



A PRACTICAL FRAMEWORK FOR EVALUATING AI CHATBOTS IN NON-ENGLISH DRUG INFORMATION SERVICES

Inthira Kanchanaphibool^{1,*}, Panyanat Aonpong², Thanaphat Dabngoen¹, Krittin Kongmee¹, Thanawat Wongcharee¹, Tanut Tantasuparuk¹, Sanita Hirunrassamee³

¹ Division of Pharmaceutical Care, Faculty of Pharmacy, Silpakorn University, Sanamchandra Palace Campus, Nakhon Pathom

² Department of Computing, Faculty of Science, Silpakorn University, Sanamchandra Palace Campus, Nakhon Pathom

³ Drug Information and Consumer Protection Center, Center of Excellence in Pharmacy Practice and Management Research, Faculty of Pharmacy, Thammasat University, Pathum Thani

* Corresponding author: Kanchanaphibool_i@silpakorn.edu

ABSTRACT

Artificial intelligence, particularly AI-based chatbots, has gained increasing attention as a tool to support pharmacists in providing drug information services, especially in settings with pharmacist shortages. This study compared the competencies of ChatGPT, ChatGPT Pro, Perplexity, and Perplexity Pro in responding to drug-related clinical questions, and examined the effect of an Enhanced Task Translation (ETT) technique on chatbot performance, after English technical terms were embedded with Thai-language prompts. An experimental comparative design was employed. Two sets of Thai-language multiple-choice questions (MCQs), each comprising 120 items based on the Thai Pharmacy Licensing Examination, were administered: one standard set and one ETT set. Cardiology-focused clinical case scenarios were additionally presented using a SOAP-note format. Performance was assessed using a rubric adapted from the American Society of Health-System Pharmacists (ASHP) guidelines across four domains: question classification, source citation, evidence application, and communication. All four chatbots surpassed the 60% passing threshold: 72.08–83.33% for the standard set and 73.75–80.83% for the ETT set. Perplexity Pro demonstrated the highest MCQ performance, while the inclusion of English technical terms did not consistently improve results across models. In the clinical case assessment, ChatGPT Pro achieved the highest rubric score, showing strong evidence use and clinical reasoning. These findings suggest that AI chatbot competency in drug-related queries is broadly comparable to that of pharmacists. The ETT technique did not produce notable performance differences between Thai-only and mixed-language prompts. AI chatbots may support pharmacists by generating initial structured responses to drug information requests. The proposed evaluation framework provides a potential model for assessing AI competency in non-English settings and underscores the importance of jointly evaluating knowledge accuracy and clinical judgment to ensure safe integration into drug information services.

Keywords: AI chatbot, language competency assessment, drug information, hospital pharmacy practice

Received: 20 December 2025; Revised: 12 June 2026; Accepted: 16 June 2026

Introduction

The global shortage of pharmacists has become a critical concern, affecting healthcare systems in both high- and low-to-middle-income countries. Developed nations such as those in Europe and the United States have reported increasing strain on pharmaceutical services due to workforce limitations^{1,2} while developing countries, including India, Ethiopia, Uganda, and Taiwan have also identified pharmacist shortages as a key barrier to effective medication-related care.³⁻⁶ This challenge has direct impact for pharmaceutical service delivery, particularly in areas requiring professional expertise in drug information provision and patient medication monitoring.

To address these challenges, digital technologies, particularly artificial intelligence (AI), have been proposed as adjuncts to alleviate workload and improve service efficiency. In the pharmaceutical area, AI chatbots have emerged as effective tools for healthcare communication, reducing repetitive inquiries and enhancing service delivery. Evidence from Thailand demonstrates their success in regulatory contexts, such as cosmetic advertising guidance and private clinic licensing^{7,8}, while broader applications highlight their transformative role in Industry 4.0.⁹ Several studies have also highlighted the integration of AI into drug information and pharmacy education, reinforcing its role in supporting clinical decision-making and professional training.¹⁰⁻¹² AI applications seem to have gained increasing traction in health-related fields, supporting functions such as symptom assessment, clinical triage, patient education, and research synthesis.¹³ In pharmacy, AI-powered chatbots have been developed to assist in patient drug information services, offering scalable solutions to meet rising demands.^{14,15} However, concerns persist regarding the accuracy, reliability, and contextual relevance of AI-generated responses, especially when such tools are deployed in clinical

contexts requiring nuanced pharmaceutical judgment.

These limitations are further compounded in non-English-speaking environments, including Thai language, where AI performance may be constrained by limited multilingual training. In fact, most large language models are trained predominantly on English-language datasets collected from digital sources such as websites, books, academic articles, and social media. As a result, their ability to process technical terms and provide culturally appropriate responses in other languages, such as Thai, remains uncertain.¹⁶ This linguistic constraint may result in inaccurate or misleading recommendations, thereby posing potential risks to patient medication safety.

Despite emerging research evaluating chatbot performance in healthcare—focusing on metrics such as readability, completeness, and user satisfaction^{15,17}—there remains a paucity of studies assessing the clinical competency of AI chatbots using standardized frameworks. In particular, there is limited evidence comparing AI performance directly against human professionals in drug information tasks, constituting a key knowledge gap.

A second gap concerns the feasibility of improving chatbot performance in resource-limited settings. Strategies such as model fine-tuning and expanded training datasets, while potentially effective, require technical expertise that healthcare professionals may not possess. Therefore, there is a need for user-friendly, accessible methods to enhance chatbot functionality without relying on specialized computer science knowledge.

To address these challenges, the present study compared the drug information competency of widely adopted AI chatbots in medical and clinical practices in Thailand with that of pharmacists using clinical questions derived from professional licensing frameworks. Additionally, the study evaluated the effectiveness of Enhanced Task Translation (ETT)—a linguistic prompting technique embedding English technical terms within Thai-language queries—in

improving chatbot performance in a non-English-speaking context such as Thai language, with potential generalizability to other languages.

Methods

Study Design

This study employed an experimental comparative design to evaluate the competency of AI chatbots commonly employed in healthcare practice in providing drug information responses. Each chatbot's performance was compared against the benchmark accuracy of 60% set by the Pharmacy Council of Thailand for the national licensure core competency examination, and clinical case scenario responses were reviewed by pharmacy experts using a structured rubric.

Rationale for Selection of AI Chatbots

This study selected ChatGPT and Perplexity based on both academic literature and real practice in Thailand during 2022–2024. The research team examined available articles in PubMed and Scopus, which indicated that nearly all research articles, reviews, and conference papers focused on ChatGPT. This predominance reflects its role as the first widely launched AI chatbot in late 2022 and its rapid global adoption in hospital pharmacy research and practice. Other chatbots were rarely mentioned in the academic literature.

In contrast, within Thai medical and pharmacy practice, Perplexity gained increasing attention because of its citation-rich responses, which were highly valued in evidence-based practice. Despite limited representation in international publications, Perplexity became relevant in local contexts as healthcare professionals in Thailand frequently adopted it to trace information back to original sources. Other AI chatbots such as Gemini, Claude, and Copilot were not widely discussed or adopted by healthcare professionals during that period.

Regarding the free and paid (Pro) versions, differences in features and competencies were expected to influence performance and utility.

ChatGPT Free (GPT 4o mini) provided general conversational support with limited reasoning depth, while ChatGPT Pro (GPT 4o) offered enhanced generative capacity for complex reasoning, writing, and coding.¹⁸ Similarly, Perplexity Free (Standard) enabled real-time search and citation, whereas Perplexity Pro emphasized faster retrieval, improved accuracy, and research-oriented tools.¹⁹

Accordingly, the study included four chatbots to comprehensively reflect differences in accessibility, functionality, and contextual relevance in hospital pharmacy practice:

- 1) ChatGPT (Free version, GPT-4o mini model)
- 2) ChatGPT Pro (ChatGPT Plus, GPT-4o model)
- 3) Perplexity (Free version, Default Standard model)
- 4) Perplexity Pro (Perplexity Pro, Default Pro model).

No prior training, prompt refinement, or familiarization sessions were conducted before formal testing. All AI chatbots were evaluated in their default configurations, using standardized prompts without pre-adjustment, to ensure that the results reflected typical user interactions in real-world healthcare practice. The evaluation included both quantitative and qualitative components.

Instruments and Competency Metrics

1. Multiple-Choice Questions (MCQs):

The development of 120 MCQs was guided by the Pharmacy Council of Thailand's examination framework (Pharmacy Licensure Examination – Pharmaceutical Care, PLE-PC1)²⁰ and validated through expert review. Five faculty experts contributed across domains: two clinical pharmacy specialists (therapeutic items), one pharmaceutical sciences expert (pharmaceutics and medicinal chemistry), one healthcare systems expert (public health), and one pharmacoeconomics expert (economic evaluation). This distribution ensured comprehensive blueprint coverage and alignment of item difficulty with PLE-PC1 standards. Two parallel sets of the examination were developed to

investigate the effect of language on test performance. Both sets contained the same 120 MCQs, constructed according to the PLE-PC1 blueprint and validated by the same panel of experts.

- **Set 1 (Standard format):** Questions in Thai only.

- **Set 2 (Enhanced format):** Same content as Set 1, but with English technical terms in parentheses following key Thai pharmaceutical terms to simulate the Enhanced Task Translation (ETT) technique commonly used by pharmacists in non-English speaking contexts. This design allowed direct comparison of candidate performance across languages. Content validity and cognitive level were controlled to ensure equivalence between the two examination sets.

Each chatbot's performance was assessed using the following indicators:

1. Accuracy Metrics

Two metrics were used to evaluate chatbot performance:

1.1 %Accuracy: Calculated as the number of correctly answered items divided by the total number of items.

$$\% \text{Accuracy} = \frac{\text{Number of correct responses}}{\text{Total number of items}} \times 100$$

1.2 %Weighted Accuracy: Calculated by multiplying the average %accuracy within each therapeutic area or knowledge domain by the number of items in that domain, then dividing by the total number of items.

$$\% \text{Weighted accuracy} = \frac{\sum (\% \text{Accuracy}_{\text{domain}} \times \text{Number of items in domain})}{\text{Total number of items}} \times 100$$

Example: If the average %accuracy in the therapeutics domain is 75% and the domain contains 57 items, then

$$\% \text{Weighted Accuracy} = \frac{75 \times 57}{120} = 35.6\%$$

which contributes proportionally to the overall weighted accuracy.

Rationale for Weighted Accuracy:

Using % accuracy alone may overestimate performance in domains with fewer items. For instance, if a chatbot answers all questions correctly in a small domain (e.g., 3 items), it would yield 100% accuracy, which does not realistically reflect overall competency. Weighted accuracy accounts for the unequal distribution of items across domains, ensuring that larger domains (e.g., therapeutics with 57 items) contribute proportionally to the overall score.

Weight Assignment:

Weights were assigned based on the number of items in each domain. For example, the therapeutics domain contained 57 items, so its weight was set to 57. This approach ensured that domains with more items had a greater influence on the weighted accuracy, reflecting their relative importance in the examination blueprint.

2. Subgroup analysis to identify chatbot strengths and weaknesses by therapeutic area and knowledge domain

3. A 60% minimum passing threshold in line with national licensure standards

2. Clinical Case Scenario Evaluation:

Cardiology domain was selected as the clinical case scenario for chatbot evaluation because cardiovascular diseases are both highly prevalent and clinically complex, requiring integration of pharmacotherapy, comorbidity management, and risk assessment. In addition, cardiology is one of the disease areas explicitly represented with a substantial number of MCQs in the PLE-PC1 blueprint²⁰, reflecting the importance of this domain in the licensure examination. Cardiology is also included in the scope of the case-based examination (PLE-PC2 blueprint).²¹ These factors collectively provided a strong rationale for selecting cardiology as the specialty for clinical case evaluation in this study.

To assess performance in complex clinical scenarios, a cardiology case study was developed following the structure of the national licensing examination. Each AI chatbot was required to respond in the SOAP-note format (Subjective, Objective, Assessment, and Plan) and address three stages of the pharmaceutical care process:

- Identification of drug-related problems
- Therapeutic decision-making
- Follow-up and monitoring plan

Chatbot responses were independently assessed by five clinical pharmacy experts, comprising faculty members and hospital pharmacists actively engaged in pharmaceutical care. A rubric-based assessment adapted from the American Society of Health-System Pharmacists (ASHP) guidelines²² was employed. The rubric encompassed four competency domains:

1. Classification of clinical inquiries
2. Appropriateness of reference sources
3. Application of evidence-based data
4. Completeness and clarity of information communicated

Scores within each domain were recorded using a standardized scale and summarized descriptively, reported as mean $\pm$ SD to reflect variability across raters. The rubric was designed to capture competency-oriented expert judgment rather than to support inferential statistical testing.

Instrument Development and Validation

The sample size for each MCQ set ($n = 120$) was determined using a one-proportion sample size calculation, based on an expected accuracy of 54% and a 10% margin of error. The expected accuracy was selected within the general range of 50–55% reported in comparable health professional licensure examinations²³, which aligns closely with the passing threshold of the PLE-PC1 written examination (60%). This provided a conservative, yet relevant, benchmark for the study. The margin of error of 10% yielded a required sample size of approximately 96 items, which is comparable to the actual number of

MCQs in the PLE-PC1 (120 items), thereby supporting the appropriateness of this assumption.

Content validity for both the MCQs and the case study was evaluated using the Index of Item-Objective Congruence (IOC) by a panel of experts in clinical pharmacy and assessment design. Revisions were made according to their feedback.

Data Analysis

Quantitative data were analyzed by comparing overall scores across chatbot models and between two examination formats:

- **Standard format:** MCQs written mostly in Thai, with technical terms presented in varied ways and without a fixed rule for item construction. Exceptions were made for drug names, disease names, and technical terms without common Thai equivalents. The scores obtained from this format are referred to as the **Standard Score**.

- **ETT format:** MCQs written in Thai with English technical terms are systematically added in parentheses following each Thai term, e.g., “อันตรกิริยา” (drug interaction). This format reflects the Enhanced Task Translation (ETT) approach, which provides bilingual cues to support comprehension. The scores obtained from this format are referred to as the **Enhanced Score**.

This distinction enabled direct comparison of chatbot performance between a purely Thai format and a bilingual Thai+English format, while controlling for content validity and cognitive level.

Clinical case evaluation scores were derived from independent assessments by five clinical pharmacy experts. Each expert applied a standardized rubric to rate chatbot responses across therapeutic domains. To account for variability among raters, scores were summarized descriptively and reported as mean $\pm$ SD, which was appropriate given the small number of evaluators.

The research protocol was approved by the Ethics Committee for Human Research of Silpakorn University (COE 67.1101-073, November 1, 2024).

Results

This study evaluated and compared the competency of four AI chatbots—ChatGPT, ChatGPT Pro, Perplexity, and Perplexity Pro—in delivering drug information in Thai. The evaluation was conducted in two components: (1) multiple-choice question (MCQ) testing and (2) qualitative analysis through a clinical case study.

1. MCQ-Based Evaluation

Each chatbot was assessed using two sets of a 120-item MCQ, developed in accordance with the Pharmacy Council of Thailand's examination framework.

The first set contained standard Thai-language questions, while the second used the same questions with English technical terms added in parentheses to simulate Enhanced Task Translation (ETT). Performance was measured using percentage accuracy and weighted accuracy.

All four chatbots surpassed the 60% minimum passing threshold established by the national licensing standard. Perplexity Pro achieved the highest overall accuracy in both the standard (83.33%) and enhanced (80.83%) formats (Table 1 and Figure 1).

The accuracy rates of all AI chatbots in both standard and enhanced MCQ formats exceeded the 60% minimum passing threshold established by the Pharmacy Council of Thailand for licensure examinations, reflecting baseline competency in drug information services and evidence-based reasoning that are expected of entry-level pharmacists.

Subsequent analysis of weighted accuracy across 17 therapeutic areas and 17 pharmaceutical knowledge domains revealed consistently high performance in the infectious diseases and therapeutics categories across all models (Figures 2–3).

Table 1 Accuracy of AI Chatbots in Standard and Enhanced MCQ Formats

AI Chatbot	Items	Standard Score	Accuracy (%)	Enhanced Score	Accuracy (%)
ChatGPT	120	95.00	79.17	94.00	78.33
ChatGPT Pro	120	91.00	75.83	94.00	78.33
Perplexity	120	86.50	72.08	88.50	73.75
Perplexity Pro	120	100.00	83.33	97.00	80.83

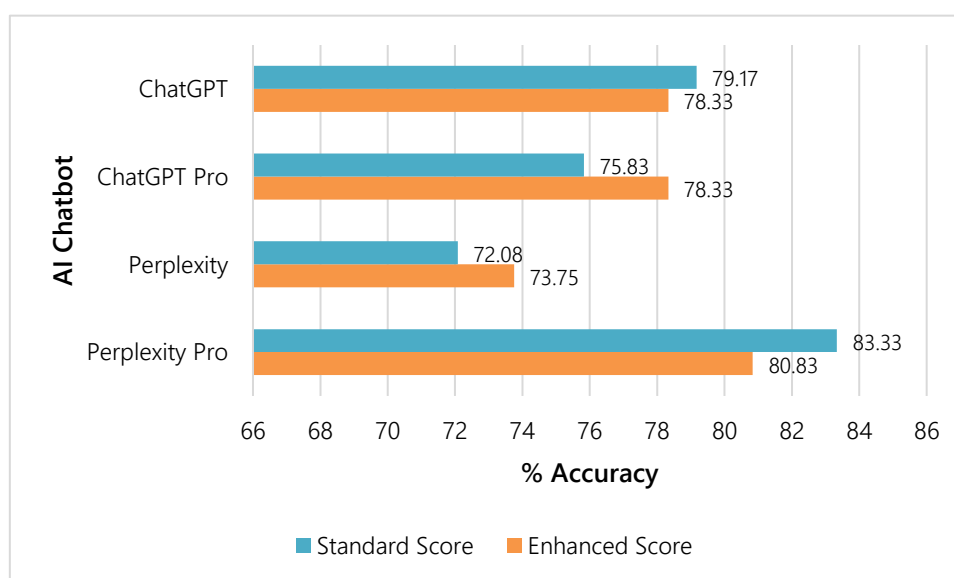


Figure 1 Overall accuracy (%) by chatbot

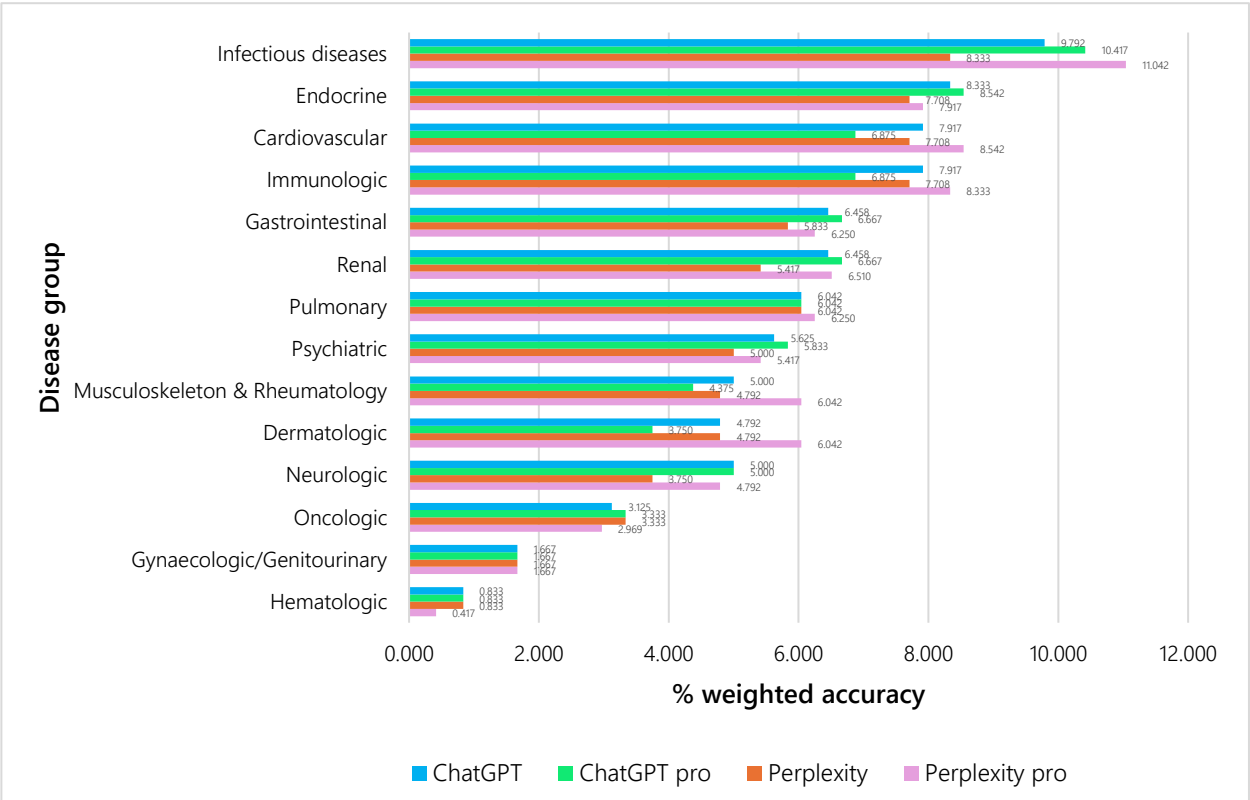


Figure 2 Weighted accuracy by therapeutic category

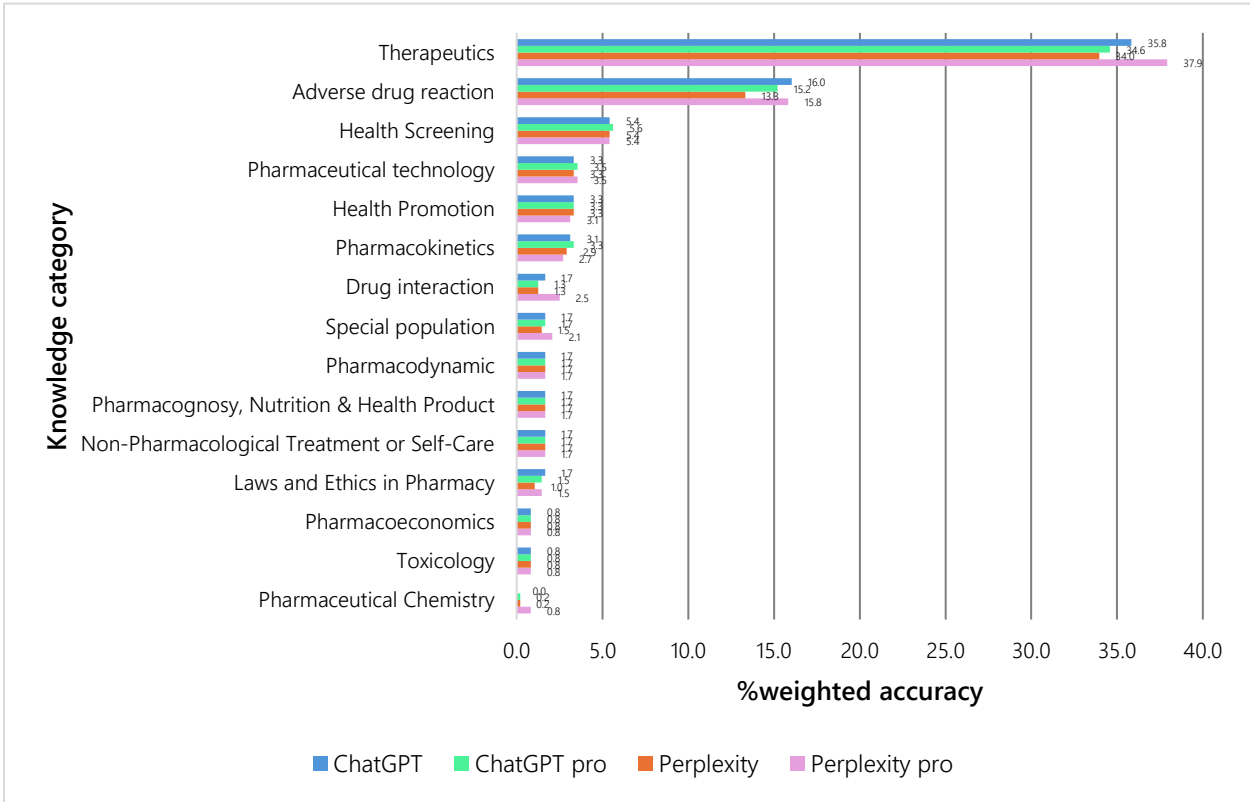


Figure 3 Weighted accuracy by knowledge domain

2. Clinical Case Evaluation

A cardiology case scenario was used to assess clinical reasoning and information provision. Responses were structured in SOAP-note format and independently evaluated by four clinical pharmacists who practiced in hospital settings using a rubric based on the ASHP guidelines. The rubric assessed four dimensions: question classification, source referencing, evidence application, and communication clarity.

ChatGPT Pro received the highest overall rubric score ($11.06 \pm 0.83/12$), outperforming the other models, particularly in the domains of reference use and evidence-based application ($3.00 \pm 0.00/3$ and $2.63 \pm 0.32/3$, respectively). (Table 2).

Discussion

This study demonstrates that all four AI chatbots—ChatGPT, ChatGPT Pro, Perplexity, and Perplexity Pro—achieved performance levels comparable to the minimum professional competency threshold for Thai pharmacists, as defined by the national licensure examination ($\geq 60\%$ passing criteria). Notably, Perplexity Pro achieved the highest overall scores in both the standard (83.33%) and enhanced (80.83%) multiple-choice formats.

Interestingly, the inclusion of English technical terms was observed that the use of either Thai or English did not result in differences in AI performance. Application of the Enhanced Task Translation (ETT) technique confirmed that the quality of responses was not influenced by the language of input. For instance, ChatGPT Pro and

Perplexity scored lower on the standard Thai-only version, suggesting that these models possess moderate proficiency in processing Thai-language inputs. This outcome may reflect limitations in bilingual fine-tuning or context-switching across languages.^{24,25}

Subgroup analyses revealed that the infectious diseases and therapeutics domains were the consistently highest performing areas across all AI models. This may be attributed to the abundance of open-access training data available in these fields, such as WHO guidelines and evidence-based protocols.^{26,27}

In the clinical case analysis, ChatGPT Pro outperformed the other models, particularly in reference citation and evidence application, achieving the highest rubric score. However, other models showed limitations in context interpretation and communication, despite meeting the minimum competency threshold. These findings align with prior literature indicating variable AI performance, which can be influenced by language, training data, and healthcare context.^{16, 28}

Therefore, AI competency should not be generalized across all settings. Instead, it requires localized evaluation tools that reflect national language, practice standards, and clinical expectations.

Another limitation of this study is the reliance on a single specialty case for chatbot evaluation. Cardiovascular diseases, as non-communicable diseases (NCDs), are both highly prevalent and extensively studied. This evidence density may have

Table 2 Rubric Scores by AI Chatbot (maximum score per evaluation criteria = 3), reported as mean $\pm$ SD

Evaluation Criteria	ChatGPT	ChatGPT Pro	Perplexity	Perplexity Pro
Question Classification	2.81 ± 0.24	2.88 ± 0.25	2.63 ± 0.48	2.75 ± 0.50
Source Referencing	2.75 ± 0.35	3.00 ± 0.00	2.63 ± 0.43	2.50 ± 0.35
Evidence Application	1.94 ± 0.52	2.63 ± 0.32	2.00 ± 0.54	2.13 ± 0.43
Communication Clarity	2.06 ± 0.38	2.56 ± 0.38	2.19 ± 0.43	1.88 ± 0.52
Total (score = 12)	9.56 ± 1.25	11.06 ± 0.83	9.44 ± 1.51	9.25 ± 1.54

contributed to higher accuracy of chatbot responses and greater availability of supporting references compared to less-studied disease areas. Such factors could potentially bias performance outcomes and limit the generalizability of findings to other therapeutic domains. To strengthen external validity, future research should expand to include multiple disease specialties, thereby providing a broader and more balanced assessment of chatbot performance across diverse clinical contexts.

Recommendations

1. Localized Competency Frameworks:

Countries where English is not the primary language should consider adopting the evaluation framework used in this study. By combining national competency-based MCQs with contextualized clinical scenarios, AI chatbot performance can be systematically assessed within the specific healthcare context of each country.

2. Strategic Use of ETT: Countries where English is not the primary language should first evaluate the competency of AI within the context of their local language before its practical implementation.

3. National Regulatory Guidance: Health regulatory bodies and professional institutions should consider developing national-level AI evaluation guidelines. These guidelines should assess accuracy, clinical appropriateness, and potential impact, using national licensing-style assessments combined with rubric-based methods to capture both knowledge and clinical judgment.

Conclusion

1. This study demonstrates that all four AI chatbots—ChatGPT, ChatGPT Pro, Perplexity, and Perplexity Pro—exhibited competency levels in drug information provision that aligned to the national professional standards. Among them, Perplexity Pro achieved the highest accuracy scores in both the

standard Thai-language multiple-choice test and the enhanced version that included English technical terms.

2. Although AI chatbot models are predominantly developed in English, the application of the ETT technique may allow non-English-speaking countries to recognize potential language effects on AI performance and to design effective strategies for their integration into practice.

3. In the qualitative clinical case analysis, ChatGPT Pro performed best, particularly in evidence-based reasoning and clinical communication, highlighting its relative strength in complex decision-making scenarios. This finding underscores the potential of AI to assist in generating a preliminary outline of responses within drug information services, thereby saving pharmacists' time and enhancing workflow efficiency.

4. These findings support the adoption of localized, competency-based evaluation frameworks for AI chatbots, especially in non-English-speaking countries. The proposed framework—combining standardized national examination questions with real-world clinical scenarios—can serve as a practical tool for assessing the readiness and safety of AI tools in pharmacy practice.

Nevertheless, AI should be regarded only as a supportive tool for pharmacists. Clinical judgment must always remain under the responsibility of pharmacists to ensure patient safety.

Acknowledgements

The authors gratefully acknowledge the insightful guidance and precise English-language editing provided by Associate Professor Dr. Waranya Wongwit

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

Partial funding for this research was provided by the Faculty of Pharmacy, Silpakorn University.

References

1. The Pharmaceutical Group of the European Union (PGEU). Pharmacists workforce shortages: PGEU position paper [Internet]. 2024 [cited 2025 Feb 14]. Available from: <https://www.pgeu.eu/wp-content/uploads/2024/11/PGEU-position-on-Pharmacists-Workforce-Shortages.pdf>
2. Arya V, Doucette WR, Kreling DH, Nadi S, Bakken BK, Gaither CA, et al. 2024 National pharmacist workforce study: Final report [Internet]. 2025 [cited 2025 Sep 24]. Available from: <https://www.aacp.org/sites/default/files/2025-06/2024-npws-final-report-5.27.25.pdf>
3. Kuruvilla AV, Madhan R, Chandagal Puttaswamy M. Clinical pharmacist-initiated assessment and amelioration of appropriate antibiotic use in surgical units at a South Indian tertiary care hospital - A handshake approach. *J Infect Dev Ctries*. 2023;17(1):66-72.
4. Mekurianew Kelkay J, Wubneh HD, Beri HM, Tefera AM, Molla RA, Negatu AA. Adoption of telepharmacy among pharmacists, physicians, and nurses at Hawassa City Public Hospitals, Ethiopia. *PLOS Digit Health*. 2024;3(12):e0000693.
5. Obua TO, Adome RO, Kutyabami P, Kitutu FE, Kamba PF. Factors associated with occupancy of pharmacist positions in public sector hospitals in Uganda: A cross-sectional study. *Hum Resour Health*;15(1):1.
6. Liao HC. Feature: Big hospitals are facing a shortage of pharmacists [Internet]. 2025 [cited 2025 Sep 24]. Available from: <https://www.taipetimes.com/News/taiwan/archives/2025/05/04/2003836302>
7. Hadsoong W, Pamonsinlapatham P. LINE chatbot for cosmetic advertising information using Dialogflow and Google Sheets. *Thai Bull Pharm Sci*. 2025;20(2):145-59. (in Thai)
8. Pamonsinlapatham K. 'Miss Clinic' AI chatbot assistant for answering questions and inquiries of medical facility operators in Nakhon Pathom Province. *Thai Bull Pharm Sci*. 2026;21(1):111-24. (in Thai)
9. Suriyaamporn P, Sila-on W, Kansom T, Opanasopit P. The application of artificial intelligence in pharmaceutical technology and advancements in the pharmaceutical industry 4.0. *Thai Bull Pharm Sci*. 2025;20(1):17-36. (in Thai)
10. Meepradista Y, Chokchaitam S, Teerakulkittipong N, Sukpornsawan P. Assessing accuracy and completeness of drug information in mobile apps compared with Micromedex®: An AI-based study in Thai online communities. *Res Social Adm Pharm*. 2026;22(3):474-81.
11. Taesotikul S, Singhana W, Taesotikul T. ChatGPT vs pharmacy students in the pharmacotherapy time-limit test: A comparative study in Thailand. *Curr Pharm Teach Learn*. 2024;16(4):404-10.
12. Chinwong D, Penthinapong T, Chinwong S. Integrating ChatGPT for smoking cessation counseling practice in pharmacy education: A single-group quasi-experimental study. *Tob Induc Dis*. 2025;23:182.
13. Hayat J, Lari M, AlHerz M, Lari A. The utility and limitations of artificial intelligence-powered chatbots in healthcare. *Cureus*. 2024;16(11):e73127.
14. Ramadhani CN. Chatbots in pharmacy: A boon or a bane for patient care and pharmacy practice? *Sci Pharm*. 2023;2(3):1-23.
15. Andrikyan W, Sametinger SM, Kosfeld F, Jung-Poppe L, Fromm MF, Maas R, et al. Artificial intelligence-powered chatbots in search engines: a cross-sectional study on the quality and risks of drug information for patients. *BMJ Qual Saf*. 2025;34(2):100-9.
16. Jin Y, Chandra M, Verma G, Hu Y, De Choudhury M, Kumar S. Better to ask in English: Cross-lingual evaluation of large language models for healthcare queries. In: *Proceedings of the ACM Web Conference 2024 (WWW '24)*; 2024 May 13-17; Singapore. New York: ACM; 2024. p. 2627-38.
17. Si Y, Meng Y, Chen X, An R, Mao L, Li B, et al. Quality, safety and disparity of an AI chatbot in managing chronic diseases: Simulated patient experiments. *npj Digit Med*. 2025;8:574.
18. OpenAI. ChatGPT Pricing [Internet]. 2024 [cited 2025 Sep 24]. Available from: <https://openai.com/chatgpt/pricing>
19. Perplexity AI. Which Perplexity subscription plan is right for you? [Internet]. 2024 [cited 2025 Sep 24]. . Available from: <https://help.perplexity.ai/en/articles/which-perplexity-subscription-plan-is-right-for-you>

20. Pharmacy Council of Thailand. Announcement No. 6/2566: written examination requirements for pharmaceutical care competency (PLE-PC1) for pharmacist licensure [Internet]. 2023 [cited 2025 Sep 24]. Available from: https://www.plecenter.org/index.php?option=content_detail&view=detail&itemid=3233&catid=86 (in Thai)
21. Pharmacy Council of Thailand. Announcement No. 9/2566: examination guidelines for pharmaceutical care competency (PLE-PC2) for pharmacist licensure [Internet]. 2023 [cited 2025 Sep 24]. Available from: https://www.plecenter.org/index.php?option=content_detail&view=detail&itemid=3288&catid=86 (in Thai)
22. American Society of Health-System Pharmacists. ASHP guidelines on documenting pharmaceutical care in patient medical records. *Am J Health Syst Pharm*. 2003;60(7):705-7.
23. Zong H, Li J, Wu E, Wu R, Lu J, Shen B. Performance of ChatGPT on Chinese national medical licensing examinations: a five-year examination evaluation study for physicians, pharmacists and nurses. *BMC Med Educ*. 2024;24:143.
24. OpenAI. GPT-4 technical report. arXiv [Preprint]; 2023 Mar 15 [cited 2025 Feb 14]. Available from: <https://arxiv.org/abs/2303.08774>
25. Lovenia C, Cahyawijaya S, Winata GI, Mahendra R, Koto F, Purwarianti A, et al. SEACrowd: Southeast Asian benchmark for multilingual and multimodal evaluation. In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024)*; 2024 Nov; Miami, FL. Stroudsburg (PA): Association for Computational Linguistics; 2024. p. 5349-65.
26. Sallam M. ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)*. 2023;11(6):887.
27. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44-56.
28. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2(2):e0000198.