ORIGINAL RESEARCH

DOI: 10.4274/globecc.galenos.2024.06025
Glob Emerg Crit Care 2024;3(3):132-139

Assessing the Performance of ChatGPT in Medical Toxicology Through Simulated Case Scenarios

İbrahim Altundağ[1], Semih Korkut[2], Ramazan Güven[2], Aynur Şahin[2]

[1]University of Health Sciences Türkiye, Haydarpaşa Numune Training and Research Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye
[2]University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye

Abstract

Objective: The insufficient number of medical toxicologists and poison information centers worldwide limits the accessibility of adequate medical recommendations for the management of poisoned patients. This study aimed to assess the effectiveness of Chat Generative Pretrained Transformers (GPTs) medical recommendations in medical toxicology and evaluate its accuracy as a valuable resource when accessing medical toxicologists or poison information centers is limited.

Materials and Methods: A toxicologist created 10 different toxicology-simulated case scenarios based on the possible presentations of poisoned patients in an emergency department setting. The categories of general approach and stabilization, diagnostic activities, and medical treatments and follow-up were used to measure case assessment and ChatGPT's medical recommendation capacity.

Results: ChatGPT-4o achieved an average success rate of 90.88% across the simulated case scenarios. ChatGPT-4o received a passing grade in 9 cases (90%) and received "improvable" in only 1 case (10%). ChatGPT-4o's average success rate in all categories and across all cases increased from 90.88% to 97.22% with the secondary test.

Conclusion: Our study indicates that it is possible to improve the success rate of ChatGPT in providing medical toxicology recommendations. The ability to query current medical toxicology information through ChatGPT-4o demonstrates the potential of ChatGPT to serve as a next-generation poison information center.

Keywords: Artificial intelligence (AI), ChatGPT-4o, clinical decision support systems, generative pretrained transformer, poison control center, toxicology

Introduction

The significant mortality and illness rates resulting from poisonings, particularly in developing countries, have made poisoning an escalating public health issue requiring specialized attention and care. The establishment of medical toxicology as a subspecialty, along with the increasing prevalence of toxicologists and poison information centers, can facilitate a decrease in mortality and morbidity rates among poisoned patients [1]. This approach can also help reduce unnecessary hospital admissions and shorten hospitalization periods. However, the insufficient numbers of medical toxicologists and poison information centers worldwide limit the access of poisoned patients to adequate medical support.

One of the main objectives of developing technology and computer systems is to eliminating dependence on human labor and create autonomous systems. Natural language processing (NLP) enables computers to comprehend texts and spoken words in a manner similar to humans [2]. NLP technology has made a noteworthy contribution to the advancement of clinical decision support (CDS) systems. CDS systems are designed to ensure complete, timely, efficient, and accurate data presentation in the healthcare industry, with

Address for Correspondence: İbrahim Altundağ, University of Health Sciences Türkiye, Haydarpaşa Numune Training and Research Hospital, Clinic of Emergency Medicine, İstanbul, Türkiye
E-mail: dr.ibrahimaltundag@gmail.com **ORCID-ID:** orcid.org/0000-0002-0880-7218
Received: 27.07.2024 **Accepted:** 11.09.2024

the aim of supporting healthcare professionals in making the most accurate decisions [3]. Currently, although not usable for CDS, a notable example of an NLP model is the Generative Pretrained Transformer (GPT) model [4]. The popularity of artificial intelligence (AI) has increased recently in mainstream media and literature, with the emergence of GPT-3, a language model capable of producing human-like text [5]. The latest version of this system, ChatGPT-4o, was made available for free use on May 13, 2024.

Although the existence of numerous sources of information related to medicine and healthcare is an indication of advanced medical literature, it complicates accessing accurate information. Despite the widespread use of the internet, there can still be challenges in obtaining specific and accurate information. This is particularly vital for clinicians operating in chaotic settings, such as emergency departments, where swift access to accurate information is of significant importance.

The aim of this study was to test the feasibility of AI applications (ChatGPT-4o) in the field of medical toxicology using simulated case scenarios. Our study is significant because it is the first to assess the utility of AI in the analysis and management of toxicology in patients encountered in daily clinical practice. In contrast to previous AI studies, our investigation employed realistic case scenarios involving patients who might present to the emergency department. The effectiveness of ChatGPT's medical recommendation in medical toxicology has been evaluated to assess its accuracy as a resource when accessing a medical toxicologist and poison information centers are limited.

Materials and Methods

This study was conducted at University of Health Sciences Türkiye, Çam and Sakura City Hospital, Clinic of Emergency Medicine, Medical Toxicology Intensive Care Unit. Since the study involved the use of simulated case scenarios (the data of any animals, patient or patient group was not used), Ethics Committee approval and individual consent were not required.

Study Design

This study was designed to test the usability of AI applications (ChatGPT) in medical toxicology. The latest version of ChatGPT available for free use is ChatGPT-4o. Due to its ease of accessibility worldwide without cost, the ChatGPT-4o version was used. To measure ChatGPT-4o's level of knowledge about current patient approaches and practices in medical toxicology, we asked open-ended question-answer pairs based on case scenarios that specifically focused on commonly encountered important types of poisoning. ChatGPT-4o underwent testing in two stages, each consisting of 10 different simulated case scenarios tailored for individual common poisoning types, which encompassed questions related to frequently encountered poisoning types.

In the first stage, the interpretation of the case and the success of recommendations from ChatGPT-3.5 were tested using open-ended questions. In the second stage, the questions that were not successfully answered in the first test were converted into knowledge-based questions and presented to ChatGPT-4o for a second attempt, with the answers then scored. This enabled ChatGPT-4o to directly assess the level of medical toxicology knowledge, independent of NLP and CDS features (without the need for case analysis). The success of ChatGPT-4o for the posed questions and received answers was evaluated by an experienced, blinded medical toxicologist with at least 5 years of experience in medical toxicology. The accuracy rates of ChatGPT-4o were evaluated both at the case level and under three categories: general approach and stabilization, diagnostic activities, and medical treatments and follow-up.

Simulated Case Scenarios

A total of 10 different toxicology simulated case scenarios were designed by experienced toxicologists based on the possible presentation of poisoned patients in an emergency department setting. The case scenarios included overdose of acetaminophen, tricyclic antidepressant (TCA) overdose, methanol toxicity, organophosphate toxicity, digoxin poisoning, sympathomimetic toxidrome, lithium overdose, carbon monoxide (CO) poisoning, calcium channel blocker (CCB) overdose, and snake bite. An example of a simulated case scenario is shown in Figure 1. The simulated case scenarios are available in Supplemental Material-1.

Assessment and Statistical Analysis

ChatGPT's knowledge of poisoning was assessed using questions addressing the differential diagnosis, stabilization, diagnosis, treatment, and follow-up of simulated case scenarios. Examples of questions that assess the evaluation and recommendation capacity of ChatGPT-4o in simulated case scenarios are shown in Figure 2.

Each simulated case scenario was evaluated in three different categories. The categories of "general approach and stabilization," "diagnostic activities," and "medical treatments and follow-up" were used to measure the case assessment and ChatGPT-4o's medical recommendation capacity. A medical toxicologist evaluated each category of each simulated case scenario, awarding a maximum score of 100 points. Additionally, the medical toxicologist identified the key recommendations and points, assigning scores based on the importance of each answer. The simulation case scenarios and scoring table (scoring that shows only high-rated, key points) are shown in Table 1. The remaining points for parameters outside the key points were distributed equally. The scoring of the simulated case scenarios and the results was uploaded as Supplemental Material-2. Each of the three categories within a case was assessed independently to evaluate the accuracy

Case 4: Organophosphate Toxicity

42-year-old male patient weighing 70kg and measuring 175cm. No known medical conditions or medication use. He was working as a farmer and entered the field three hours ago. His friend became concerned when he didn't see him for a long time and found him lying on the ground when he entered the field. He noticed that this patient had vomited and there was foaming in his mouth, and his skin was sweaty. He was immediately transported to the hospital by ambulance. The other friend who entered the field also experienced headache, nausea/vomiting, and sweating, but recovered shortly after. This patient was intubated when his Glasgow Coma Scale (GCS) dropped from 12 in the ambulance to 8 in the emergency room. Significant secretion was observed from the mouth and tube. His overall condition was fair to poor. Vital signs revealed blood pressure of 90/50 mmHg, pulse rate of 40/min, respiratory rate of 15/min, oxygen saturation of 87%, and blood sugar level of 88 mg/dl. Bilateral rales were heard in lung sounds. Pupils were bilaterally myotic. His skin was diaphoretic. No pathology was observed in other system examinations.

Blood gas analysis showed pH of 7.34, bicarbonate of 22 mEq/L, lactate of 2.4 mmol/L, and BE of -4. Other laboratory parameters were normal. EKG showed sinus bradicardia of 42/min, QRS duration of 75ms, and QTc duration of 430ms. No pathology was observed in the brain tomography. Bilateral ground-glass opacities were present in thorax tomography.

Figure 1. A case presentation of a simulated case scenario example prepared in accordance with the poisoning table

1. What pathologies should be considered in the differential diagnosis?
2. What is the most likely diagnosis for this patient?
3. How should the treatment approach be in this patient? List the general principles of approach and what needs to be done?
4. Should decontamination methods (skin decontamination, gastric lavage, activated charcoal) be used for this patient?
5. What diagnostic tests should be ordered? Explain the reason for each diagnostic test.
6. Is there a specific diagnostic tool? When should it be used? How does it contribute to the treatment process?
7. How should I treat this patient? What treatment steps should be used?
8. Is there an antidote that can be used to treat this patient? If so, what is it and how is it administered? What's the dose of antidote?
9. Is there an indication for the use of elimination-enhancing methods (repeated dose of activated charcoal, urine alkalization, extracorporeal treatment)?

Figure 2. Example question list designed to assess the evaluation and recommendation capacity of ChatGPT in simulated case scenarios in medical toxicology

of ChatGPT-4o responses. Responses were scored according to a predefined scale to obtain category scores. The questions that were answered incorrectly in the first part were repeated in the second part. Here, the questions were presented as direct knowledge questions rather than case scenarios. Thus, ChatGPT-4o's knowledge level was directly assessed rather than its analytical thinking skills. The evaluation results for the second part (questions answered incorrectly in the first part) have been uploaded as Supplemental Material 3.

Table 1. The scoring system for key points used in the evaluation of simulated case scenarios

Simulated cases	General evaluation and stabilization		Diagnostic acitivies		Medical treatments and follow-up	
	Points and parameters		Points and parameters		Points and parameters	
1. Acetaminophen overdose	50	Recognition of APAP overdose	25	Serum acetaminophen level	25	Antidote=NAC
					25	Dose of NAC
			25	Request for serum APAP at 4 h	25	Serum ALT-AST-INR levels
2. TCA overdose	50	Recognition of TCA overdose	25	ECG	25	Antidote=bicarbonate
			25	Recognition of Na channel blockade	25	Dose of sodium bicarbonate
3. Methanol toxicity	50	Recognition of toxic alcohol ingestion	25	Serum ethanol level	25	Antidote = fomepizole antidote = ethanol
			25	Blood gas	25	Elimination using ECTR
4. Organophosphate toxicity	50	Recognition of OF poisoning	x	N/A	25	Antidote= atropine antidote=pralidoxime
					25	Atropinization
5. Digoxin poisoning	50	Recognition of digoxin poisoning	25	Serum digoxin level	25	Digifab / digibind
			25	ECG		
6. Sympathomimetic toxidrome	50	Recognition of sympathomimetic toxidrome	25	Serum CK level	25	IV fluid benzodiazepine
				UDS	25	Cooling methods
7. Lithium overdose	25	Ineffectivity of activated charcoal	25	Serum lithium level	25	Elimination using ECTR
	25	WBI	25	ECG	25	Indication for ECTR
8. CO poisoning	50	Recognition of CO poisoning	50	Blood gas	25	Non-rebreathing mask
					25	HBOT
9. CCB overdose	25	Decontamination	25	Echocardiography	15	Vasopressors
					15	IV calcium
			25	ECG	15	HDI
					15	ECMO
10. Snake bite	50	Snake bite grading	25	CBC, routine biochemistry	25	Antidote=anti-venom
	25	Supportive treatment of extremity edema	25	Serum CK level	25	Dose of anti-venom

ALT: Alanine aminotransferase, APAP: Acetaminophen (n-acetyl-p-aminophenol), CBC: Complete blood count, AST: Aspartate aminotransferase, CCB: Calcium channel blocker, CK: Creatinine kinase, CO: Carbon monoxide, ECG: Electrocardiography, ECMO: Extracorporeal membrane oxygenation, ECTR: Extracorporeal treatment, HBOT: Hyperbaric oxygen therapy, HDI: High-dose insulin treatment, INR: International normalized ratio, IV: Intravenous, NAC: N-acetylcysteine, OF: Organophosphate, TCA: Tricyclic anti-depressant, UDS: Urine drug screen, WBI: Whole bowel irrigation

The average case score was determined by computing the mean of the three category scores for each case. The mean score of cases was employed as an indicator of ChatGPT-4 performance on a case-by-case basis. Subsequently, the overall performance of ChatGPT-4o in simulated scenarios was evaluated by calculating the mean success score across all cases. Bloom's 80% cut-off value was used to determine the success of ChatGPT-4o [6]. According to Bloom's cut-off value, achieving a success rate of over 80% was considered a "success", a success rate between 60-80% was considered "improvable", and a success rate below 60% was considered a "failure". A performance with a correct answer rate exceeding 60% is defined as a passing grade.

Outcomes

The primary objective of this study was to evaluate the appropriateness of the medical recommendations provided by ChatGPT-4o regarding current medical toxicology practices. The secondary goal is to evaluate ChatGPT-4o's capacity to interpret medical toxicology. The accuracy of the answers to the questions was evaluated using medical toxicology guidelines and reference sources, as used by toxicologists [7,8].

Results

In this study, the medical toxicology knowledge level of ChatGPT-4o was tested through a two-stage process using 10

different simulated case scenarios. In the first part of the study, it was observed that ChatGPT-4o achieved an average success rate of 90.88% across the simulated case scenarios. ChatGPT-4o was successful in 9 cases (90%) and received a "improvable" in only 1 case (10%). ChatGPT-4o achieved an excellent success rate (100%) in cases of poisoning with TCA in the first part. ChatGPT-4o also achieved the second highest success rate in the case of methanol toxicity, with an average success rate of 96.67% across the three categories. In addition, ChatGPT-4o achieved a high success rate in the following cases: digoxin poisoning, with an average of 95.83% success rate, CO poisoning with an average of 94.45% success rate. The only case in which ChatGPT-4o performed poorly was a snake bite, with a success rate of 67.5%. The average success rates of ChatGPT-4o in the simulated case scenarios are presented in Table 2.

The average success rate of ChatGPT-4o in the simulated case scenarios was evaluated separately for each of the three categories, and it was observed that ChatGPT-4o was successful in all categories. ChatGPT-4o achieved a success rate of 91.25% in the "general evaluation and stabilization" category, 86.15% in the "diagnostic activities" category, and 95.25% in the "medical treatment and follow-up" category. The average success rates of ChatGPT-4o at the case level, based on categories, were calculated, and the results are shown in Table 2. The results indicate that ChatGPT-4o achieved a perfect (100%) success rate in the "general evaluation and stabilization" category in all cases, except for snake bite. In the case of snake bites, ChatGPT-4o was unsuccessful in the "general evaluation and stabilization" category, with a success rate of 12.5%.

ChatGPT-4o was successful in 6 cases (60%) in the diagnostic activities category. We achieved a perfect (100%) success rate in cases of TCA poisoning, organophosphate toxicity, digoxin poisoning, and snake bite. In the category of "diagnostic activities", although ChatGPT-4o received a passing grade in the remaining cases, it demonstrated limited success, particularly in lithium poisoning (66.67%) and CCB poisoning (67.86%). The category in which ChatGPT-4o performed the least successfully was diagnostic activities, with an average success rate of 86.15%.

ChatGPT-4o was successful in 9 cases (90%) in the "medical treatments and follow-up" category. It achieved a perfect (100%) success rate in all cases (70%) except organophosphate toxicity, digoxin poisoning, and snake bite. ChatGPT-4o received a "passing grade" in the "medical treatments and follow-up" category for organophosphate poisoning with a 75% success rate. The statistical data on the success rates achieved by ChatGPT-4o in the simulated case scenarios are presented in Table 2.

During the initial phase of the study, questions that ChatGPT-4o failed to answer accurately in the simulated case scenarios were subsequently presented as knowledge-based questions in a separate second test. During the secondary test, the correct responses given by ChatGPT-4o for each case were scored in accordance with the evaluation criteria applied in the assessment of the simulated case scenarios. The overall success rate was then determined based on the results of the secondary tests. The success rates obtained from the secondary test and the simulated case scenarios are comparatively presented in Table 3.

It was observed that ChatGPT-4o's average success rate in all three categories and across all cases increased from 90.88% to 97.22% with the secondary test. In the secondary test,

Table 2. Analyzing the performance of ChatGPT-4o in simulated case scenarios across different categories

Simulated cases	General evaluation and stabilization		Diagnostics acitivies		Medical treatments and follow-up		Total	
Case number	**Points**	**Results**	**Points**	**Results**	**Points**	**Results**	**%**	**Result**
1. Acetaminophen overdose	100	Success	78.6	Improvable	100	Success	92.86	Success
2. TCA overdose	100	Success	100	Success	100	Success	100	Success
3. Methanol toxicity	100	Success	90	Success	100	Success	96.67	Success
4. Organophosphate toxicity	100	Success	100	Success	75	Improvable	91.67	Success
5. Digoxin poisoning	100	Success	100	Success	87.5	Success	95.83	Success
6. Sympathomimetic toxidrome	100	Success	75	Improvable	100	Success	91.67	Success
7. Lithium overdose	100	Success	66.67	Improvable	100	Success	88.89	Success
8. CO poisoning	100	Success	83.34	Success	100	Success	94.45	Success
9. CCB overdose	100	Success	67.86	Improvable	100	Success	89.29	Success
10. Snake bite	12.5	Failure	100	Success	90	Success	67.5	Improvable
Total (%)	**91.25**	**Success**	**86.15**	**Success**	**95.25**	**Success**	**90.88**	**Success**

CCB: Calcium channel blocker, CO: Carbon monoxide, TCA: Tricyclic antidepressant

Table 3. Comparison of success rates of ChatGPT-4o after simulated case scenarios and overall success rates following the secondary tests

	Success in simulated case scenarios				Overall success following the secondary test			
Simulated case scenarios	**Category I**	**Category II**	**Category III**	**Total (%)**	**Category I**	**Category II**	**Category III**	**Total (%)**
1. Acetaminophen overdose	100	78.6	100	92.86	100	85.74	100	95.25
2. TCA overdose	100	100	100	100	100	100	100	100
3. Methanol toxicity	100	90	100	96.67	100	100	100	100
4. Organophosphate toxicity	100	100	75	91.67	100	100	100	100
5. Digoxin poisoning	100	100	87.5	95.83	100	100	100	100
6. Sympathomimetic toxidrome	100	75	100	91.67	100	100	100	100
7. Lithium overdose	100	66.67	100	88.89	100	66.67	100	88.89
8. CO poisoning	100	83.34	100	94.45	100	100	100	100
9. CCB overdose	100	67.86	100	89.29	100	92.86	100	97.62
10. Snake bite	12.5	100	90	67.5	81.25	100	90	90.42
Total (%)	**91.25**	**86.15**	**95.25**	**90.88**	**98.13**	**94.53**	**99**	**97.22**

Category I: General evaluation and stabilization, Category II: Diagnostic activities, Category III: Medical treatments and follow-up, CCB: Calcium channel blocker, CO: Carbon monoxide, TCA: Tricyclic antidepressant

ChatGPT-4o achieved an average success rate of over 80% in all cases. As a result of the second test, ChatGPT-4o achieved 100% success in TCA overdose, methanol toxicity, organophosphate toxicity, digoxin poisoning, sympathomimetic toxidrome, and CO poisoning by answering all questions correctly in all three categories. Specifically, it improved the success rate from 67.5% to 90.42% in the case of snake bite. As a result, ChatGPT-4o's success grade in snake bite cases increased from "improvable" to "success". After the second test, ChatGPT-4o failed to increase success rates in cases of lithium overdose. Secondary tests did not contribute to the success rate in this case.

The success rates of ChatGPT-4o in the simulated case scenarios and after the secondary tests were compared, and the results are shown in Table 3. As indicated in the table, ChatGPT-4o demonstrated success in all three categories after the secondary tests, with a score of over 80%. By category, the success rates were 99% in category III (medical treatments and follow-up), 98.13% in category I (general assessment and stabilization), and 94.53% in category II (diagnostic activities). As a result of the secondary tests, the success rate of snake bite cases in category 1 increased from 12.5% to 81.25%. This resulted in a change of the success grade from "failure" to "success" in the case of snake bite. In category II, the success rates of acetaminophen overdose and CCB overdose increased to 85.74% and 92.86%, respectively. In these cases, ChatGPT-4o's success grade increased from "improvable" to "success" in category II. Similarly, in category III, the success rate of organophosphate toxicity increased from 75% to 100%.

Discussion

This study is the pioneering evaluation of ChatGPT-4o proficiency in medical toxicology through the use of simulated case scenarios. The results of our study indicate that ChatGPT-4o achieved high success rates in toxicology case scenarios. However, in a very small number of instances, it did not perform as well in specific categories. Given the potential for improvement in this aspect, it has emerged that ChatGPT has the potential to be used in areas where access to poison information centers and medical toxicologists is limited in the future.

The integration of AI with ChatGPT's advanced assessment and response capabilities demonstrates the contribution of AI in overcoming human workforce and time limitations. This contribution has prompted current research to focus on AI applications similar to ChatGPT. Recent studies utilizing ChatGPT encompass various areas, such as AI-assisted article writing, medical problem-solving, case analysis, exam/test solutions, triage, and generating differential diagnosis lists [9-12]. ChatGPT has no official approval for use in medicine and health [13]. In studies employing ChatGPT, there is frequently an emphasis on ChatGPT's CDS capabilities [10-12]. In these studies, ChatGPT draws conclusions about hypothetical variables, fictional cases, or scenarios and subsequently assesses the accuracy of the conclusions.

In our study, ChatGPT-4o demonstrated high success in providing medical recommendations for the diagnosis, treatment, and follow-up of overall poisoning cases. However,

it was also identified that there were scenarios in which it could be further improved and cases in which it was unsuccessful. In the "diagnostic activities" category, ChatGPT-4o gave insufficient responses concerning the diagnostic parameters that should be routinely requested and utilized for differential diagnosis in every overdose patient arriving at the emergency department. These parameters include serum acetaminophen, urine drug screen, serum salicylate level, serum ethanol level, and serum beta-HCG level in reproductive-age women during the prodromal period. Therefore, ChatGPT-4o did not achieve a perfect (100%) in the diagnostic activities stage of poisoning cases. The lack of appropriate diagnostic tests that should be routinely requested in toxicology had a significant impact on the overall success rate. However, even in such a situation, ChatGPT-4o provided highly successful responses in specific aspects of the diagnostic activities tailored to the individual case.

ChatGPT-4o failed in the "general evaluation and stabilization" category for snake bites. However, it received a satisfactory rating in the secondary tests. ChatGPT-4o demonstrated variable success rates in the categories of different cases. Some questions that were incorrectly answered in the initial assessment were correctly answered in the secondary tests (direct queries). In the secondary tests, the success rate increased from 90.88% to 97.22%. This demonstrates that ChatGPT-4o is capable of delivering highly accurate responses when directly queried with factual inquiries. The observation that previously unsuccessful phases yielded accurate responses when posed with direct factual inquiries emphasizes the importance of further enhancing ChatGPT's already well-regarded analytical prowess and interpretive capabilities.

The high analytical power of ChatGPT-4o stems from its ability to process the concrete data provided by users. Another noteworthy attribute of ChatGPT-4o is its capacity to generate intuitive or inferential judgments, in addition to processing concrete data, which is a capability that humans, as the most intelligent beings on earth, are also capable of. This feature plays a pivotal role in enhancing performance in case scenarios. The most illustrative examples of poisoning in our study are methanol poisoning, CO poisoning, and snake bites. In the case of methanol toxicity, it is explicitly stated that the individual consumed homemade alcohol with friends. Based on the patient's history, clinical, and laboratory findings, ChatGPT-4o hypothesized that the individual consumed methanol. In the simulated case scenario of CO poisoning while smoking hookah with friends at a café, ChatGPT-4o was able to diagnose CO poisoning. Similarly, in the simulated case scenario in which the person expressed feeling pain in her foot and the snake moved away afterwards, ChatGPT was able to infer that a snake bite had occurred.

One of the main drawbacks limiting the use of AI applications is that they give wrong answers to questions. In our study, ChatGPT-4o did not give incorrect answers to any questions. This supports that ChatGPT-4o can be used as a reference source. In our study, the reason for the low ChatGPT-4o score was its inability to provide the desired answer. For example, serum paracetamol level and serum ethanol level were not requested in patients with intentional overdose, and B-HCG was not requested in young women of childbearing age. These factors reduced the success of ChatGPT-4o. Another shortcoming is that ChatGPT-4o reacts to poisoning by ignoring the possibility that the patient may have taken more than one drug at the same time. This can be overcome by the clinical experience of the clinician. However, these shortcomings can also be overcome by training ChatGPT on topics such as "treatment approaches in poisoning patients" and "diagnostic approaches in poisoning patients."

GPTs (generative pre-trained transformers) represent one of the most significant indicators that ChatGPT can be used as a reliable reference in the future. On November 6, 2023, OpenAI announced the launch of custom versions of ChatGPT designed for specific purposes, which are referred to as GTPs. Thus, with a single click (chatgpt.com/create), users can train AI on a specific topic without coding knowledge. As in other professions, GPTs have been established in medicine and health. GPTs can also be trained in specific sub-branches of medicine, such as anatomy, biochemistry, and cardiology. The subject matter of the GPTs is limited only by the user's imagination. This indicates that if GPTs are developed in the field of medical toxicology, they could serve as a gateway for assessing patients with poisoning, offering the same level of expertise as a poison information center but with greater responsiveness.

Despite not being specifically trained in medical toxicology or medicine, ChatGPT's high analytical power and accurate response rates demonstrate its potential in medical decision-making. Although the reliability of AI in medical decision-making remains a topic of debate, the results of our study suggest that ChatGPT can be used by healthcare professionals in areas where access to poison information centers and medical toxicologists is limited. Furthermore, we believe that this study will shed light on the possibility of using ChatGPT as a portal for instant access to accurate and up-to-date information based on textbooks and guidelines.

Study of Limitations

The main limitation of this study is that ChatGPT-4o is not clinically approved for obtaining health information. Although ChatGPT-4o is competent in medical topics, it lacks specific training in medical toxicology. Another limitation of this study is that it was conducted using English questions. It is possible

that the results may be positive or negative depending on whether the clinician conducts the query in their native language.

Conclusion

The ability to query current medical toxicology information through ChatGPT showcases the potential of ChatGPT-4o to serve as a next-generation poison information center, providing a function that physicians can easily access, especially in areas where access to medical toxicologists is limited. The development of algorithms and innovations has made it possible to significantly enhance the success rate of ChatGPT in offering medical toxicology recommendations.

Ethics

Ethics Committee Approval: Not required.

Informed Consent: Not required.

Footnotes

Authorship Contributions

Concept: İ.A., A.Ş., Design: İ.A., S.K., R.G., A.Ş., Data Collection or Processing: İ.A., R.G., A.Ş., Analysis or Interpretation: İ.A., S.K., R.G., Literature Search: İ.A., S.K., Writing: İ.A., S.K., R.G., A.Ş.

Conflict of Interest: Ramazan Güven, MD, is a Section Editor in the Emergency and Critical Care. He had no involvement in the peer-review of this article and had no access to information regarding its peer-review. Other authors have nothing to disclose.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Friedman LS, Krajewski A, Vannoy E, Allegretti A, Wahl M. The association between U.S. Poison Center assistance and length of stay and hospital charges. Clin Toxicol (Phila). 2014;52:198-206.
2. Hao T, Huang Z, Liang L, Weng H, Tang B. Health natural language processing: methodology development and applications. JMIR Med Inform. 2021;9:23898.
3. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med. 2020;3:17.
4. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Family Med Prim Care. 2019;8:2328-31.
5. Nath S, Marie A, Ellershaw S, Korot E, Keane PA. New meaning for NLP: the trials and tribulations of natural language processing with GPT-3 in ophthalmology. Br J Ophthalmol. 2022;106:889-92.
6. Larsen TM, Endo BH, Yee AT, Do T, Lo SM. internal assumptions of the revised bloom's taxonomy. CBE Life Sci Educ. 2022;21:ar66.
7. Nelson LS, Howland MA, Lewin NA, Smith SW, Goldfrank LR, Hoffman RS. Goldfrank's toxicologic emergencies, 11th ed. McGraw Hill, 2019.
8. The Extracorporeal Treatments in Poisoning Workgroup (EXTRIP), The information available at the online source https://www.extrip-workgroup.org/
9. Macdonald C, Adeloye D, Sheikh A, Rudan I. Can ChatGPT draft a research article? An example of population-level vaccine effectiveness analysis. J Glob Health. 2023;13:01003
10. Wang YM, Shen HW, Chen TJ. Performance of ChatGPT on the pharmacist licensing examination in Taiwan. J Chin Med Assoc. 2023;86:653-8.
11. Bhattaram S, Shinde VS, Khumujam PP. ChatGPT: The next-gen tool for triaging? Am J Emerg Med. 2023;69:215-7.
12. Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic accuracy of differential-diagnosis lists generated by generative pretrained transformer 3 chatbot for clinical vignettes with common chief complaints: a pilot study. Int J Environ Res Public Health. 2023;20:3378.
13. Baumgartner C, Baumgartner D. A regulatory challenge for natural language processing (NLP)-based tools such as ChatGPT to be legally used for healthcare decisions. Where are we now?. Clin Transl Med. 2023;13:1362.

SUPPLEMENTAL MATERIAL LINKS-1: https://d2v96fxpocvxx.cloudfront.net/2a4f1576-691d-4c9c-9173-1686c7aa9aea/documents/6-SUPPLEMENTAL-1-06025.pdf

SUPPLEMENTAL MATERIAL LINKS-2: https://d2v96fxpocvxx.cloudfront.net/2a4f1576-691d-4c9c-9173-1686c7aa9aea/documents/6-SUPPLEMENTAL-2-06025.pdf

SUPPLEMENTAL MATERIAL LINKS-3: https://d2v96fxpocvxx.cloudfront.net/2a4f1576-691d-4c9c-9173-1686c7aa9aea/documents/6-SUPPLEMENTAL-3-06025.pdf