Bridging the Gap or Widening the Divide: A Call for Capacity-Building in Artificial Intelligence for Healthcare in the Philippines

Kurt Lenard T. Gutierrez,[1] *Paulene Miriel L. Viacrusis*[1]

ABSTRACT

The emerging field of generative artificial intelligence (GAI) and some of its well-known technologies: ChatGPT, Google Bard and Claude, have gained substantial popularity due to their enormous potential in healthcare applications, as seen in medically fine-tuned models such as Med-PaLM and ChatDoctor. While these advancements are impressive, the dependence of AI development on data volume and quality raises questions about the generalizability of these models. Regions with lower medical research output risk bias and misrepresentation in AI-generated content, especially when used to assist clinical practice. Upon testing of a prompt concerning the isoniazid dosing of Filipinos versus other ethnic and racial groups, responses from GPT-4, GPT-3, Bard and Claude resulted in 3 out of 4 outputs showing convincing but false content, with extended prompting illustrating how response hallucination happens in GAI models. To address this, model refinement techniques such as fine-tuning and prompt ensembles are suggested; however, refining AI models for local contextualization requires data availability, data quality and quality assurance frameworks. Clinicians and researchers in the Philippines and other underrepresented regions are called to initiate capacity-building efforts to prepare for AI in healthcare. Early efforts from all stakeholders are needed to prevent the exacerbation of health inequities, especially in the new clinical frontiers brought about by GAI.

Keywords Artificial Intelligence, Bias, ChatGPT, Healthcare, Philippines

INTRODUCTION

From automated scribing to clinical decision support tools — the fast-paced development of artificial intelligence (AI) and virality of user-friendly tools like ChatGPT opens a transformative frontier in clinical practice. High-level advances in healthcare applications like generalist biomedical AI systems (GMAI) and multimodal models appear to be significant steps towards bridging the care gap;[1–3] however, the inherent dependency of AI development on data quality and availability may exacerbate inequalities in medicine, especially in communities with scarce medical literature. This raises concerns about the generalizability of healthcare AI systems, especially considering that most of the development in this space is built with Western-oriented datasets. Unlike clinical practice guidelines that consider local context, AI models are built from pretraining on

✉ Kurt Lenard T. Gutierrez
kurtlenard.gutierrez.med@ust.edu.ph

1 Student, Doctor of Medicine, Faculty of Medicine & Surgery, University of Santo Tomas, España, Manila, Philippines

Academic editor: Leilani B. Mercado-Asis
Submitted date: October 25, 2023
Accepted date: October 31, 2023

vast amounts of data that may or may not consider clinical nuances in underrepresented localities.[4] Despite the availability of AI-based tools, healthcare professionals in the Philippines must be wary of the risks of utilizing them in clinical practice and build initiatives to contextualize AI tools with local clinical nuances in mind.

What is Artificial Intelligence?

Artificial intelligence (AI) may be defined in a multitude of ways. However, for the purposes of understanding its impact on clinical practice, AI is defined as a collection of technologies that utilize computational learning methods (eg, machine learning, neural networks, deep learning, natural language processing) in interpreting, analyzing and generating data, aiming to replicate human-level intelligence.[4,5]

The most relevant type of AI in this discussion is GAI, which is differentiated from other types of AI by its autonomous ability to flexibly generate seemingly novel data without explicit human programming.[4] Under GAI, there are multiple types of applications like diffusion models, large language models and multimodal models, which focus on image, text and a combination of multiple data types (eg, audio, video, image, text) for interpretation and generation, respectively.

ChatGPT and the Development of Generative AI for Healthcare Applications

GAI traces back to the first neural networks developed in 1943, and one of the groundworks that laid the foundation for generative models roots from a 2017 paper published by Google titled "Attention is all you need," — giving birth to *transformer models*, the core model used in Generative Pre-Trained Language Models (GPT) that we all know today as the technology behind ChatGPT.[6,7]

ChatGPT is an AI-powered conversational model that produces human-like responses to multimodal inputs released by OpenAI in November 2022.[8] It earned 100 million users in 2 months, the fastest uptake of users in any emerging technology.[9] Its fame comes from its ability to read, summarize, interpret and produce human-like output based on human-produced input — giving it the capacity to hold conversations with human prompts, thus opening a world of possibilities in various domains, including medicine.[9–11]

Since ChatGPT, public interest and rapid development of healthcare-focused tools powered by GAI models like GPT, PaLM and LLaMA have rapidly been advancing and are continuously gaining popularity due to their enormous potential in clinical use. As of October 2023, there are 692 AI and machine learning-enabled medical devices filed with the US Food and Drug Administration; however, no GAI-enabled projects have been approved just yet.[12] A list of notable developments in GAI for healthcare built in connection with technology giants Google, Meta, Microsoft and Amazon are summarized in Table 1.

The list was limited to initiatives and projects linked to GAI tools from major technology companies with publicly-recognized in-house AI foundation models for concise information delivery. As of writing, the pacing of AI development consists of hundreds to thousands of initiatives in early-stage startups and nationally-backed research worldwide; hence, the scope of this table was limited to a few notable models.

Bias, Underrepresentation and Hallucination in GAI

While recent developments are indeed revolutionary, it is unlikely for AI tools to be a plug-and-play technology. One particular risk that AI carries is inherently rooted in the underlying technology of its creation. Generative AI relies on *Foundation Models,* which are large models built from processing massive amounts of data using self-supervised pretraining methods that learn the patterns and relationships of data from the provided content.[18]

The pretraining methods that build foundation models are observed to be scale-dependent, and model competitiveness improves with greater data volume.[19] However, this poses a significant risk: data-poor communities, especially in developing regions with lower medical research output, lag in AI utilization due to models being skewed to represent available data from developed regions. The domination of Western data in medical literature, and consequently, AI pretraining datasets, risks the failure to represent the unique clinical presentations of patients from minority groups with scarce medical literature.[20,21]

Table 1 Google, Meta, Microsoft and Amazon-based initiatives on Generative Artificial Intelligence in Healthcare.

Name	Description	Source
Med-PaLM	A fine-tuned large language model (LLM) built from the Pathways Language Model (PaLM) as the baseline foundation and its instruction-tuned Flan-PaLM. When assessed with MedQA, a USMLE-styled question bank, Med-PaLM reportedly scored 67.1% in answer accuracy. In MedMCQA, a dataset on India's medical entrance exam, Med-PaLM scored 57.6%.	Singhal, et al. 2023 [2] Google Research *Preprint first published on 26 Dec 2022, Journal article published on 11 July 2023*
Med-PaLM 2	Improved version of Med-PaLM, scoring 86.5% on the USMLE-based MedQA dataset, a +19% improvement from earlier models.	Singhal, et al., 2023 [13] Google Research *Preprint first published on 16 May 2023*
Med-PaLM Multimodal (Med-PaLM M)	A proof-of-concept multimodal generative biomedical AI system that can interpret and integrate clinical language, radiology, genomics and other data. Tested on a task-based multimodal biomedical benchmark, *MultiMedBench,* consisting of question-answering, radiology report generation tasks, etc. Participating clinicians reportedly expressed a preference for radiology reports generated by Med-PaLM Multimodal over human radiologists in up to 40.5% of cases.	Tu, et al. 2023 [1] Google Research & Google DeepMind *Preprint first published on 26 July 2023*
ChatDoctor	A medically fine-tuned large language model refined from the Large Language Model Meta AI (LLaMA) as the baseline foundation. Adaptation and refinement were done using a dataset of 100,000 anonymized patient-doctor dialogues. Built as a chat assistant for patient medical advice.	Li, et al. 2023 [14] *LLaMA model is a foundation model developed by Meta* [15] *Preprint first published on 24 Mar 2023, Journal Article published on 24 June 2023*
HealthScribe	A HIPAA-compliant AI-enabled tool to automate the generation of clinical notes from patient-doctor conversations based on Amazon Bedrock, an application programming interface (API)-based *base model* tool for GAI applications.	Amazon Inc., 2023 [16] *Announced in an Amazon Press Release last 27 July 2023*
Dragon Ambient eXPerience Express (DAX Express)	An automated clinical documentation tool powered by OpenAI's GPT-4 *foundation model*.	McGuinness & Nole, 2023 [17] Microsoft and Nuance Communications *Announced in an official Microsoft post on 20 Mar 2023*

Simply put, foundational models are trained on vast amounts of data and follow the simple concept of 'garbage in, garbage out'. If data is either unavailable or is of poor quality — then the output of an AI model may not be of clinical use. Bluntly put, it is unlikely for US-based AI research groups to include updated clinical practice guidelines and medical literature from the Philippines in training their models.

Before the ChatGPT-induced AI boom, concerns about bias in healthcare AI had already been pointed out.[20,21] However, what makes the era of GAI different is its capacity not only to interpret but also analyze and generate novel data. What comes with its ability to create new data is the capability to generate seemingly true but fabricated data, achieved in various ways, like by citing research papers that don't exist or convincing claims without factual evidence, a phenomenon commonly referred to as *hallucination*.[22,23] This multiplies the risk of bias, especially when these tools assist in clinical decision-making.

To illustrate this, we show the AI-generated responses to a prompt that is very relevant to the Philippine healthcare situation, as shown in Table 2.

No prior chats, custom instructions or prompt tuning methods were performed. GPT-3.5 and GPT-4 were set on default settings and internet access

Table 2 AI-Generated Responses on Isoniazid dosing variability, retrieved from models by OpenAI (ChatGPT GPT-3.5 and GPT-4), Google (Bard) and Anthropic (Claude). Generated on 20 October 2023.

PROMPT: In less than 100 words, explain why isoniazid dosing tends to differ for Filipinos when compared to other racial and ethnic groups.

ChatGPT (GPT-3.5)
Isoniazid dosing can differ for Filipinos due to genetic variations in drug metabolism. *Many Filipinos have a higher prevalence of slow acetylation*, a genetic trait that affects how the body processes isoniazid. This can result in slower drug clearance and increased risk of adverse effects or toxicity at standard dosages. Therefore, lower initial dosages or alternative regimens are often recommended to ensure safety and effectiveness in this population.

ChatGPT (GPT-4)
Isoniazid dosing can vary for Filipinos due to their genetic makeup, which often includes a *higher prevalence of slow acetylators*. The acetylation process, governed by the NAT2 gene, affects how quickly isoniazid is metabolized in the body. Slow acetylators metabolize the drug more slowly, leading to higher drug levels in the blood and increased risk of side effects. Consequently, healthcare providers may adjust dosages or monitor these patients more closely to mitigate the risks of toxicity and other adverse effects while ensuring the medication is effective.

Google Bard
Isoniazid is an antibiotic used to treat and prevent tuberculosis (TB). The dosing of isoniazid varies depending on a number of factors, including age, weight and kidney function. However, there is also evidence that ethnicity can play a role in isoniazid dosing.
Filipinos are more likely to experience side effects from isoniazid than other racial and ethnic groups. This is thought to be due to a genetic difference that affects the way Filipinos metabolize isoniazid. As a result, Filipinos typically receive a lower dose of isoniazid than other groups.

Claude AI
Isoniazid dosing often needs adjustment in Filipino patients because they tend to *metabolize the drug more rapidly due to genetic differences influencing liver enzymes*. This leads to lower isoniazid levels, reducing efficacy. Lower initial doses are recommended for Filipinos to avoid subtherapeutic levels while minimizing side effects.

was not enabled. The response from Google Bard is a snippet from a 97-word answer for conciseness on this paper; see Appendix 1 for full response and settings applied.

In literature, it is generally agreed that Filipinos are rapid acetylators, which comes as a consequence of being a country with high tuberculosis disease burden.[24,25] Despite this, three out of the four AI-generated responses, as shown in Table 2, convincingly claim otherwise. When GPT-4, the latest model from OpenAI as of October 2023, was retested with a request to provide peer-reviewed citations for the claim, the response provided produced a fake citation with a DOI link that leads nowhere, as shown in Table 3 below:

No prior chats, custom instructions or prompt tuning methods were performed. GPT-4 was set on default settings.

Upon further assessment, the real-world paper that closely resembles the cited article in the response found in Table 3 is:

Cai Y, Yi J, Zhou C, Shen X. Pharmacogenetic Study of Drug-Metabolising Enzyme Polymorphisms on the Risk of Anti-Tuberculosis Drug-Induced Liver Injury: A meta-analysis, *PLoS One*. 2012;7(10):e47769. DOI: 10.1371/journal.pone.0047769.

Clearly, GPT-4 was able to retrieve a reference; however, it falsely modified the title, journal and links in a very convincing format. When reviewing the studies included in the cited meta-analysis, Filipinos were not even part of the included samples,[26] yet it was utilized to generate the response. If not double-checked otherwise, then this claim may appear to be as good as any material containing a complete citation.

While we acknowledge that elaborate prompt-tuning or modification of model weights and settings can create responses with better precision, these *hallucinations* still pose a significant danger, especially when applied in clinical practice. Assessing the risk of isoniazid-induced liver injury relies on the patient's rapid or slow acetylator status.[24,25,27] If AI-assisted clinical decision support tools generate *hallucinating* responses like the one shown above, potentially influencing a physician's dosing regimen, then patients are placed under unnecessary harm.

Capacity-building for artificial intelligence in the Philippines

GAI and its applications in healthcare are in their infancy stages, and there are a multitude of

Table 3 Retest of Isoniazid dosing query with citation prompt for GPT-4. Generated on 20 October 2023.

PROMPT: In less than 100 words, explain why isoniazid dosing tends to differ for Filipinos when compared to other racial and ethnic groups. Cite peer-reviewed references for your claims.
RESPONSE from ChatGPT (GPT-4): Isoniazid dosing can differ for Filipinos due to genetic variations that affect drug metabolism. A higher prevalence of slow acetylators is found among Filipinos and other Asian populations leading to slower rate of isoniazid clearance from the body. Slow acetylators may require adjusted dosing to minimize the risk of drug toxicity while ensuring treatment efficacy. Reference: Cai Y, Yi J, Zhou C, et al. NAT2 polymorphisms and susceptibility to anti-tuberculosis drug-induced liver injury: a meta-analysis. *Int J Tuberc Lung Dis*. 2012;16(5):589-595. DOI: 10.5588/ijtld.11.0568. Please note, always refer to the most recent and pertinent literature for up-to-date information.

ways to ensure the applicability of internationally developed AI models in the Philippine setting. Just like how clinical practice guidelines stem from a combination of internationally accepted standards applied to the local context, *foundation models* can further be trained on local datasets using methods like fine-tuning, word embeddings, content injection, prompt ensembles and a multitude of other techniques.[2,3,19,28–30] Despite the existence of these methods, model refinement continues to rely on fundamentals like high-quality datasets and validation frameworks.

With this in mind, clinicians and researchers in the Philippines need to develop quality assurance (QA) and validation frameworks to assess models similar to the likes of MedQA, PubMedQA and MultiMedQA, which were used to assess medically fine-tuned foundation models like Google's Med-PaLM.[13,31] Health facilities should be prepared to employ recurring local validation strategies that take site-specific context into assessing AI implementation [32] and ensure safe data-gathering methods with a privacy-first approach to iterating AI models.

Early initiatives like employing electronic health record systems, building and validating QA frameworks using updated clinical practice guidelines from medical societies and creating anonymized patient datasets from tertiary hospitals using Philippine data should be further welcomed and supported at a policy level.

The state of artificial intelligence projects in the Philippines

The Philippines is not completely behind the AI wave: The Department of Health, together with the Philippine Health Insurance Corporation, is pioneering the National Health Data Repository (NHDR), which would be the country's first integrated architecture for medical data.[33] While the NHDR Framework is already available online, there has yet to be an official statement on its projected completion.

Among other projects is the CANDLE Study by the National Health Institute, which includes phenotypically scoring CT scans for early liver cancer detection [34]; the CHERISH2 App of the Department of Health in partnership with a tertiary hospital for screening COVID-Pneumonia [35]; and a Dengue prediction model using Long Short-term Memory (LSTM) made by researchers from the University of the Philippines [36].

Overall, progress is being made with the collaborative efforts of both the private and public sectors in developing Filipino-oriented AI, and this serves as a waypoint towards building locally contextualized resources for emerging AI models.

CONCLUSION

It is clear that AI is here to stay and developing countries risk lagging behind. In 2023 alone, major tech giants have been launching medically fine-tuned models left and right, and are actively integrating these into applications for consumer use. Though impressively scoring in clinical question banks, these models still grapple with biases, hallucinations and misinformation. This is rooted in how these models are trained, provided the limited data of patients from the lesser end of the digital divide.

Fundamentally, AI is dependent on data quality, volume and validation. The lack of Filipino representation in the pretraining data poses substantial risks for these models to perpetuate false or potentially harmful clinical advice. These cutting-edge technologies are not plug-and-play, and capacity-building is needed to prepare ourselves to welcome this emerging field. The Philippines must employ policies to ensure the safety and quality of emergent AI technologies; clinicians and researchers should initiate the development of validated datasets and

quality assurance frameworks, and medical schools must include AI in classroom discourse early on.

The pursuit of representation in this emerging field is paramount in determining the course of AI in Philippine healthcare. Our early initiatives will shape the future of this emerging technology: will AI become a stepping stone towards universal healthcare, or add to our list of health inequities to overcome?

The tightrope between bridging the gap or widening the divide relies entirely on how we collectively decide to tip the balances of this technology — and we are at the forefront of building its future.

REFERENCES

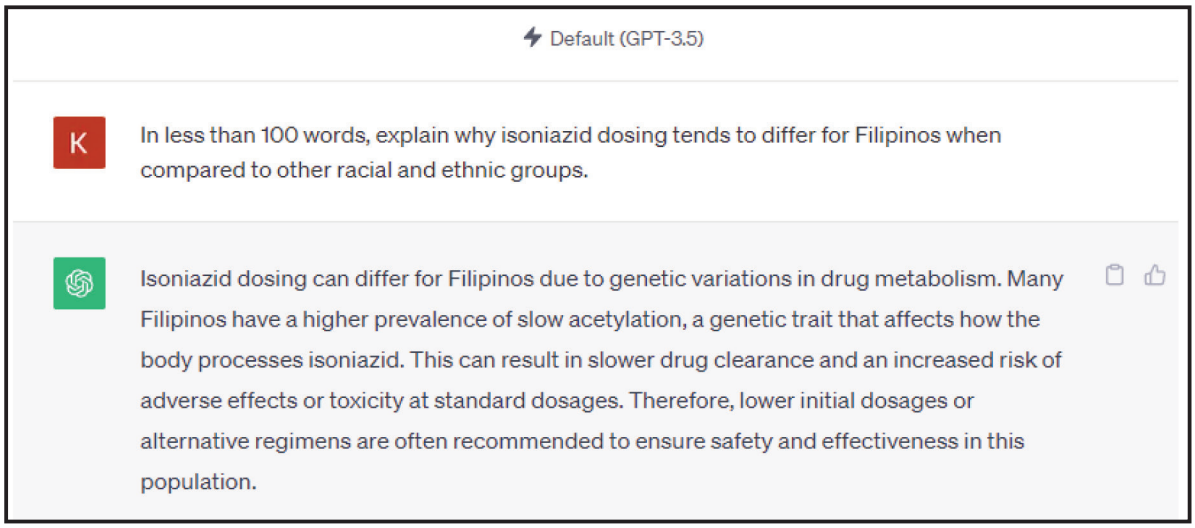
1. Tu T, Azizi S, Driess D, Schaekermann M, Amin M, Chang PC, et al. Towards generalist biomedical AI [Internet]. arXiv; 2023. Available from: http://arxiv.org/abs/2307.14334
2. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. *Nature*. 2023 Aug;620(7972):172–80.
3. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. *Nature*. 2023 Apr;616(7956):259–65.
4. Wang P. On defining artificial intelligence. *J Artif Gen Intell*. 2019 Jan 1;10(2):1–37.
5. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. 2019 Jun;6(2):94–8.
6. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need [Internet]. arXiv; 2023. Available from: http://arxiv.org/abs/1706.03762
7. Emmert-Streib F, Yang Z, Feng H, Tripathi S, Dehmer M. An introductory review of deep learning for prediction models with big data. *Front Artif Intell* [Internet]. 2020;3. Available from: https://www.frontiersin.org/articles/10.3389/frai.2020.00004
8. Introducing ChatGPT [Internet]. San Francisco (US): OpenAI; 2022 Nov 30 [cited 2023 Oct 10]. Available from: https://openai.com/blog/chatgpt
9. Hu K. ChatGPT sets record for fastest-growing user base - analyst note. Reuters [Internet]. 2023 Feb 2. Available from: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/
10. OpenAI. GPT-4 technical report. arXiv [Preprint]. 2023 Mar 27 [cited 2023 Aug 6]. Available from: http://arxiv.org/abs/2303.08774
11. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell*. 2023 May 4;6:1169595.
12. U.S. Food and Drug Administration. Artificial intelligence and machine learning (AI/ML)-enabled medical devices. Digit Health Cent Excell [Internet]. 2023 Oct 19. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices
13. Singhal K, Tu T, Gottweis J, Sayres R, Wulczyn E, Hou L, et al. Towards expert-level medical question answering with large language models. arXiv [Preprint]. 2023 May 16. Available from: http://arxiv.org/abs/2305.09617
14. Li Y, Li Z, Zhang K, Dan R, Jiang S, Zhang Y, et al. ChatDoctor: A medical chat model fine-tuned on a Large Language Model Meta-AI (LLaMA) using medical domain knowledge. *Cureus* [Internet]. 2023 Jun 24;15(6). Available from: https://www.cureus.com/articles/152858-chatdoctor-a-medical-chat-model-fine-tuned-on-a-large-language-model-meta-ai-llama-using-medical-domain-knowledge
15. Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, Lacroix T, et al. LLaMA: Open and efficient foundation language models.
16. Amazon.com, Inc. AWS announces AWS HealthScribe, a new generative AI-powered service that automatically creates clinical documentation [Internet]. Amazon Press Center. 2023. Available from: https://press.aboutamazon.com/2023/7/aws-announces-aws-healthscribe-a-new-generative-ai-powered-service-that-automatically-creates-clinical-documentation
17. McGuinness, Tom, Nole, Diana. Breaking new ground in healthcare with the next evolution of AI [Internet]. The Official Microsoft Blog. 2023. Available from: https://blogs.microsoft.com/blog/2023/03/20/breaking-new-ground-in-healthcare-with-the-next-evolution-of-ai/
18. Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, et al. On the opportunities and risks of foundation models. arXiv [Preprint]. 2022 Jul 12. Available from: http://arxiv.org/abs/2108.07258
19. Lester B, Al-Rfou R, Constant N. The power of scale for parameter-efficient prompt tuning. arXiv [Preprint]. 2021 Sep 2. Available from: http://arxiv.org/abs/2104.08691
20. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. *Patterns*. 2021 Oct 8;2(10):100347.
21. Kleinberg G, Diaz MJ, Batchu S, Lucke-Wold B. Racial underrepresentation in dermatological datasets leads to biased machine learning models and inequitable healthcare. *J Biomed Res*. 2022;3(1):42–7.
22. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus*. 15(2):e35179.
23. Goddard J. Hallucinations in ChatGPT: A cautionary tale for biomedical researchers. *Am J Med* [Internet]. 2023 Jun 25;0(0). Available from: https://www.amjmed.com/article/S0002-9343(23)00401-1/fulltext
24. Gregorio G, Cutiongco-dela Paz EM, Gonzales ML, dela Paz EM, Rocamora FM. N-acetyltransferase2 genotype and its association with hepatitis among Filipino children treated with first line anti-tuberculosis drugs. *Acta Med Philipp* [Internet]. 2010 Dec 31;45(1). Available from: https://www.herdin.ph/index.php/partner/journal?view=research&cid=39818
25. Autus-Geniston LV, Matias RR, Tuazon AO. Allelic frequencies of single nucleotide polymorphisms (SNPs) in the N-acetyltransferase 2 (NAT2) gene of Filipinos. *Acta Med Philipp* [Internet]. 2010;44(3). Available from: https://actamedica-philippina.upm.edu.ph/index.php/acta/article/view/7702
26. Cai Y, Yi J, Zhou C, Shen X. Pharmacogenetic study of drug-metabolising enzyme polymorphisms on the risk of anti-tuberculosis drug-induced liver injury: A meta-analysis. *Plos One*. 2012 Oct 17;7(10):e47769.
27. Wang P, Pradhan K, Zhong X bo, Ma X. Isoniazid metabolism and hepatotoxicity. *Acta Pharm Sin B*. 2016 Sep;6(5):384–92.
28. Li Y, Li Z, Zhang K, Dan R, Jiang S, Zhang Y. ChatDoctor: A medical chat model fine-tuned on a Large Language Model Meta-AI (LLaMA) using medical domain knowledge. arXiv [Preprint]. 2023 Jun 24. Available from: http://arxiv.org/abs/2303.14070
29. Ding N, Qin Y, Yang G, Wei F, Yang Z, Su Y, et al. Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nat Mach Intell*. 2023 Mar;5(3):220–35.
30. Varghese J, Chapiro J. ChatGPT: The transformative influence of generative AI on science and healthcare. *J Hepatol* [Internet]. 2023 Aug 4;0(0). Available from: https://www.journal-of-hepatology.eu/article/S0168-8278(23)05039-0/fulltext
31. Zhang X, Wu J, He Z, Liu X, Su Y. Medical exam question answering with large-scale reading comprehension. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Education Advances in Artificial Intelligence; 2018 February; Louisiana (USA): AAAI Press. Available from https://dl.acm.org/doi/abs/10.5555/3504035.3504735

32. Youssef A, Pencina M, Thakur A, Zhu T, Clifton D, Shah NH. External validation of AI models in health should be replaced with recurring local validation. *Nat Med*. 2023 Oct 18;1–2.
33. Department of Health, Philippine Health Insurance Corporation. National Health Data Repository Framework [Internet]. Department of Health; 2022. Available from: https://www.philhealth.gov.ph/about_us/NationalHealthDataRepositoryFramework03282022.pdf
34. Tiangco, Beatrice. Early cancer detection in the liver of Filipinos with chronic hepatitis B using AI-driven integration of clinical and genomic biomarkers (CANDLE Study) [Internet]. National Institutes of Health, University of the Philippines - Manila; 2019. Available from: https://bidsandawards.upm.edu.ph/sites/default/files/PROFESSIONAL%20SERVICES%20Scan.pdf
35. Baltazar LR, Manzanillo MG, Gaudillo J, Viray ED, Domingo M, Tiangco B, et al. Artificial intelligence on COVID-19 pneumonia detection using chest x-ray images. *Plos One*. 2021 Oct 14;16(10):e0257884.
36. Ligue, Kim, Ligue, Kristine. Deep learning approach to forecasting dengue cases in Davao City using long short-term memory (LSTM). *Philipp J Sci*. 2022 Jun;151(3):797–812.

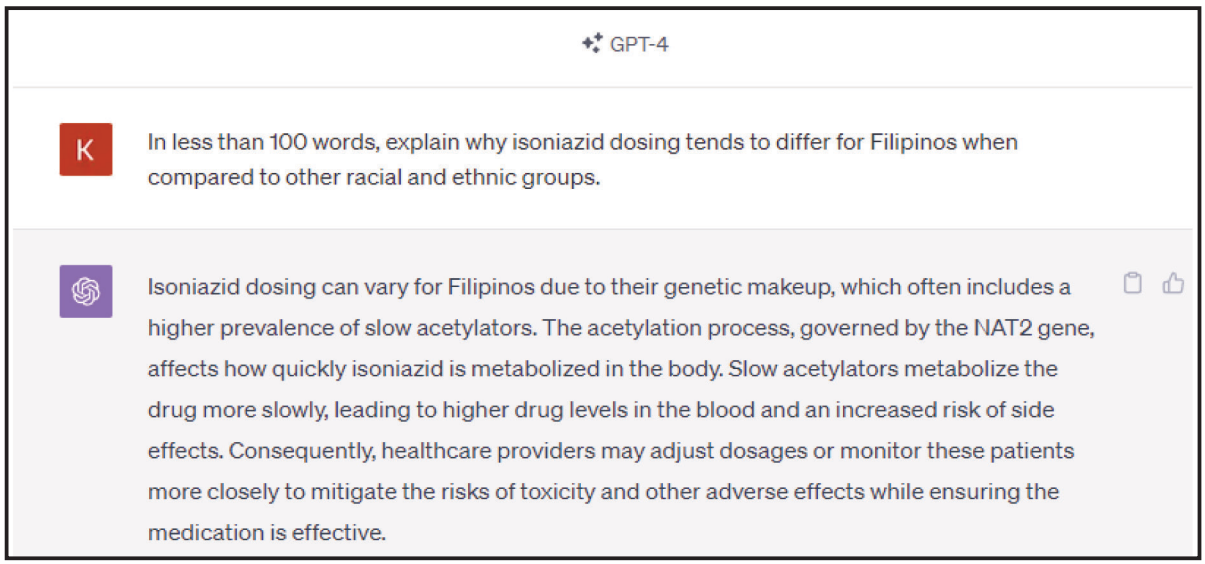
APPENDIX 1

Language model output on prompts concerning isoniazid dosing, results summarized in Tables 2 and 3.

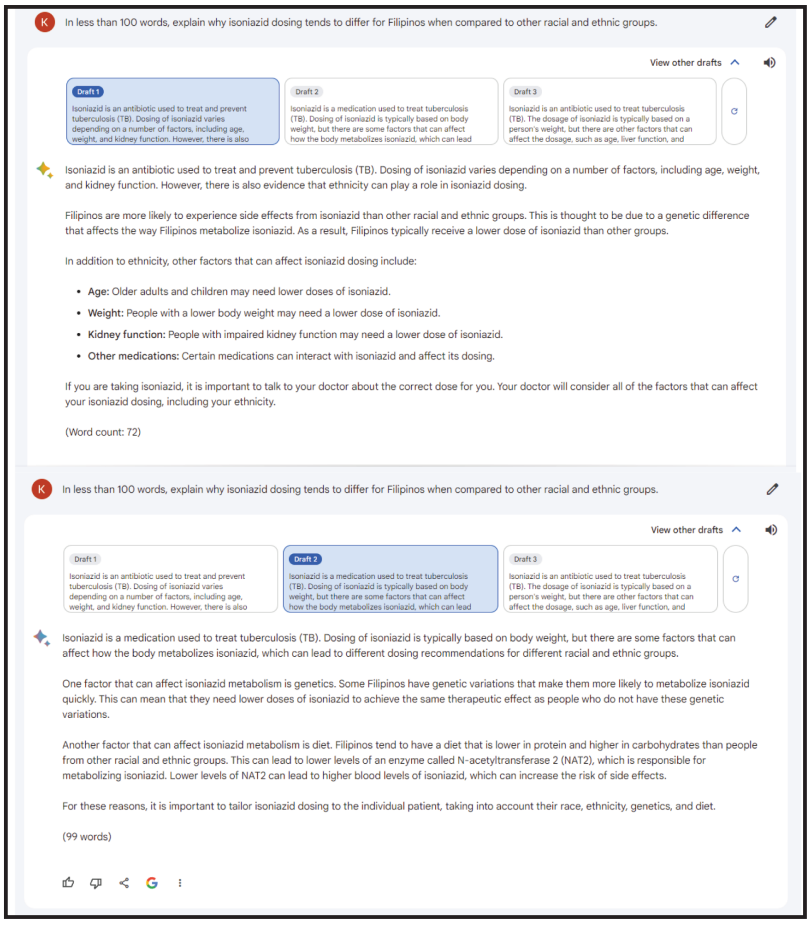
A. OpenAI ChatGPT (GPT-3.5). ChatGPT September 25, 2023 version. Response retrieved on 20 October 2023 UTC+8 from Metro Manila, Philippines. Default settings were utilized in generating the response.

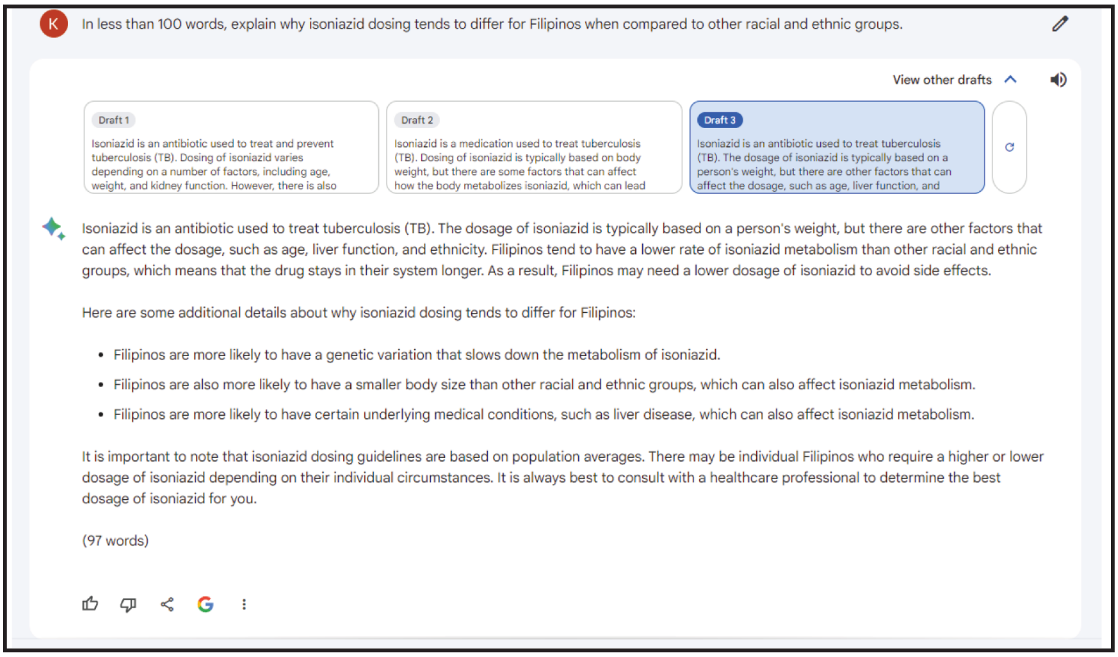


B. OpenAI ChatGPT (GPT-4). ChatGPT September 25, 2023 version. Response retrieved on 20 October 2023 UTC+8 from Metro Manila, Philippines. Default settings were utilized in generating the response, GPT-4 was accessed from a ChatGPT Plus account.



C. Google Bard (Bard AI 'Experiment'). 2023.09.27 Version. Response retrieved on 20 October 2023 UTC+8 from Metro Manila, Philippines. Google Bard generated three draft responses; result of Draft 1 is featured in **Table 2** as it is the first featured response automatically shown by the chatbot. Responses from Draft 2 and 3 are shown below for reference:





D. Anthropic Claude AI. Claude AI Free Version. Anthropic does not publicly provide versioning information for Claude on its chat page. Response retrieved on 20 October 2023 UTC+8 from Metro Manila, Philippines.