

Original Research Articles

# Brief Communication: More Chat about Chatbots Assisting with Routine Clinical Biochemistry Queries

Rucita Severaj, Nareshni Moodley and Verena Gounden\*

Department of Chemical Pathology, Inkosi Albert Luthuli Central Hospital, University of KwaZulu Natal and National Health Laboratory Service, 800 Vusi Mzimela Road, Cator Manor, Durban 4091, Kwa-Zulu Natal, South Africa

\* Correspondence: [verenagounden@yahoo.com](mailto:verenagounden@yahoo.com); Tel.: +353-9182800

**How To Cite:** Severaj, R.; Moodley, N.; Gounden, V. Brief Communication: More Chat about Chatbots Assisting with Routine Clinical Biochemistry Queries. *International Journal of Clinical and Translational Medicine* 2025, 1(2), 4. <https://doi.org/10.53941/ijctm.2025.1000011>.

Received: 9 March 2025

Accepted: 25 March 2025

Published: 9 April 2025

**Abstract:** Background: Artificial intelligence (AI) technologies have the potential to transform many aspects of patient care and is playing an increasing role in health care in diagnostics and patient management. Several studies have already demonstrated that AI can perform as well as or better than humans at key healthcare tasks. The aim of this study was to evaluate and compare the utility of the AI chatbots ChatGPT and Deepseek to answer everyday chemical pathology queries as handled by the registrars in the Chemical Pathology Department at Inkosi Albert Luthuli Central Hospital in Durban, South Africa. Method: All queries received by the registrars over a month period were documented and answers from traditional sources such as laboratory procedures and sample handbooks were documented. The queries were later asked to ChatGPT and Deepseek and further evaluated and the answers evaluated by two blinded senior pathologists in terms of suitability and accuracy of responses and key word recognition. Results: A total of 37 queries were asked to the chatbots. Based on average scoring of the two reviewers 97% (n = 36/37) of queries for ChatGPT responses and 73% (n = 27) for Deepseek were ranked 3 and above for suitability and accuracy. The poorly scoring responses from both Deepseek and ChatGPT related to questions that were very specific to the local laboratory or testing in the laboratory. Conclusion: A longer-term evaluation and verification of ChatGPT as a resource to assist with lab related queries is required. It may be a useful resource not only to the trainee chemical pathology registrar but to under-resourced health care settings where pathologist support may not be present.

**Keywords:** artificial intelligence; clinical biochemistry

## 1. Introduction

Artificial intelligence (AI) is revolutionizing many disciplines and its application in health care is being explored with recognition that it has tremendous potential to optimize healthcare and make it more efficient by improving diagnostics, detecting medical errors and reducing the burden of paperwork [1]. Chatbots are based on language patterns rather than data extracted from verified or objective databases. They use algorithms to draw from already generated human textual content to find the best matches to the queries generated by the user. This newer generation of chatbots are also designed to accept unstructured or non-standardised output [2,3].

ChatGPT is an open-source chatbot developed by OpenAI (San Francisco, CA, USA) and launched in November 2022. It has positioned itself as the fastest-growing consumer application in history with the hype surrounding the application being justified given that it is free, easy to use and able to authentically converse on many subjects. Additionally, ChatGPT's ability to engage people with human-like conversation highlights the



importance of language and communication to the human experience and can help forge relationships between patients and healthcare professionals [4].

Studies have demonstrated that AI algorithm systems like ChatGPT perform relatively well on knowledge-based tests despite the lack of domain specific training and performed at or near the passing threshold on the United States Medical Licensing Exam. However, it has been demonstrated that these systems are particularly bad at context and nuance, both of which are critical for safe and effective patient care [3].

AI is playing an increasing role in laboratory medicine where the ability to provide accurate, readily available and contextualised data is crucial. Much of the interest of AI in laboratory medicine is related to its use in interpretation of results, diagnosis and quality control systems [5]. Deepseek is a recently launched open-sourced large language model that has been marketed as a competitor to ChatGPT.

The aim of this study was to evaluate the utility of the AI chatbots ChatGPT and Deepseek to answer everyday clinical biochemistry queries and to compare the performance of these two systems by evaluation performed by experienced pathologists.

## 2. Materials and Methods

The study was carried out at the National Health Laboratory Service (NHLS) Department of Chemical Pathology at Inkosi Albert Luthuli Central Hospital, Durban, South Africa. This is a tertiary academic laboratory that is part of the NHLS which supplies laboratory testing to all public health care facilities, serving 80% of the national population. The department receives numerous queries from clinical staff telephonically, via email or in person. All queries received by the trainee pathologists (registrars) are documented with details of the clinician, the actual query received and how the query was answered. The answers utilised for such queries come from various resources such as (but not limited to) laboratory procedures, Laboratory Handbook, discussion with pathologists and internet search. These documented queries received over a four-week period from 1 March 2023 to 31 March 2023 were assessed retrospectively. Accounts were created for ChatGPT and Deepseek and the queries that were received during the study period were then asked to the chatbots over a one week period in February 2025 and the answers obtained from both were tabulated against those received from the traditional sources. Two blinded senior pathologists reviewed the responses from the chatbots against the responses from the usual references. ChatGPT and Deepseek responses were then ranked on a scale of 1–5 regarding suitability and accuracy of responses; with a scale of 1 being dangerous/not relevant/unable to answer, a scale of 2 meaning unacceptable practice, 3 being acceptable practice/scientifically sound, a scale of 4 means it gave 3 additional information points known to the typical junior resident and a scale of 5 means 4 or more additional relevant information that could be used as learning points was provided by the chatbot versus the trainee. Additionally, key words/phrases in the traditional response were identified in the AI generated responses and scored on a scale of 1 to 3 with 1 being no matching key words; 2 being 50% of keywords and 3 being all keywords present. Data review was carried out using Microsoft Excel and Medcalc (Medcalc version 23.2.1, Belgium) software”

Ethical considerations: This is a retrospective study involving reviewing of existing queries received by the laboratory, no related patient demographic data or any other identifying data for patients and personnel was recorded or utilised for the purpose of this study.

## 3. Results

A total of thirty-seven queries were documented over the study period. The findings for each reviewer are summarised in Table 1. Based on average scoring of the two reviewers 97% (n = 36/37) of queries for ChatGPT responses and 73% (n = 27) for Deepseek were ranked 3 and above for suitability and accuracy. For the ChatGPT responses more than 75% (reviewer 1—78.3%; reviewer 2—86.4%) of responses were judged to have a keyword matching score of 2 or more. This was found to be lower on assessment of the Deepseek responses for reviewer 1 (67.5%) but was similar for reviewer 2 (84.8%). It was noted that for ChatGPT only 3 responses (8%) received a key word score of 3 and a suitability score of >3 from both reviewers.

The poorly scoring responses from both ChatGPT and Deepseek were related to questions that were very specific to the local laboratory service, for example, if a test was performed in a certain lab and expected turnaround times). Notably no response was judged to be dangerous by any of the reviewers for both ChatGPT and Deepseek, but more responses were considered to be irrelevant by both the reviewers for Deepseek (Table 1). The highest rated responses across both chatbots were ones dealing with sample type, stability and collection queries. Between reviewers there was a concordance of 30% (n = 11) for suitability & accuracy scoring and 51% (n = 19) for key word scoring for the ChatGPT responses evaluated. The concordance between reviewers for the Deepseek responses was higher 46% (n = 17) and 67% (n = 25) for keyword scoring. The inter-rater agreement between

reviewers was good (kappa statistic 0.877) for Deepseek responses and poorer for the ChatGPT responses (kappa statistic <0.20) for accuracy and suitability evaluation (Table 2 shows examples of the responses received from both chatbots).

**Table 1.** Distribution of Scoring for ChatGPT and DeepSeek responses per reviewer.

Reviewer 1	Suitability and Accuracy Rating	ChatGPT Number of Responses (%)	Deepseek Number of Responses (%)	Keyword Rating	Chat GPT Number of Responses (%)	Deepseek Number of Responses
	1	2 (5.4%)	8 (21.6%)	1	8 (21.6%)	12 (32.4%)
	2	2 (5.4%)	5 (13.5%)	2	14 (37.8%)	13 (35.1%)
	3	7 (18.9%)	14 (37.8)	3	15 (40.5%)	12 (32.4%)
	4	15 (40.5%)	5 (13.5%)			
	5	11 (29.7%)	5 (13.5%)			

  

Reviewer 2	Suitability and Accuracy Rating	ChatGPT Number of Responses (%)	Deepseek Number of Responses (%)	Keyword Rating	ChatGPT Number of Responses (%)	Deepseek Number of Responses
	1	0 (0%)	9 (24.3%)	1	5 (13.5%)	13(35.1%)
	2	0 (0%)	2 (5.4%)	2	16 (43.2%)	9 (24.3%)
	3	22 (59.5%)	12 (32.4%)	3	16 (43.2%)	15 (40.5%)
	4	9 (24.3%)	2 (5.4%)			
	5	6 (16.2%)	12 (32.4%)			

**Table 2.** Examples of queries asked to the chatbots.

Examples of Queries That Scored <3 for Accuracy and Suitability and <2 for Keyword Assessment by Both Reviewers					
Query	Answer to Query	Traditional Method of Determining Response to Query	Chat GPT Response	Deepseek Response	
Can I still add a request for polyethylene-glycol (PEG) precipitation on a prolactin sample received in the lab 6 weeks ago?	No, the samples get discarded after a month of storage in freezer	Laboratory user guide and PEG SOP	<p>Recommended Action:</p> <ul style="list-style-type: none"> <li>• Yes, you can add PEG but be cautious and interpret results with consideration of potential protein degradation.</li> <li>• If results appear unusual, consider re-running the test with a fresh sample.</li> </ul> <p>Would you like protocol details or troubleshooting steps for PEG precipitation?</p>	<p>Yes, you can still add PEG to a prolactin sample that has been stored in the freezer for 6 weeks, provided the sample was stored properly and shows no signs of degradation. However, if there are concerns about sample integrity, it may be worth considering a fresh sample for accurate results. Always validate the assay after PEG treatment to ensure reliability.</p> <p>If this is for clinical or diagnostic purposes, consult with a laboratory specialist or chemical pathologist to ensure the best approach</p>	
Is Vitamin B6 testing offered by the IALCH NHLS laboratory?	No, this is a referral test out of the NHLS laboratory service to a	NHLS national lab user guide	The National Health Laboratory Service (NHLS) in South Africa offers a wide range of diagnostic tests. However, specific information about the availability of	The National Health Laboratory Service (NHLS) in South Africa provides a wide range of laboratory tests, including many specialized tests for vitamins and nutrients. However, the availability of specific tests, such as vitamin B6	

private laboratory in South Africa

vitamin B6 (pyridoxal 5'-phosphate) testing is not readily accessible online.

For accurate and up-to-date information, it's recommended to contact your local NHLS laboratory directly. They can provide details on test availability, sample requirements, and any necessary preparations. If the NHLS does not offer this test, they may be able to refer you to an alternative facility or provide guidance on how to proceed.

For general context, vitamin B6 testing typically involves measuring pyridoxal 5'-phosphate levels in the blood to assess nutritional status or diagnose deficiencies. Proper sample handling, such as protecting the specimen from light and processing it promptly, is crucial for accurate results.

(pyridoxine) testing, may depend on the particular laboratory or facility within the NHLS network.

Key Points to Consider:

1. Vitamin B6 Testing:
  - Vitamin B6 testing is not as commonly requested as tests for other vitamins (e.g., vitamin D or B12), so it may not be routinely offered at all NHLS laboratories.
  - Specialized tests like vitamin B6 may require sending the sample to a central or reference laboratory within the NHLS network.
2. How to Confirm:
  - Contact your local NHLS laboratory or healthcare provider to confirm whether vitamin B6 testing is available.
  - If the test is not routinely offered, the NHLS may be able to arrange for the sample to be sent to a reference laboratory that performs the test.
3. Alternative Options:
  - If the NHLS does not offer vitamin B6 testing, private laboratories in South Africa may provide this service. Examples include Pathcare, Ampath, or Lancet Laboratories.
4. Clinical Indications:
  - Vitamin B6 testing is typically requested in cases of suspected deficiency (e.g., in patients with malnutrition, malabsorption, or certain metabolic disorders) or toxicity (e.g., due to excessive supplementation).

Steps to Take:

- Consult your healthcare provider to determine if vitamin B6 testing is necessary.
- Ask your provider to contact the NHLS to confirm the availability of the test and the appropriate sample collection and handling procedures.

If you need more specific or up-to-date information, I recommend reaching out directly to the NHLS or your healthcare provider.

**Example of query that scored >3 for accuracy and suitability and ≥2 for keyword assessment by both reviewers**

Query	Answer to Query	Traditional Method of Determining Response to Query	Chat GPT Response	Deepseek Response
What is the colour of the tube for ACTH collection?	Collect sample in K-EDTA tube and transport on ice to the laboratory immediately	Laboratory user guide	<p>Adrenocorticotrophic Hormone (ACTH) testing requires a lavender (purple) top tube, which contains EDTA (Ethylenediaminetetraacetic acid) as an anticoagulant.</p> <p>Key Handling Requirements for ACTH Sample:</p> <ol style="list-style-type: none"> <li>Pre-Chilled Tube: <ul style="list-style-type: none"> <li>Ideally, the collection tube should be pre-chilled to preserve ACTH stability.</li> </ul> </li> <li>Immediate Processing: <ul style="list-style-type: none"> <li>The sample must be placed on ice immediately after collection to prevent ACTH degradation.</li> </ul> </li> <li>Centrifugation &amp; Plasma Separation: <ul style="list-style-type: none"> <li>The sample should be centrifuged as soon as possible (within 30 min) at 4 °C, and plasma must be separated and frozen at -20 °C or lower if not analyzed immediately.</li> </ul> </li> <li>Avoid Hemolysis: <ul style="list-style-type: none"> <li>Hemolysis can affect ACTH results, so careful handling is crucial.</li> </ul> </li> </ol> <p>Would you like a step-by-step protocol for ACTH sample handling</p>	<p>The adrenocorticotrophic hormone (ACTH) test requires specific sample collection and handling procedures to ensure accurate results. The blood sample for ACTH must be collected in a pre-chilled EDTA tube (typically a lavender-top tube) and kept on ice immediately after collection. Here's why:</p> <p>Key Points for ACTH Sample Collection:</p> <ol style="list-style-type: none"> <li>Tube Type: <ul style="list-style-type: none"> <li>Lavender-top tube (EDTA tube): This tube contains the anticoagulant ethylenediaminetetraacetic acid (EDTA), which prevents clotting and stabilizes the sample.</li> <li>The tube must be pre-chilled before use to minimize degradation of ACTH, which is a very labile hormone.</li> </ul> </li> <li>Handling: <ul style="list-style-type: none"> <li>After collection, the tube must be placed on ice immediately and transported to the lab as soon as possible.</li> <li>The sample should be centrifuged at 4 °C (refrigerated centrifuge) to separate plasma from blood cel</li> </ul> </li> <li>Storage: <ul style="list-style-type: none"> <li>If the sample cannot be processed immediately, the plasma should be frozen at -20 °C or lower until analysis.</li> </ul> </li> </ol> <p>Why These Precautions?</p> <ul style="list-style-type: none"> <li>ACTH is highly unstable and can degrade rapidly at room temperature or in the presence of proteases in the blood.</li> </ul>

- Pre-chilling the tube and keeping the sample on ice helps preserve the hormone’s integrity until it can be processed.

Common Errors to Avoid:

- Using a non-chilled tube.
- Delaying transportation or processing of the sample.
- Allowing the sample to warm up before centrifugation.

If these precautions are not followed, the ACTH levels measured may be artificially low, leading to inaccurate diagnostic results. Always follow your laboratory’s specific protocols for ACTH testing

**Examples of queries where discrepant answers noted between both ChatBots**

Query	Answer to Query	Traditional Method of Determining Response to Query	Chat GPT Response	Deepseek Response
Can cardiac markers (Trop I and CKMB) be performed on yellow top tubes	Yes	Laboratory user guide, relevant test SOP and manufacturer IFU	<p>Cardiac markers like Troponin I (Trop I) and CK-MB are not ideally performed on yellow-top tubes (SST—Serum Separator Tubes) because of the gel separator, which can sometimes interfere with results.</p> <p>Preferred Tubes for Cardiac Markers:</p> <ol style="list-style-type: none"> <li>1. Troponin I (Trop I)                             <ul style="list-style-type: none"> <li>○ Preferred: Green top (Lithium Heparin—Plasma) or Red top (Plain—Serum)</li> <li>○ Acceptable: Yellow top (SST) if serum is separated promptly after clotting.</li> </ul> </li> <li>2. CK-MB (Creatine Kinase MB Isoenzyme)                             <ul style="list-style-type: none"> <li>○ Preferred: Green top (Lithium Heparin—Plasma) or Red top (Plain—Serum)</li> <li>○ Acceptable: Yellow top (SST) if serum is separated immediately after centrifugation.</li> </ul> </li> </ol> <p>Why Yellow Top Tubes Are Not Ideal:</p> <ul style="list-style-type: none"> <li>• SST tubes contain a gel separator, which can adsorb some proteins and alter test results if the sample is not separated promptly.</li> </ul>	<p>Yes, cardiac markers (Troponin I and CK-MB) can generally be performed on yellow-top tubes, as long as the sample is handled and processed according to standard laboratory protocols. Always confirm with the laboratory performing the test to ensure compatibility with their specific assays</p>

- Delayed processing may lead to degradation or diffusion of analytes.

Best Practice:

- If using a yellow-top SST tube, ensure immediate centrifugation and separate the serum to avoid interference.
- Green-top lithium heparin tubes (without gel) are preferred for faster processing in emergencies.

Would you like guidance on sample handling to ensure accurate results

Is SCA genetic testing available in South Africa/NHLS

Yes—performed at Groote Schuur Genetics

NHLS and referral lab user guide, Lab Referral test SOP

Spinocerebellar ataxia (SCA) testing is available in South Africa. The Division of Human Genetics at the University of Cape Town (UCT) offers genetic testing for several SCA subtypes, including SCA1, SCA2, SCA3, SCA6, and SCA7.

UCT Health Sciences

Regarding the National Health Laboratory Service (NHLS), specific information about SCA testing availability—is not readily accessible online. For detailed information, it’s advisable to contact the NHLS directly or consult with a healthcare professional who can provide guidance on testing options within the NHLS network.

It’s important to note that SCA7 has been observed predominantly in families of indigenous Black African descent in South Africa.

PubMed

For comprehensive genetic counseling and testing services, reaching out to specialized centers like UCT’s Division of Human Genetics is recommended.

Spinocerebellar ataxia testing is available in South Africa, primarily through private laboratories and academic institutions. For comprehensive testing, samples may be referred to international labs. If you or a patient requires SCA testing, consult a neurologist or geneticist to initiate the process and ensure proper genetic counseling is provided. While the NHLS may not directly offer spinocerebellar ataxia testing, they can facilitate the process by collecting samples and referring them to specialized laboratories. For definitive testing, private labs or international reference labs are typically used. If you or a patient requires SCA testing, consult a neurologist or geneticist to initiate the process and determine the best testing pathway.

#### 4. Discussion

The applications of AI in healthcare and laboratory medicine are many and when used appropriately can revolutionize and optimize the way we practice medicine. Much of the focus has been in implementing AI in diagnostics and interpretation of results as well as quality management, however AI can also be implemented in communications. Being a language model, one of its most impressive abilities is the ability to reproduce human language in a convincing and human-like way. This can potentially be used to assist communications between patients and healthcare workers but also between healthcare professionals. Pathologists and other laboratory staff are the interface between clinicians and the laboratory and effective communication is essential to ensure good understanding and productive relationships between disciplines for the benefit of patient care. ChatGPT has been one of the applications on the forefront of language-based AI tools and has gained its reputation as it is easy to use, and some have found it to generate appropriate and contextually relevant responses across a broad spectrum of prompts [6]. Other studies have found that while ChatGPT performs well in knowledge-based assessments, its performance is subpar in situations where context and nuance is required [3]. Deepseek is relatively new and no studies examining similar knowledge based assessments were available at the time of this study.

There are also some ethical issues to consider regarding conversational AI in medical practice. Training a model requires a large amount of high-quality data, and current algorithms are often trained on biased data sets. The models are susceptible to availability, selection and confirmation bias. Other ethical issues are related to legal responsibility and who is to blame if AI makes a mistake [3].

Clinical personnel from various levels of healthcare and specialties contact the laboratory departments with a variety of queries such, as how to investigate a patient in a particular clinical circumstance, specimen requirements for a particular test and if a particular test is offered by the laboratory. These queries often require consultation with the pathologists, review of standard operating procedures and laboratory handbooks and internet searches, which may at times be time consuming. This study investigated the utility of ChatGPT and Deepseek to answer these queries and compared the AIs' answers to those acquired by the registrars/trainees within the department. We found that majority of responses from both chatbots were in keeping with acceptable practice and was scientifically sound. Importantly, none of the responses were found to be dangerous and the poorly scoring responses were all related to highly specific, institution dependent queries. Notably, the most highly rating response were related specimen requirements and collection procedures which is relatively universal. These findings support other findings that noted ChatGPT performed best in knowledge-based areas rather than areas that require contextual information.

Limitations of this study include the use of only 2 blinded reviewers and the short study period.

#### 5. Conclusions

A longer term and more thorough evaluation and ongoing verification of these chatbots as a resource to assist with laboratory related queries is required. This should be done across laboratory departments in different settings as they may be a useful resource not only to the clinical chemistry trainee but to under-resourced health care settings where pathologist support may not be readily available.

#### Author Contributions

V.G. and N.M. generated the idea for this publication. R.S. and V.G. undertook the statistical analyses. R.S. generated the original version and V.G. edited multiple iterations. All authors have read and agreed to the published version of the manuscript.

#### Funding

This research received no external funding.

#### Institutional Review Board Statement

Not applicable.

#### Informed Consent Statement

Not applicable.

#### Data Availability Statement

The data is available from the senior author for review on reasonable request.

### Conflicts of Interest

The authors declare no conflict of interest.

### Use of AI and AI-Assisted Technologies

These were not used by the authors submitting this paper.

### References

1. Homolak, J. Opportunities and risks of ChatGPT in medicine, science, and academic publishing: A modern Promethean dilemma. *Croat. Med. J.* **2023**, *64*, 1–3. <https://doi.org/10.3325/cmj.2023.64.1>.
2. Sng, G.G.R.; Tung, J.Y.M.; Lim, D.Y.Z.; et al. Potential and Pitfalls of ChatGPT and natural language artificial intelligence models for Diabetes Education. *Diabetes Care* **2023**, *10*, e1–e3.
3. Bibault, J.-E.; Chaix, B.; Guillemasse, A.; et al. A chatbot versus physicians to provide information for patients with breast cancer: Blind, randomized controlled noninferiority trial. *J. Med. Internet Res.* **2019**, *21*, e15787.
4. Editorial: Will ChatGPT transform healthcare? *Nat. Med.* **2023**, *29*, 505–506. <https://doi.org/10.1038/s41591-023-02289-5>.
5. Paranjape, K.; Schinkel, M.; Hammer, R.D.; et al. The Value of Artificial Intelligence in Laboratory Medicine: Current Opinions and Barriers to Implementation. *Am. J. Clin. Pathol.* **2021**, *155*, 823–831. <https://doi.org/10.1093/ajcp/aqaa170>.
6. Dave, T.; Athaluri, S.A.; Singh, S. ChatGPT in medicine: An overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front. Artif. Intell.* **2023**, *6*, 1169595. <https://doi.org/10.3389/frai.2023.1169595>.