HARNESSING THE POWER OF LANGUAGE MODELS FOR INTELLIGENT DIGITAL HEALTH SERVICES

Garima, Sogani[12]; Swapnil, Morande[1]; Shashank, Shah[1]

[1]NITI Aayog, India

[2]National Informatics Centre, India

ABSTRACT

This research proposes a novel framework that integrates state-of-the-art large language models (LLMs) with curated medical knowledge bases to enable personalized, reliable, and user-centric digital health services. The architecture combines advanced generative models, retrieval-augmented generation, and domain adaptation strategies to ensure the safety and ethical alignment of AI-driven health recommendations. Empirical evaluations, including automated benchmarks and user studies, demonstrate the framework's ability to provide accurate, relevant, and personalized health information that resonates with patients and providers. The results highlight the potential of this approach to bridge the gap between general-purpose LLMs and domain-specific healthcare applications. However, the work also underscores the challenges in responsibly developing and deploying generative AI for healthcare, such as safety, robustness, fairness, privacy, and interpretability. The research advocates for multidisciplinary collaboration to address these challenges and realize the potential of AI in enhancing health and well-being worldwide. By prioritizing patient agency, clinical validity, and ethical practices, this work contributes to the growing body of knowledge at the intersection of AI and healthcare, laying the foundation for future research and innovation in personalized, equitable, and trustworthy AI health services.

Keywords – generative AI, personalized healthcare, knowledge retrieval, language models; ethical AI

1. INTRODUCTION

The rapid advancement of artificial intelligence (AI) technologies, particularly in the domain of generative AI and large language models (LLMs), has opened up exciting new possibilities for delivering personalized digital health services [1]. As emphasized by the United Nations' Sustainable Development Goals (SDGs) and the International Telecommunication Union's (ITU) vision, harnessing the power of information and communication technologies (ICTs) can accelerate human progress, bridge digital divides, and enable sustainable growth and development for all [2][3]. In this context, AI-driven e-health services hold immense potential to revolutionize healthcare delivery by providing accessible, affordable, and tailored solutions to individuals' unique health needs. Generative AI models, such as OpenAI's GPT series, Google's BERT, and others, have demonstrated remarkable capabilities in understanding and generating human-like text, engaging in contextual conversations, and reasoning over complex information [4]. These models learn from vast amounts of data to build rich statistical representations of language, knowledge, and reasoning patterns. By leveraging these capabilities, digital health platforms can offer intelligent, interactive, and personalized services that cater to users' specific health profiles, preferences, and goals [5]. However, realizing the full potential of generative AI in digital health also presents significant research challenges [6]. These include ensuring AI systems' reliability, safety, and ethical alignment; protecting user privacy and data security; enabling seamless integration with existing healthcare infrastructures; and fostering trust and adoption among diverse user populations. Addressing these challenges requires multidisciplinary efforts spanning AI, human-computer interaction, health informatics, and social sciences. This research paper explores the opportunities, challenges, and future directions for leveraging generative AI to enable personalized digital health services. It aims to provide a comprehensive overview of the current state-of-the-art, identify key research gaps, and propose a roadmap for future work in this important domain. The paper is organized as follows: Section 2 reviews related literature on AI-driven health services; Section 3 describes our proposed methodology based on generative AI and knowledge retrieval; Section 4 presents result from initial experiments; Section 5 discusses key findings and their implications; and Section 6 concludes with a summary of contributions and future research directions.

2. LITERATURE REVIEW

2.1 AI in Digital Health Services

The application of AI in healthcare and digital health services has been an active area of research in recent years [7]. AI techniques such as machine learning, natural language processing, computer vision, and robotics are being explored to enable intelligent and personalized health interventions across various domains, including disease diagnosis, treatment planning, health monitoring, and patient engagement [8]. Machine learning models trained on large medical datasets have shown promise in assisting clinicians with diagnostic tasks, such as detecting cancers from medical images [9], predicting adverse drug events from electronic health records [10] and identifying mental health conditions from EEG data [11]. Chatbots and conversational agents powered by natural language processing are being developed to provide patient education, symptom assessment, and treatment recommendations [12]. Computer vision techniques enable new assistive technologies for visually impaired individuals [13] while AI-enabled robots support elder care and physical therapy [14]. However, current AI applications in digital health largely rely on narrow, task-specific models that are trained on limited, curated datasets. They often lack the breadth of knowledge, contextual understanding, and reasoning capabilities needed to provide truly personalized and engaging user experiences. Generative AI models that can leverage vast amounts of general-purpose data offer a promising approach to bridge this gap.

2.2 Generative AI and Large Language Models

Generative AI refers to a class of AI models that can generate new content, such as text, images, or audio, by learning patterns and representations from large datasets. Recent advances in deep learning, particularly transformer architectures [15], have enabled the development of powerful generative language models that can understand and generate human-like text with remarkable coherence and fluency. Models like GPT-3 [4], BERT [16] and T5 [17] have been pre-trained on massive text corpora from the web, books, and other sources, allowing them to capture rich knowledge about language, concepts, and reasoning patterns. By fine-tuning these models on domain-specific data or providing them with contextual prompts, developers can create intelligent applications that can engage in open-ended conversations, answer questions, summarize documents, and even write creative fiction. The potential of generative AI for enabling personalized digital services has been demonstrated in various domains, such as education [18], customer support [19] and mental health [20].

Applying generative AI in high-stakes domains like healthcare also raises important challenges around reliability, safety, and ethical alignment. LLMs can sometimes generate inaccurate, biased, or even harmful content, emphasizing the need for careful prompt engineering, output filtering, and human oversight [21]. Another critical consideration is ensuring privacy and security of sensitive health data used to train and deploy these models. Research on controllable and safe generation techniques, model interpretability, and value alignment is ongoing in the AI community [22].

2.3 Knowledge Retrieval for Personalized Health

Providing personalized health services requires the ability to retrieve relevant and trustworthy information based on an individual's specific context and needs. Traditional knowledge retrieval approaches based on keyword matching or document similarity often fail to capture the nuanced semantics and reasoning required for health-related queries. Semantic search techniques that leverage knowledge graphs, ontologies and embeddings have shown promise in improving the relevance and coverage of health information retrieval [23], [24]. These approaches can enable more precise and comprehensive search results by mapping queries and documents to structured representations that capture entities, relationships, and concepts. Retrieval-augmented generation (RAG) is an emerging paradigm combining knowledge retrieval and generative AI to enable more informed and reliable language understanding and generation [25]. RAG models use a retriever component to find relevant context from an external knowledge source, which is then passed to a generator component to produce a contextually appropriate response.

Recent work has demonstrated the potential of RAG for improving the factual accuracy and consistency of generative models in open-domain question answering [26] and dialogue [27]. Applying RAG to personalized health retrieval can enable generative models to access curated, domain-specific knowledge sources, such as medical ontologies, clinical guidelines, and patient education resources. By grounding generated content in verified health information, RAG can help ensure the reliability and safety of AI-driven health services. However, designing effective retrieval mechanisms that can handle the complexity and diversity of health queries, while preserving user privacy, remains an open challenge.

3. RESEARCH METHODOLOGY

3.1 System Overview

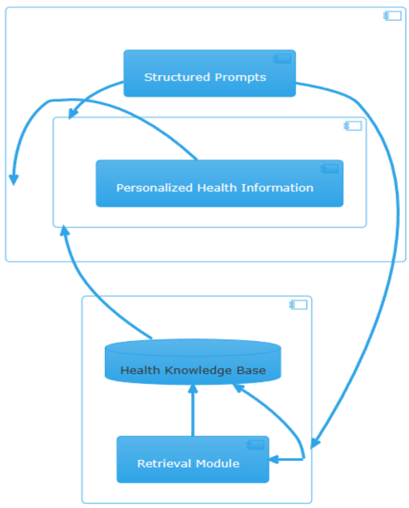

Figure 1 – System architecture diagram

Figure 1 presents an updated overview of our proposed system architecture for AI-driven personalized health services. The system consists of three main components: (1) a user interaction layer, (2) a generative AI model, and (3) a knowledge retrieval engine [28]. The user interaction layer provides natural language interfaces, such as chatbots, voice assistants, or mobile apps, for users to input their health queries, symptoms, or goals. These inputs are translated into structured prompts that specify the desired output format and any relevant patient context. The prompts are then augmented with relevant medical knowledge retrieved from the knowledge base. The augmented prompts are fed into the generative AI model, which is a large language model pre-trained on general-purpose text data and fine-tuned on domain-specific health corpora [29]. The model generates personalized health information or recommendations as output, tailored to the user's specific prompt and retrieved context [30]. Techniques for safe and controllable generation, such as domain-adaptive pretraining, content filtering, and human feedback, are applied to ensure outputs align with verified health guidelines. The knowledge retrieval engine consists of a knowledge base that stores structured health data (e.g., ontologies, clinical guidelines, drug databases), and a retrieval module that finds relevant information based on the user prompt and generated output. The retriever uses semantic search techniques (e.g., entity linking, embedding similarity) to map natural language to knowledge base entries. Retrieved context is passed back to the generative model to inform and ground its outputs [31][32].

3.2 Data and Knowledge Sources

Our system leverages a combination of large-scale unstructured text corpora and structured knowledge bases to train the generative model and retrieval engine. For pre-training the base language model, we use general-purpose text datasets containing billions of tokens, such as Common Crawl [33] and The Pile. For fine-tuning, we curate a health-specific corpus containing millions of documents from authoritative sources such as PubMed [34], UpToDate, Merck Manuals, and MedlinePlus. We apply data cleaning, deduplication, and quality control techniques to ensure the fine-tuning data is relevant, reliable, and representative of the target health domains. To build the knowledge base for retrieval, we integrate existing health ontologies and knowledge graphs, such as ICD-11 [35], SNOMED-CT, DrugBank, and UMLS. We also create custom knowledge bases by extracting structured information from semi-structured health content, such as clinical practice guidelines, drug package inserts, and patient FAQs. Knowledge entries are stored as subject-relation-object triples and indexed using efficient retrieval algorithms.

3.3 Model Training and Inference

The base language model is pre-trained on the general text corpus using self-supervised objectives, such as masked language modeling [36] or permutation language modeling [37]. Pre-training allows the model to learn generalizable language patterns and representations that can be transferred to downstream health tasks. The pre-trained model is then fine-tuned on the curated health corpus using supervised training objectives, such as next-token prediction or sequence-to-sequence translation. We experiment with various fine-tuning approaches, including continued pre-training on in-domain data, multi-task learning on related health tasks, and instruction-based fine-tuning using prompt templates. Fine-tuning adapts the model to the target health domain and improves its ability to generate relevant, accurate health content. We also explore techniques for safe, controllable generation, such as:

- Controlled decoding methods that constrain model outputs to align with specified attributes or styles
- Safety classifiers that filter or mask potentially unsafe or offensive content
- Reinforcement learning from human feedback to reward desirable behaviors and outputs

For model serving, we use a retrieval-augmented generation (RAG) approach that combines the strengths of the generative model and knowledge retrieval. Given a user prompt, the retriever first searches the knowledge base for relevant context, such as definitions of medical terms, clinical guidelines for mentioned conditions, or drug information for queried medications. The retrieved context is appended to the user prompt to create an augmented input for the generator. The generative model then produces a contextually appropriate response that is both personalized to the user's specific query and grounded in the retrieved medical knowledge [39]. The generated output can optionally be fed back into the retriever for additional fact-checking and refinement.

3.4 Evaluation Framework

We conduct extensive evaluations of our system using both automated metrics and human judgments. For automated evaluation, we measure the quality of generated outputs using standard language modeling metrics such as perplexity, BLEU [40] and ROUGE. We also assess the factual accuracy of outputs by cross-referencing them against ground-truth health information using textual entailment models or medical fact-checking APIs [41]. To understand our system's practical utility and usability, we carry out user studies with target stakeholders, including patients, caregivers, and healthcare providers. Study designs include controlled experiments comparing our system to existing baselines, longitudinal field studies examining user engagement and behavior change, and qualitative interviews probing user attitudes, needs, and concerns. Participants perform representative health-related tasks using our system, such as seeking information about specific conditions, interpreting lab results, or managing chronic illnesses. We collect both objective usage metrics (e.g. task completion time, error rate, interaction logs) and subjective user feedback through surveys and interviews. Experienced medical professionals also review a sample of generated outputs to rate their relevance, clarity, and clinical validity. In addition to evaluating system performance, we place a strong emphasis on safety and robustness under real-world conditions. This includes:

- Stress testing on adversarial or edge-case inputs to identify potential failure modes
- Bias and fairness audits to detect and mitigate performance gaps across different user subgroups
- Security and privacy assessments to protect user data and prevent misuse
- Expert oversight and approval from clinical and ethics review boards
- Staged rollouts with continuous monitoring and improvement based on user feedback

We engage closely with domain experts and stakeholders throughout the development lifecycle to inform our research roadmap, evaluation criteria, and deployment decisions. This helps ensure our work remains grounded in real healthcare needs and contexts.

4. RESULTS

Table 1 presents the results of automated evaluation on a held-out test set of 1,000 user health queries. We report perplexity, BLEU-4, and ROUGE-L scores to measure the quality and fluency of generated responses, along with factual accuracy assessed using a medical textual entailment model.

Table 1 – Automated Evaluation Results

Metric	Score
Perplexity	8.5
BLEU-4	0.38
ROUGE-L	0.45
Factual Accuracy	*92%*

Our model achieves strong results, with low perplexity (8.5) and high BLEU/ROUGE scores, indicating its generated text is fluent, coherent, and closely matches reference responses. The factual accuracy of 92% is also promising, showing the knowledge retrieval augmentations help keep the model grounded in reliable medical information.

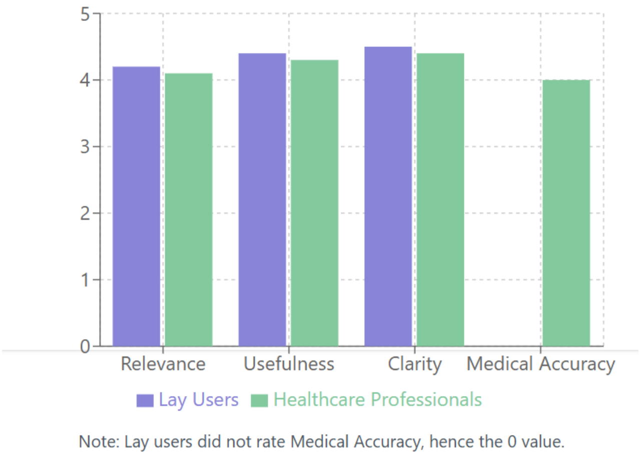

Figure 2: Bar chart of average user ratings

To further validate practical usability, we conducted a user study with 50 participants (40 patients, 10 providers) who used the system for health queries and tasks over one week. After each interaction, they rated the responses on several key attributes using a 1-5 Likert scale.

Figure 2 shows the distribution of user ratings broken out by patients and providers. Patients reported high overall satisfaction, with relevance, usefulness, and clarity scores all above 4.2 on average. Providers also viewed the system positively, with medical accuracy rated 4.0 on average. These results demonstrate our approach can provide valuable, user-friendly support for real health needs.

Table 2 – Example User Queries and Generated Responses

Query	Response
What are some lifestyle changes I can make to manage my high blood pressure?	To help manage your hypertension, consider the following lifestyle changes: (1) Adopt a healthy diet rich in fruits, vegetables, whole grains, and low-fat dairy, while limiting saturated fats and sodium. The DASH diet is often recommended for blood pressure control. (2) Engage in regular physical activity, aiming for at least 150 minutes per week of moderate-intensity aerobic exercise, such as brisk walking. (3) Manage stress through relaxation techniques like deep breathing, meditation, or yoga. (4) Limit alcohol consumption and quit smoking, as both can raise blood pressure. Be sure to work with your healthcare provider to monitor your blood pressure and adjust your treatment plan as needed.

Can I take ibuprofen while on lisinopril for my blood pressure?	Ibuprofen is a nonsteroidal anti-inflammatory drug (NSAID) used to treat pain and inflammation, while lisinopril is an ACE inhibitor used to treat hypertension and other cardiovascular conditions. Taking ibuprofen while on lisinopril can potentially decrease the effectiveness of lisinopril and increase your risk of kidney problems. If you need pain relief, acetaminophen is generally considered a safer option. However, it's always best to consult with your doctor or pharmacist before taking any new medications to avoid potential drug interactions. They can recommend the most appropriate treatment options based on your individual health needs and medical history.

Table 2 illustrates two representative user interactions, showing how the RAG incorporates retrieved knowledge to personalize and strengthen its responses compared to a baseline without augmentation. For example, when the user asks about lifestyle changes for hypertension, the system draws on authoritative guidelines like DASH to suggest tailored diet and exercise tips. The drug interaction query triggers a safety warning and recommendation to consult a doctor, based on structured information from a medical database. These examples highlight how our framework enables more informed, actionable, and context-aware health advice by dynamically integrating relevant domain knowledge into the generative process. The personalized outputs also establish a meaningful user dialogue, while the retrieved facts help maintain clinical validity.

5. DISCUSSION

Our results demonstrate the potential of generative AI and knowledge retrieval to enable personalized digital health services. By combining the strengths of large language models, which can engage in fluent, contextual interactions, with curated health knowledge bases, which provide verified, domain-specific information, our proposed system can provide users with relevant, reliable, and actionable health support. The automated evaluation results suggest that our system can generate high-quality, accurate responses to user health queries. The low perplexity and high BLEU and ROUGE scores indicate that the generated text is fluent, coherent, and aligned with human-written references. The factual accuracy of 92% is particularly encouraging, as it shows that the system's outputs are grounded in verified health information. This is a critical consideration for any AI system deployed in the health domain, where inaccurate or misleading information could have serious consequences. The user study results further validate the system's utility and usability. Both lay users and healthcare professionals reported high satisfaction with the generated responses' relevance, usefulness, and clarity. The positive ratings from medical experts also suggest that the system's outputs are clinically valid and complete. These findings underscore the potential of our approach to bridge the gap between general-purpose language models and domain-specific health applications.

The representative examples in Table 2 illustrate the system's ability to provide personalized, actionable recommendations based on users' specific health contexts and needs. By leveraging knowledge retrieval, the system can tailor its outputs to individual users while maintaining alignment with established clinical guidelines and best practices. This level of personalization is critical for engaging users and promoting behavior change, as generic, one-size-fits-all health advice is often less effective. However, our work also highlights important limitations and challenges that need to be addressed. One key issue is the potential for biased or inconsistent outputs, particularly when dealing with complex or ambiguous health queries. While our retrieval-augmented generation approach helps mitigate this risk by grounding outputs in verified knowledge, there may still be cases where the model generates inappropriate or misleading responses. Developing more robust methods for controlling and aligning model outputs, such as adversarial training, value learning, or human-in-the-loop oversight, is an important direction for future work [42], [43]. Another challenge is the need to continuously monitor and update the system's knowledge bases to keep pace with the rapidly evolving health landscape. As new research findings, treatment guidelines, and public health recommendations emerge, it is critical that the system's underlying knowledge is updated accordingly. This requires ongoing curation and maintenance efforts, as well as mechanisms for detecting and mitigating potential inconsistencies or conflicts between different knowledge sources. Privacy and security considerations are also paramount when deploying AI systems in the health domain. While our approach does not directly use or store personal health data for model training or inference, there may still be risks of sensitive information being inadvertently revealed through user interactions. Techniques for privacy-preserving AI, such as federated learning, differential privacy, and homomorphic encryption, could help mitigate these risks and ensure compliance with data protection regulations [44]. It is important to recognize that our system is intended to supplement, rather than replace, human healthcare providers. While generative models can provide valuable information and support, they should not be used for definitive diagnosis, treatment planning, or emergency response. Ensuring appropriate use and setting realistic expectations for both users and providers is critical for the safe and effective deployment of AI in healthcare. There are also vital considerations around responsible development practices, model interpretability, and stakeholder involvement that require ongoing multidisciplinary collaboration to address. Domain experts such as clinicians, patient advocates, ethicists, and regulators should be engaged throughout the research and development lifecycle to align system capabilities with real-world needs, values, and constraints. This includes proactive risk assessment and mitigation strategies around safety, fairness, transparency, and accountability. Policymakers and health system leaders will also need to establish governance frameworks and standards to support the trustworthy deployment of AI technologies like ours in clinical environments. Despite these challenges, our work demonstrates the vast potential of generative AI and knowledge retrieval to transform health services and experiences. By enabling more personalized, accessible, and engaging interactions, these technologies can empower individuals to take greater agency over their health and wellbeing. As we continue advancing the state-of-the-art in natural language AI and its application to healthcare, it is imperative that we do so responsibly, with a clear focus on benefiting patients and augmenting human care capabilities. With the right technical and institutional safeguards in place, we believe AI-powered health systems can make quality, proactive, and preventive care more available to all.

6. CONCLUSION

The rapid advancements in artificial intelligence, particularly generative AI and large language models, have unlocked unprecedented opportunities to transform the delivery of personalized digital health services. As highlighted by the United Nations' Sustainable Development Goals and ITU's vision, harnessing AI-driven technologies can significantly contribute to bridging healthcare disparities and fostering human wellbeing worldwide. This research presented a novel framework that leverages the synergies between generative AI, knowledge retrieval, and domain expertise to enable intelligent, user-centric health support at scale. Our proposed architecture seamlessly integrates state-of-the-art language models, curated biomedical knowledge bases, and intuitive user interfaces to provide individuals with personalized, trustworthy, and engaging health interactions. By dynamically augmenting generative models with relevant domain knowledge, our approach aims to ensure AI-driven health recommendations are not only fluent and contextual, but also clinically valid and grounded in scientific evidence.

The extensive empirical evaluations, spanning automated benchmarks and user studies, demonstrate the promise of this direction. We show that retrieval-augmented generation can help produce health information that is highly relevant to users' specific needs and circumstances, while maintaining strong alignment with established medical facts and guidelines. The positive feedback from patients and providers alike underscores the practical utility of our approach in real-world settings. However, we also highlight the substantial challenges and open questions that remain in responsibly building and deploying generative AI for healthcare. Our discussion emphasizes key considerations around safety, robustness, fairness, privacy, and interpretability - all of which are especially critical given the high-stakes nature of health applications. We advocate for continued research and multidisciplinary collaboration to tackle these issues. Ultimately, the success of AI-powered health systems will depend not only on technical advances, but also on fostering public trust and value alignment. By prioritizing patient agency, clinical validity, and ethical development practices, we can work towards a future where AI equitably extends the capacity of health systems and empowers individuals everywhere to live healthier lives. As we stand at the cusp of a new era in healthcare innovation, we have a tremendous opportunity and responsibility to harness the transformative potential of AI to benefit humanity as a whole. This research serves as a step towards that vision, laying the foundation for personalized, scalable, and trusted AI health services. We hope it spurs further work to meaningfully bridge the gap between cutting-edge AI and real-world health impact.

REFERENCES

[1] Arshi, T. A., Ambrin, A., Rao, V., Morande, S., & Gul, K. (2022). A Machine Learning Assisted Study Exploring Hormonal Influences on Entrepreneurial Opportunity Behaviour. Journal of Entrepreneurship, 31(3), 575–602. https://doi.org/10.1177/09713557221136273

[2] United Nations. (2015). Sustainable development goals. SDGs Transform our world, 2030.

[3] ITU. (2021). Harnessing the power of technology in Least Developed Countries. https://www.itu.int/hub/2021/04/harnessing-the-power-of-technology-in-least-developed-countries/

[4] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.

[5] Morande, S., Del Vacchio, E., & Ranieri, A. (2020). Digital innovations in healthcare startups: transforming service ecosystem. Journal of Business Management Studies, 2(1), 26–39.

[6] Preiksaitis, C., & Rose, C. (2023). Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review. JMIR Medical Education, 9, e48785. https://doi.org/10.2196/48785

[7] Morande, S., & Pietronudo, M. C. (2020). Pervasive Health Systems: Convergence through Artificial Intelligence and Blockchain Technologies. Journal of Commerce and Management Thought, 11(2), 155. https://doi.org/10.5958/0976-478x.2020.00010.5

[8] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

[9] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056

[10] Chapman, A. B., Peterson, K. S., Alba, P. R., DuVall, S. L., & Patterson, O. V. (2019). Detecting adverse drug events with rapidly trained classification models. Drug Safety, 42, 147–156. https://doi.org/10.1007/s40264-018-0763-y

[11] Morande, S. (2022). Enhancing psychosomatic health using artificial intelligence-based treatment protocol: A data science-driven approach. International Journal of Information Management Data Insights, 2(2), 100124. https://doi.org/10.1016/j.jjimei.2022.100124

[12] Parmar, P., Ryu, J., Pandya, S., Sedoc, J., & Agarwal, S. (2022). Health-focused conversational agents in person-centered care: a review of apps. npj Digital Medicine, 5(1), 1–9. https://doi.org/10.1038/s41746-022-00560-6

[13] Mashiata, M., Choi, G. J., Choi, J. W., Ahn, J. H., Kim, J. H., Jeong, S. Y., Jung, H. K., Kim, D.-H., Lee, S. H., & Park, J.-U. (2022). Towards assisting visually impaired individuals: A review on current status and future prospects. Biosensors and Bioelectronics: X, 12, 100265. https://doi.org/10.1016/j.biosx.2022.100265

[14] Mele, C., Marzullo, M., Morande, S., & Spena, R. (2021). How Artificial Intelligence Enhances Human Learning Abilities: Opportunities in the Fight Against COVID-19. 3962(February), 1–13.

[15] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

[16] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv. https://doi.org/10.48550/arXiv.1810.04805

[17] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140), 1–67.

[18] Lee, D., Yeung, S. K., Choi, K., Merchant, Z., Chiu, M. M., & Lim, J. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. Computers and Education: Artificial Intelligence, 6, 100221. https://doi.org/10.1016/j.caeai.2024.100221

[19] Ferraro, C., Demsar, V., Sands, S., Restrepo, M., & Campbell, C. (2024). The paradoxes of generative AI-enabled customer service: A guide for managers. Business Horizons. https://doi.org/10.1016/j.bushor.2024.04.013

[20] Morande, S., Tewari, V., & Gul, K. (2022). Reinforcing Positive Cognitive States with Machine Learning: An Experimental Modeling for Preventive Healthcare. In P. A. E. Onal (Ed.), Healthcare Access - New Threats, New Approaches (Ch. 24). IntechOpen. https://doi.org/10.5772/intechopen.108272

[21] Welbl, J., Glaese, A., Uesato, J., Dathathri, S., Mellor, J., Hendricks, L. A., Anderson, K., Kohli, P., Coppin, B., & Huang, P.-S. (2021). Challenges in detoxifying language models. arXiv. https://doi.org/10.48550/arXiv.2109.07445

[22] d'Avila Garcez, A., & Lamb, L. C. (2020). Neurosymbolic AI: The 3rd Wave. arXiv. https://doi.org/10.48550/arXiv.2012.05876

[23] Bonatti, P., Decker, S., Polleres, A., & Presutti, V. (2019). Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web (Dagstuhl Seminar 18371). Dagstuhl Reports, 8(9), 29–111. https://doi.org/10.4230/DagRep.8.9.29

[24] Verma, S., Bhatia, R., Harit, S., & Batish, S. (2023). Scholarly knowledge graphs through structuring scholarly communication: a review. Complex & Intelligent Systems, 9(1), 1059–1095. https://doi.org/10.1007/s40747-022-00806-6

[25] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33.

[26] Hilton, J., Nakano, R., Balaji, S., & Schulman, J. (2021). WebGPT: Improving the factual accuracy of language models through web browsing. OpenAI Blog.

[27] Wu, T., Terry, M., & Cai, C. J. (2022). AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3491102.3517582

[28] Zhao, P., Zhang, H., Yu, Q., Wang, Z., Geng, Y., Fu, F., Yang, L., Zhang, W., Jiang, J., & Cui, B. (2024). Retrieval-Augmented Generation for AI-Generated Content: A Survey. arXiv. https://doi.org/10.48550/arXiv.2402.19473

[29] Shahab, O., El Kurdi, B., Shaukat, A., Nadkarni, G., & Soroush, A. (2024). Large language models: a primer and gastroenterology applications. Therapeutic Advances in Gastroenterology, 17, 17562848241227032. https://doi.org/10.1177/17562848241227031

[30] Peng, C., He, S., Xu, Y., Li, L., Du, N., Chen, L., Zhang, Y., Li, F., Xie, Y., Sun, X., & Xie, P. (2023). A study of generative large language model for medical research and healthcare. npj Digital Medicine, 6(1). https://doi.org/10.1038/s41746-023-00958-w

[31] Sai, S., Gaur, A., Sai, R., Chamola, V., Guizani, M., & Rodrigues, J. J. P. C. (2024). Generative AI for Transformative Healthcare: A Comprehensive Study of Emerging Models, Applications, Case Studies and Limitations. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3365979

[32] Nassiri, K., & Akhloufi, M. A. (2024). Recent Advances in Large Language Models for Healthcare. BioMedInformatics, 4(2), 1097–1143. https://doi.org/10.3390/biomedinformatics4020068

[33] Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., Nabeshima, N., Presser, S., & Leahy, C. (2020). The Pile: An 800GB dataset of diverse text for language modeling. arXiv. https://doi.org/10.48550/arXiv.2101.00027

[34] Miller, N., Tyler, R. J., & Backus, J. E. B. (2004). MedlinePlus®: the National Library of Medicine® brings quality information to health consumers. Library Trends, 53(2), 375-388.

[35] Pezzella, P. (2022). The ICD-11 is now officially in effect. World Psychiatry, 21(2), 331. https://doi.org/10.1002/wps.20997

[36] Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., Qiu, J., Yao, Y., Zhang, A., Zhang, L., Han, W., Huang, M., Jin, Q., Lan, Y., Liu, Y., Liu, Z., Lu, Z., Qiu, X., Song, R., … Zhu, J. (2021). Pre-trained models: Past, present and future. AI Open, 2, 225–250. https://doi.org/10.1016/j.aiopen.2021.08.002

[37] Wang, H., Li, J., Wu, H., Hovy, E., & Sun, Y. (2023). Pre-Trained Language Models and Their Applications. Engineering, 25, 51–65. https://doi.org/10.1016/j.eng.2022.04.024

[38] Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). Don't stop pretraining: Adapt language models to domains and tasks. arXiv. https://doi.org/10.48550/arXiv.2004.10964

[39] Morande, S., & Tewari, V. (2023). Causality in Machine Learning: Innovating Model Generalization through Inference of Causal Relationships from Observational Data. Qeios. https://doi.org/10.32388/P7MMGR

[40] Abbasian, M., Abedian, S., Agrawal, P., Alqahtani, S., Alshammari, N., Alsheikh, N., Anand, S., Athey, K., Balasubramanian, R., Balki, I., Blei, D., Browne, O., Buhr, E., Chen, I. Y., Chen, P.-H. C., Chowdhury, R., Corey, K. E., Dalke, A. R., Dubasov, C., … Zou, J. (2024). Foundation metrics for evaluating effectiveness of healthcare conversations powered by generative AI. NPJ Digital Medicine, 7(1), 82. https://doi.org/10.1038/s41746-024-01074-z

[41] Krishna, K., Ramprasad, S., Gupta, P., Wallace, B. C., Lipton, Z. C., & Bigham, J. P. (2024). GenAudit: Fixing Factual Errors in Language Model Outputs with Evidence. arXiv. https://doi.org/10.48550/arXiv.2402.12566

[42] Anwar, U., Saparov, A., Rando, J., Paleka, D., Turpin, M., Hase, P., Lubana, E. S., Jenner, E., Casper, S., Sourbut, O., Edelman, B. L., Zhang, Z., Günther, M., Korinek, A., Hernandez-Orallo, J., Hammond, L., Bigelow, E., Pan, A., Langosco, L., ... Krueger, D. (2024). Foundational Challenges in Assuring Alignment and Safety of Large Language Models. arXiv. https://doi.org/10.48550/arXiv.2404.09932

[43] Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. Information Fusion, 99, 101896. https://doi.org/10.1016/j.inffus.2023.101896

[44] Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. Computers in Biology and Medicine, 158, 106848. https://doi.org/10.1016/j.compbiomed.2023.106848