on these observations. AIDx-Copilot's standard deviation of 0.158 signifies consistent efficiency across various medical tasks, emphasizing its reliability. Meanwhile, GPT-4's low standard deviation (0.005) in efficiency scores across all benchmarks underscores its pervasive inefficiency, irrespective of the medical domain.

4 Discussion and Future Research

AIDx represents a significant advance in medical informatics, signaling a shift toward more efficient and accurate clinical decision-making tools. Its ability to rival and even surpass the performance of existing large language models

Figure 3: MultiMedQA Performance (Percent Accuracy, Higher is Better), Med-PaLM 2 Results Sourced From [16]

Dataset	AIDx-Copilot	Med-PaLM	GPT-4	GPT-3.5
MedQA (USMLE)	84.60	**86.50**	81.40	50.82
PubMedQA	79.40	**81.80**	75.20	71.60
MedMCQA	70.70	72.30	**72.40**	50.08
MMLU Clinical Knowledge	**90.00**	88.70	86.40	69.81
MMLU Medical Genetics	87.00	**92.00**	**92.00**	70.00
MMLU Anatomy	78.10	**84.40**	80.00	56.30
MMLU Professional Medicine	93.40	**95.20**	93.80	70.22
MMLU College Biology	90.50	**95.80**	95.10	72.22
MMLU College Medicine	78.80	**83.20**	76.90	61.27
Mean	83.61	**86.66**	83.69	63.59
Standard Deviation	7.37	7.38	8.47	9.13

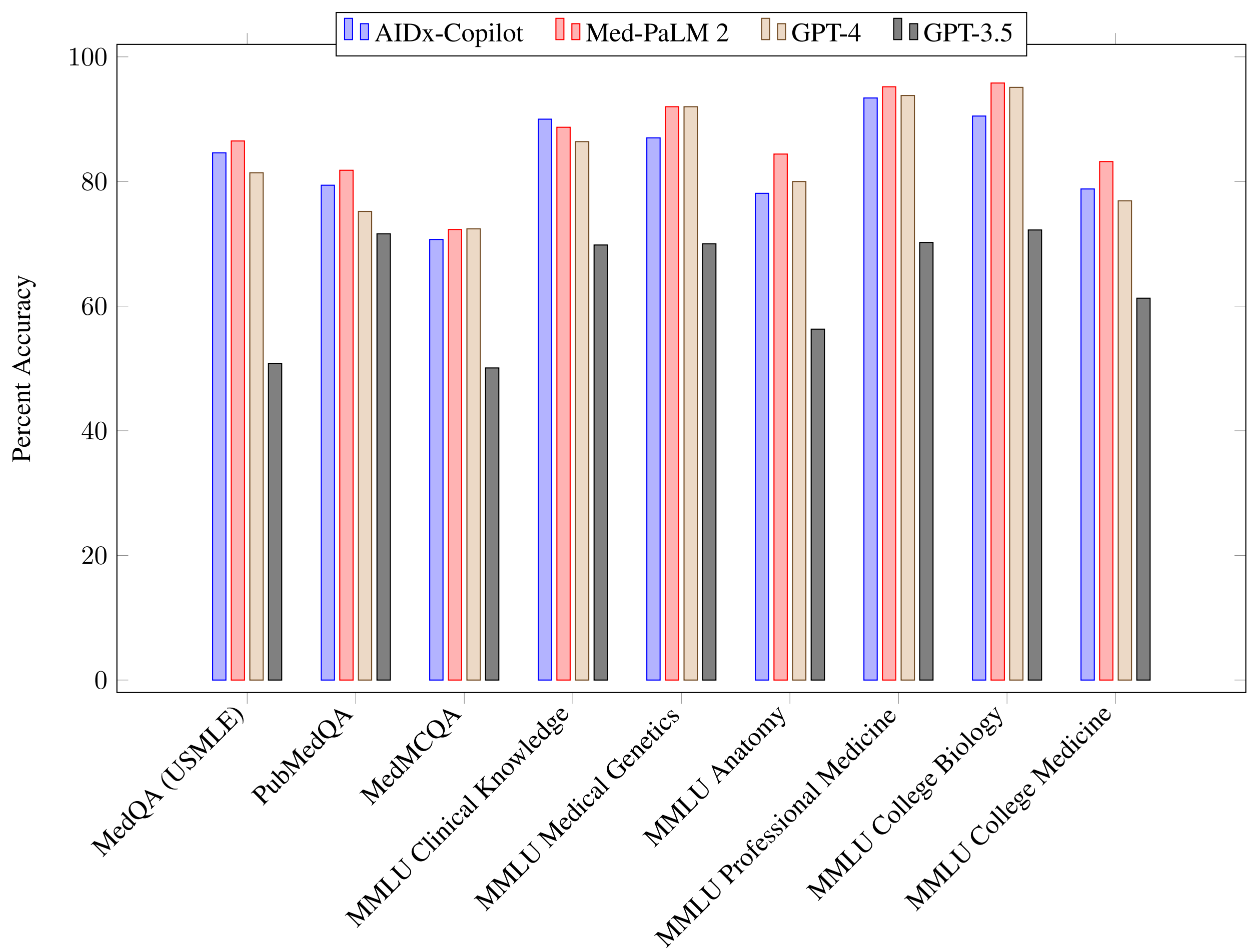
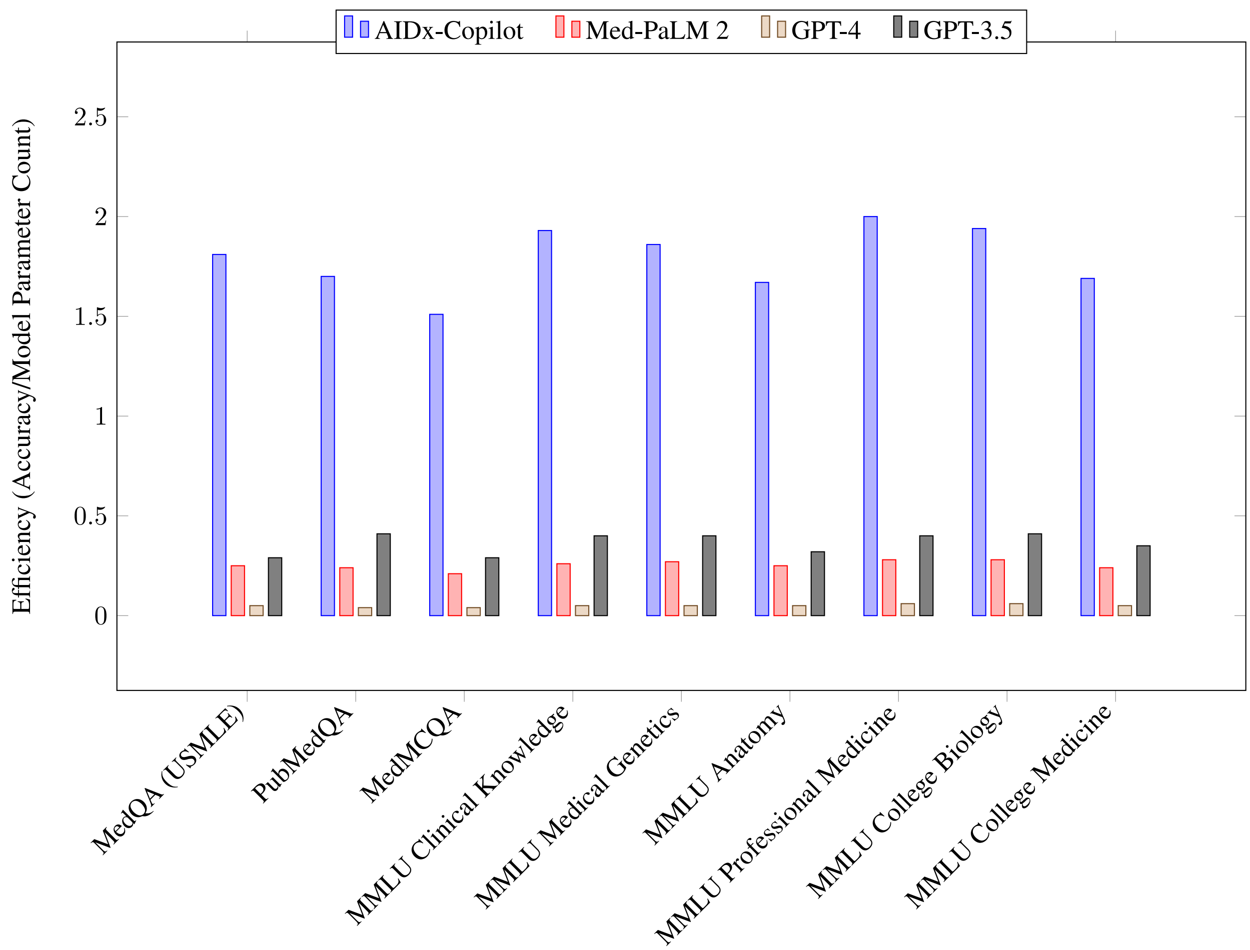


Figure 4: MultiMedQA Efficiency ($\frac{\text{Accuracy}}{\text{Model Parameter Count}}$, Higher is Better)

Dataset	AIDx-Copilot	Med-PaLM 2	GPT-4	GPT-3.5
MedQA (USMLE)	**1.81**	0.25	0.05	0.29
PubMedQA	**1.70**	0.24	0.04	0.41
MedMCQA	**1.51**	0.21	0.04	0.29
MMLU Clinical Knowledge	**1.93**	0.26	0.05	0.40
MMLU Medical Genetics	**1.86**	0.27	0.05	0.40
MMLU Anatomy	**1.67**	0.25	0.05	0.32
MMLU Professional Medicine	**2.00**	0.28	0.06	0.40
MMLU College Biology	**1.94**	0.28	0.06	0.41
MMLU College Medicine	**1.69**	0.24	0.05	0.35
Mean	**1.79**	0.25	0.05	0.36
Standard Deviation	0.158	0.022	0.005	0.052



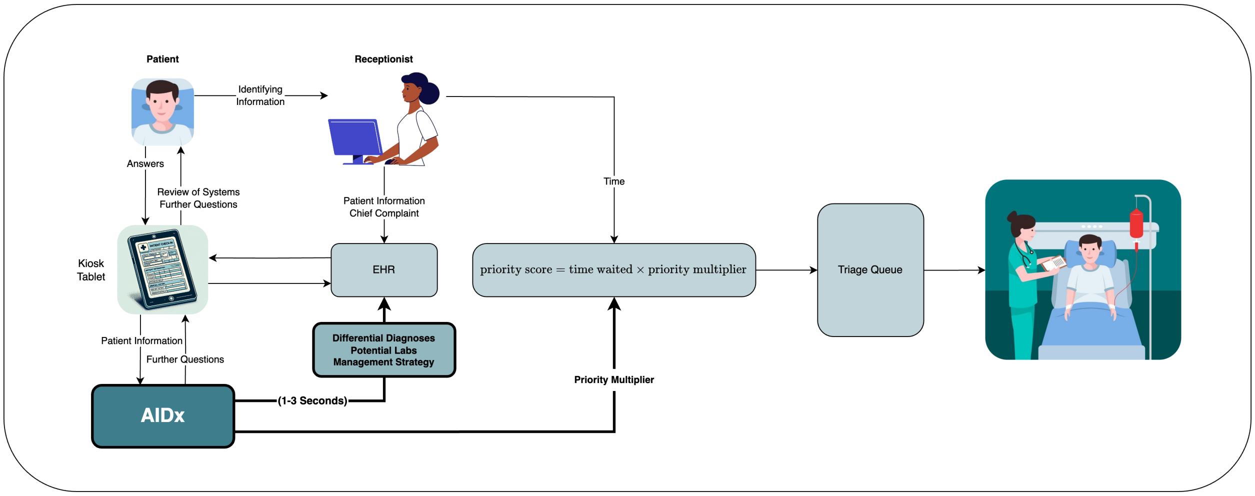


Figure 5: Application of AIDx in automating triage and enhancing patient prioritization in emergency departments

in healthcare, while maintaining a relatively small model size, establishes a new standard for efficiency in the field, translating into fast, accurate, and affordable implementation. This achievement transcends technical success, heralding a future where advanced clinical decision-support tools are universally accessible across all healthcare environments.

The implementation of AIDx has the potential to transform patient care by providing personalized treatment recommendations and improving outcomes through data-driven decision support. Its compatibility with existing EHR systems ensures that AI-enhanced insights are seamlessly integrated with patients' historical data, reducing human error and offering a dependable "second opinion" quickly and efficiently. The minimal training requirement for physicians highlights AIDx's practicality, facilitating its integration into existing clinical workflows.

AIDx's utility extends across various medical contexts, particularly in emergency departments where timely decision-making is critical. Here, AIDx could expedite diagnosis and treatment, thereby improving patient care and outcomes. The system's ability to learn and adapt from ongoing patient data introduces a new dimension of personalized healthcare, where treatments are expertly tailored to individual patient situations. Furthermore, its cost-effectiveness and operational efficiency make AIDx an attractive solution for resource-constrained environments, potentially universalizing access to quality healthcare. The model's capacity to translate complex medical terminology into language understandable by patients also enhances communication and patient engagement in their care.

In emergency departments, AIDx's integration could revolutionize triage and patient prioritization (Figure 5). Initially, upon a patient's arrival, basic information and chief complaints are logged in their chart. Subsequently, patients complete a review of systems questionnaire, aiding in initial diagnosis. Before the nurse evaluation, AIDx analyzes this data to propose additional targeted questions, identify differential diagnoses, suggest potential lab tests, and offer preliminary management strategies, all of which are written into the patient's chart. It also assigns a priority multiplier, optimizing a smart triage queue that balances urgency and waiting time, ensuring timely attention to critical cases.

The next steps for AIDx involve rigorous validation through controlled clinical trials to assess its impact on patient outcomes and clinician workflows in a real hospital environment. Assessing its generalizability across different medical institutions and patient populations will be crucial to understanding its broader applicability. Long-term studies on the efficiency and cost savings brought about by AIDx's use in healthcare systems will provide insights into its economic viability. Data security and patient privacy in EHR integration remain paramount, necessitating thorough investigation. Moreover, evolving AIDx into a multimodal and modular system that can directly analyze medical imaging and modularly integrate other specialized AI models in its patient processing could open new frontiers in AI-powered diagnostics, making it a truly holistic clinical support tool.

Future research directions for AIDx include comprehensive clinical trials to validate its effectiveness and impact on patient outcomes and healthcare workflows in real hospital environments. Understanding its adaptability across various medical settings and patient demographics is essential for assessing its widespread applicability. Longitudinal studies will be invaluable in evaluating the system's long-term efficiency and economic benefits. Ensuring data security and maintaining patient confidentiality in EHR integration is also imperative, necessitating continuous development in AIDx's implementation. Future enhancements could involve expanding AIDx into a multimodal system capable of directly analyzing medical imaging and other types of medical data, along with developing more specialized AI model tools, thus broadening its diagnostic and decision-support capabilities.

References

[1] Langchain-ai/langchain: Build context-aware reasoning applications. https://github.com/langchain-ai/langchain/tree/master.

[2] LibreTexts Medicine Library. https://med.libretexts.org, February 2016.

[3] Muhammad Aurangzeb Ahmad, Ilker Yaramis, and Taposh Dutta Roy. Creating Trustworthy LLMs: Dealing with Hallucinations in Healthcare AI, September 2023.

[4] Mistral AI. Mixtral of experts. https://mistral.ai/news/mixtral-of-experts/, December 2023.

[5] CDC. CDC FastStats - Emergency Department Visits. https://www.cdc.gov/nchs/fastats/emergency-department.htm, November 2023.

[6] Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, 101(23):e215–e220, June 2000.

[7] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-Rank Adaptation of Large Language Models, October 2021.

[8] Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. MIMIC-IV.

[9] Alistair E. W. Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J. Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, Li-wei H. Lehman, Leo A. Celi, and Roger G. Mark. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10(1):1, January 2023.

[10] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, April 2021.

[11] OpenAI. Introducing ChatGPT. https://openai.com/blog/chatgpt, November 2022.

[12] Pinecone. The vector database to build knowledgeable AI | Pinecone. https://www.pinecone.io/.

[13] Samyam Rajbhandari, Jeff Rasley, Olatunji Ruwase, and Yuxiong He. ZeRO: Memory optimizations toward training trillion parameter models. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '20, pages 1–16, Atlanta, Georgia, November 2020. IEEE Press.

[14] Jeff Rasley, Samyam Rajbhandari, Olatunji Ruwase, and Yuxiong He. DeepSpeed: System Optimizations Enable Training Deep Learning Models with Over 100 Billion Parameters. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, pages 3505–3506, New York, NY, USA, August 2020. Association for Computing Machinery.

[15] Karan Singhal, Shekoofeh Azizi, Tao Tu, S. Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, Perry Payne, Martin Seneviratne, Paul Gamble, Chris Kelly, Nathaneal Scharli, Aakanksha Chowdhery, Philip Mansfield, Blaise Aguera y Arcas, Dale Webster, Greg S. Corrado, Yossi Matias, Katherine Chou, Juraj Gottweis, Nenad Tomasev, Yun Liu, Alvin Rajkomar, Joelle Barral, Christopher Semturs, Alan Karthikesalingam, and Vivek Natarajan. Large Language Models Encode Clinical Knowledge, December 2022.

[16] Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Le Hou, Kevin Clark, Stephen Pfohl, Heather Cole-Lewis, Darlene Neal, Mike Schaekermann, Amy Wang, Mohamed Amin, Sami Lachgar, Philip Mansfield, Sushant Prakash, Bradley Green, Ewa Dominowska, Blaise Aguera y Arcas, Nenad Tomasev, Yun Liu, Renee Wong, Christopher Semturs, S. Sara Mahdavi, Joelle Barral, Dale Webster, Greg S. Corrado, Yossi Matias, Shekoofeh Azizi, Alan Karthikesalingam, and Vivek Natarajan. Towards Expert-Level Medical Question Answering with Large Language Models, May 2023.

[17] Theroyallab. Theroyallab/tabbyAPI. The Royal Lab, March 2024.

[18] turboderp. Turboderp/exllamav2, March 2024.

[19] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, January 2023.

[20] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. HuggingFace's Transformers: State-of-the-art Natural Language Processing, July 2020.