Received: 16 September 2025 / Accepted: 11 February 2026 / Published online: 27 February 2026

DOI 10.34689/SH.2026.28.1.012

UDC 616.099:004-035.258

EVALUATION OF READABILITY INDICES OF CHATGPT-4 AND GOOGLE GEMINI ABOUT MUSHROOM POISONING

Ökkeş Zortuk[1],

Cihan Bedel[2], https://orcid.org/0000-0002-3823-2929

[1] **Department of Emergency Medicine, Bandırma University Hospital, Balıkesir, Turkey;**
[2] **Department of Emergency Medicine, Health Science University Antalya Training and Research Hospital, Antalya,Turkey.**

Abstract

Aim: Mushroom poisoning is a significant public health concern. artificial intelligence (AI) technologies have shown remarkable success in mushroom classification and poisoning prevention, challenges remain. It is clear that the complexity and computational demands of deep learning models can act as a barrier for interdisciplinary researchers [10]. In this study, we wanted to compare the readability indices of AI programs in mushroom poisoning cases in the literature.

Materials and Methods: This study included 100 questions about mushroom poisoning Questions were compiled from publicly available question-and-answer platforms such as the Quaro program and ChatGPT-5 and Gemini-2.5 Pro AI models compared. Before generating the questions, we instructed both AI models to formulate their responses in language appropriate for a general audience with no prior medical knowledge on the topic. The responses generated by the models were recorded.

Results: Gemini's content showed a significantly higher Automated Readability Index (ARI) of 9.89 ± 0.9 compared to Open AI's 8.93 ± 0.78, with a p-value of <0.001. Gemini's content had a lower Fog Scale score of 9.9 ± 1.01 compared to Open AI's 11.3 ± 0.98, with a p-value of <0.001. Gemini's Flesch-Kincaid Grade Level was 8.59 ± 0.56, which was higher than Open AI's 7.97 ± 0.12, with a p-value of <0.001. Gemini's Coleman-Liau Index was 13.29 ± 1.27, significantly higher than Open AI's 11.38 ± 1.12, with a p-value of <0.001. Gemini had a higher SMOG index of 8.38 ± 0.78 compared to Open AI's 7.58 ± 1.34, with a p-value of 0.013. Open AI's Forcast readability formula score was 13.02 ± 0.12, compared to Gemini's 11.82 ± 2.04, p=0.006.

Conclusion. While some readability metrics like Average Reading Level, Flesch Reading Ease, and Linsear Write Formula showed no significant difference between ChatGPT and Gemini, other indices such as Automated Readability Index, Fog Scale, Flesch-Kincaid Grade Level, Coleman-Liau Index, SMOG index, and Forcast readability formula revealed statistically significant differences, often indicating that Gemini's output tends to be more complex or, in some cases, easier depending on the specific metric.

Keywords: *Mushroom poisoning, readability, Automated Readability Index, Fog Scale, Flesch-Kincaid Grade Level, Coleman-Liau Index, SMOG index*

For citation:
Zortuk Ö., Bedel C. Evaluation of readability indices of ChatGPT-4 and Google Gemini about mushroom poisoning // *Nauka i Zdravookhranenie* [Science & Healthcare]. 2026. Vol.28 (1), pp. 99-105. *doi 10.34689/SH.2026.28.1.012*

Резюме

ОЦЕНКА ПОКАЗАТЕЛЕЙ ЧИТАЕМОСТИ CHATGPT-4 И GOOGLE GEMINI В ВОПРОСАХ ОТРАВЛЕНИЯ ГРИБАМИ

Оккес Зортук[1],

Джихан Бедель[2], https://orcid.org/0000-0002-3823-2929

[1] **Отделение неотложной медицины, Университетская больница Бандырма, г. Балыкесир, Турция;**
[2] **Отделение неотложной медицины, Учебно-исследовательская больница при Университете медицинских наук Антальи, г. Анталья, Турция.**

Цель: Отравление грибами представляет собой значимую проблему общественного здравоохранения. Технологии искусственного интеллекта (ИИ) продемонстрировали значительные успехи в классификации грибов и профилактике отравлений, однако сохраняются определённые трудности. Очевидно, что сложность и вычислительные требования моделей глубокого обучения могут являться барьером для междисциплинарных исследователей [10]. В данном исследовании мы хотели сравнить индексы читаемости программ ИИ в случаях отравления грибами, описанных в литературе.

Материалы и методы: В исследование были включены 100 вопросов об отравлении грибами. Вопросы были собраны из общедоступных платформ формата «вопрос–ответ», таких как Quora. Были сравнены модели ИИ ChatGPT-5 и Gemini-2.5 Pro. Перед генерацией ответов обеим моделям было дано указание формулировать ответы языком, понятным широкой аудитории без медицинской подготовки. Ответы, сгенерированные моделями, были записаны.

Результаты: Контент Gemini продемонстрировал статистически значимо более высокий показатель индекс автоматической читаемости (ARI) — 9,89 ± 0,9 по сравнению с OpenAI — 8,93 ± 0,78 ($p < 0,001$). При этом показатель Fog Scale у Gemini был ниже (9,9 ± 1,01) по сравнению с OpenAI (11,3 ± 0,98), $p < 0,001$. Уровень Flesch-Kincaid Grade Level у Gemini составил 8,59 ± 0,56, что выше, чем у OpenAI — 7,97 ± 0,12 ($p < 0,001$). Индекс Coleman-Liau у Gemini был значительно выше (13,29 ± 1,27) по сравнению с OpenAI (11,38 ± 1,12), $p < 0,001$. Индекс SMOG также оказался выше у Gemini (8,38 ± 0,78) по сравнению с OpenAI (7,58 ± 1,34), $p = 0,013$. Показатель по формуле Forcast у OpenAI составил 13,02 ± 0,12, тогда как у Gemini — 11,82 ± 2,04 ($p = 0,006$).

Заключение: Хотя по некоторым показателям читабельности, таким как средний уровень чтения, Flesch Reading Ease и формула Linsear Write, статистически значимых различий между ChatGPT и Gemini не выявлено, другие индексы Automated Readability Index, Fog Scale, Flesch-Kincaid Grade Level, Coleman-Liau Index, SMOG index и Forcast readability formula продемонстрировали статистически значимые различия. В большинстве случаев результаты указывают на то, что тексты Gemini могут быть более сложными либо, в отдельных случаях, более простыми в зависимости от конкретного показателя.

Ключевые слова: *отравление грибами, читабельность, Automated Readability Index, Fog Scale, Flesch-Kincaid Grade Level, Coleman-Liau Index, SMOG index.*

Для цитирования:

Зортук О., Бедель Д. Оценка показателей читаемости ChatGPT-4 и Google Gemini в вопросах отравления грибами // *Nauka i Zdravookhranenie* [Science & Healthcare]. 2026. Vol.28 (1), pp. 99-105. *doi 10.34689/SH.2026.28.1.012*

Түйіндеме

CHATGPT-4 ЖӘНЕ GOOGLE GEMINI МОДЕЛЬДЕРІНІҢ САҢЫРАУҚҰЛАҚТАН УЛАНУ ТУРАЛЫ МӘТІНДЕРІНІҢ ОҚЫЛЫМДЫЛЫҚ КӨРСЕТКІШТЕРІН БАҒАЛАУ

Оккес Зортук[1],

Джихан Бедель[2], **https://orcid.org/0000-0002-3823-2929**

[1] **Жедел медицина бөлімі, Бандырма университеттік ауруханасы, Балыкесир қ., Түркия;**
[2] **Жедел медицина бөлімі, Денсаулық сақтау ғылымдары университетінің Анталья оқу-зерттеу ауруханасы, Анталья қ., Түркия.**

Мақсаты: Саңырауқұлақтан улану қоғамдық денсаулық сақтау саласындағы маңызды мәселе болып табылады. Жасанды интеллект (ЖИ) технологиялары саңырауқұлақтарды жіктеу және уланудың алдын алу бағытында айтарлықтай жетістіктер көрсетті, алайда белгілі бір қиындықтар сақталуда. Терең оқыту модельдерінің күрделілігі мен жоғары есептеу талаптары пәнаралық зерттеушілер үшін кедергі болуы мүмкін екені анық [10]. Осы зерттеуде біз әдебиетте сипатталған саңырауқұлақтан улану жағдайлары бойынша ЖИ бағдарламаларының оқылымдылық индекстерін салыстыруды мақсат еттік.

Материалдар мен әдістер: Зерттеуге саңырауқұлақтан улану туралы 100 сұрақ енгізілді. Сұрақтар Quora сияқты ашық «сұрақ–жауап» платформаларынан жиналды. ChatGPT-5 және Gemini-2.5 Pro жасанды интеллект модельдері салыстырылды. Жауаптарды генерациялау алдында екі модельге де медициналық білімі жоқ кең аудиторияға түсінікті тілде жауап беру туралы нұсқау берілді. Модельдер ұсынған жауаптар тіркелді.

Нәтижелер: Gemini ұсынған контенттің автоматтандырылған оқылымдылық индексі (ARI) 9,89 ± 0,9 болып, OpenAI көрсеткішінен (8,93 ± 0,78) статистикалық тұрғыдан едәуір жоғары болды ($p < 0,001$). Сонымен қатар, Fog Scale көрсеткіші Gemini-де төмен (9,9 ± 1,01), ал OpenAI-де жоғары (11,3 ± 0,98) болды ($p < 0,001$). Flesch-Kincaid Grade Level көрсеткіші Gemini-де 8,59 ± 0,56 болып, OpenAI-ден (7,97 ± 0,12) жоғары анықталды ($p < 0,001$). Coleman-Liau индексі Gemini-де 13,29 ± 1,27 болып, OpenAI көрсеткішінен (11,38 ± 1,12) айтарлықтай жоғары болды ($p < 0,001$). SMOG индексі де Gemini-де жоғары (8,38 ± 0,78), ал OpenAI-де 7,58 ± 1,34 болды ($p = 0,013$). Forcast формуласы бойынша OpenAI көрсеткіші 13,02 ± 0,12, ал Gemini-де 11,82 ± 2,04 болды ($p = 0,006$).

Қорытынды: Оқылымдылықтың кейбір көрсеткіштері, атап айтқанда орташа оқу деңгейі, Flesch Reading Ease және Linsear Write формуласы бойынша ChatGPT пен Gemini арасында статистикалық тұрғыдан маңызды айырмашылық анықталмады. Алайда Automated Readability Index, Fog Scale, Flesch-Kincaid Grade Level, Coleman-Liau Index, SMOG индексі және Forcast формуласы бойынша статистикалық тұрғыдан маңызды айырмашылықтар байқалды. Көп жағдайда нәтижелер Gemini мәтіндерінің күрделірек болуы мүмкін екенін, ал кейбір көрсеткіштер бойынша керісінше қарапайым болуы ықтимал екенін көрсетті.

Түйінді сөздер: *саңырауқұлақтан улану, оқылымдылық, Automated Readability Index, Fog Scale, Flesch-Kincaid Grade Level, Coleman-Liau Index, SMOG индексі.*

Дәйексөз үшін:

Зортук О., Бедель Д. ChatGPT-4 және Google Gemini модельдерінің саңырауқұлақтан улану туралы мәтіндерінің оқылымдылық көрсеткіштерін бағалау // Ғылым және Денсаулық сақтау. 2026. Vol.28 (1), Б. 99-105. *doi 10.34689/SH.2026.28.1.012*

Introduction

Mushroom poisoning is a significant public health concern. Toxic mushrooms can cause a range of clinical symptoms, from mild gastrointestinal distress to severe organ failure and death. The difficulty in distinguishing between edible and poisonous mushrooms is a key risk factor, as many toxic species closely resemble their edible counterparts [1]. The most common initial symptoms of mushroom poisoning are nausea, vomiting, diarrhea and abdominal pain. It is clear that these symptoms are reported in a high percentage of cases, with nausea and vomiting occurring in 81.5% to 80% of patients in various studies [2]. It is a fact that poisonous mushrooms contain various toxins, including cyclopeptides, gyromitrin, muscarine and orellanine. Cyclopeptides, particularly amatoxins found in Amanita species, are the most lethal, causing over 90% of mushroom poisoning deaths [3].The integration of artificial intelligence (AI) in the identification and classification of mushrooms has shown significant promise in reducing mushroom poisoning incidents. AI technologies, especially machine learning and deep learning, have been used to develop systems that can accurately identify edible and poisonous mushrooms [4].

These systems use various algorithms and models to analyse mushroom characteristics and provide reliable identification, thereby reducing the risk of mushroom poisoning. The following sections explore different AI approaches and their effectiveness in mushroom classification [5]. The deep learning-based mobile application MushAPP identifies mushroom species and their toxicity by analysing images captured via a smartphone camera. The application's accuracy rate of 99.8% in detecting mushroom species is undeniable proof of the potential of deep learning in real-time applications to reduce mushroom poisoning incidents [6].

CNNs are the most effective way to classify mushrooms as edible or poisonous. For instance, a study used CNNs to analyse images of mushrooms, achieving high credibility in classification and providing a decision-making basis for selecting edible fungi [7]. This machine learning algorithm was used to classify mushrooms based on attributes from 23 species. The study definitively proved the RandomForestClassifier to be a practical and accurate alternative to deep learning methods for mushroom classification [8]. We tested various machine learning algorithms, including Naive Bayes, Decision Tree, Support Vector Machine, and AdaBoost, for mushroom classification. The AdaBoost model achieved a perfect classification success rate of 100%, demonstrating its unparalleled effectiveness in distinguishing between edible and poisonous mushrooms [9]. While AI technologies have shown remarkable success in mushroom classification and poisoning prevention, challenges remain. It is clear that the complexity and computational demands of deep learning models can act as a barrier for interdisciplinary researchers [10]. In this study, we wanted to compare the readability indices of AI programs in mushroom poisoning cases in the literature.

Materials and metod

This study included 100 questions about mushroom poisoning posed online by patients and their families to healthcare professionals and artificial intelligence programs. Similar questions, along with incomplete or incorrect answers, were excluded. Questions were compiled from publicly available question-and-answer platforms such as the Quaro program. Before analysis, the questions were reviewed, corrected for grammar and meaning, and then sent to ChatGPT-5 and Gemini-2.5 Pro AI models to generate responses. Before generating the questions, we instructed both AI models to formulate their responses in language appropriate for a general audience with no prior medical knowledge on the topic. The responses generated by the models were recorded.

Readability Analysis and Statistical Evaluation

The **Flesch Reading Ease (FRE)** formula is a readability test designed to indicate how easy or difficult a passage in English is to understand. It's widely used in education, publishing, and content writing [11].

90–100	Very Easy	5th grade
80–89	Easy	6th grade
70–79	Fairly Easy	7th grade
60–69	Standard	8th–9th grade
50–59	Fairly Difficult	10th–12th grade
30–49	Difficult	College
0–29	Very Confusing	College graduate level

Use the Flesch-Kincaid Grade Level (FKGL) Formula.

The FKGL formula is used to calculate the Flesch-Kincaid Grade Level, which estimates the U.S. school grade level required to understand a text. The formula is as follows: FKGL = 0.39 (Total Sentences × Total Words) + 11.8 (Total Words × Total Syllables) – 15.59. For example:

FKGL = 8.2. This is something that can be understood by an average 8th grader.

FKGL = 12.0. This means it is suitable for 12th grade (senior in high school) [12].

Fog Scale (Gunning FOG Formula): The Gunning Fog Index is a reliable readability formula. It estimates the years of formal education required to understand a text on the first reading. It is used in English and has been adapted for other languages, such as Indonesian. The formula uses the average sentence length and the percentage of complex words, defined as those with three or more syllables, to calculate a readability score [13].

FOG = 0.4 ((Total Sentences Total Words)+100 (Total Words Complex Words))

The **SMOG Index** is a readability formula that estimates the years of education a person needs to understand a piece of writing. It is perfect for health communication and legal or public policy documents [14].

The **Automated Readability Index (ARI)** is a readability test that estimates the **U.S. grade level** required to understand a text. Unlike other readability formulas, ARI relies on **character count**, not syllables, making it easier to compute automatically.

ARI=4.71(Words/Characters)+0.5(Sentences/Words)−21.43 [15].

Coleman-Liau Index: The Coleman-Liau Index (CLI) is the go-to metric for assessing the complexity of a text by estimating the U.S. school grade level required to understand it. The CLI is different from other readability formulas. It doesn't rely on syllable counts. The CLI uses the number of characters, words and sentences to calculate its score. This makes it particularly suitable for computerized text analysis. This index is widely applied across various domains to evaluate the accessibility of written materials, from scientific articles to political manifestos and policy guides [16].

Linsear Write Formula: The Linsear Write Formula is the go-to metric for gauging the complexity of English text, ensuring it meets the needs of readers at all educational levels. It is clear that this is not directly related to the mathematical contexts provided in the papers, which focus on linear algebra and its applications [17].

Dale-Chall Readability Score: The Dale-Chall Readability Score is the go-to formula for assessing the readability of a text. It does this by evaluating its complexity and the familiarity of the words used. It is vital to use this in educational settings to determine if reading materials are appropriate for a specific grade level. The formula considers both the number of words in a sentence and the percentage of difficult words, which are defined as words not found on a list of 3,000 familiar words. The score is then converted into a grade level, indicating the educational level required to understand the text[18].

Spache Readability Formula: The Spache Readability Formula is the tool you need to assess the readability of texts, especially for primary-level readers. It is a classic readability formula. It evaluates text difficulty by analysing sentence length and the frequency of "difficult" words. "Difficult" words are words not included in a list of common words. This formula is specially tailored for texts aimed at children in the lower elementary grades, typically ranging from 1.5 to fourth grade [19]..

The scores obtained from all readability formulas were compiled. We subjected these scores to a comparative and analytical statistical process using SPSS software.

Statistical Analysis

In our study, we analysed data from artificial intelligence models using SPSS version 27 (IBM Corp., USA). The data were categorised according to type. Categorical data were defined as percentages and frequencies. The chi-squared test was used to compare these. Distribution analysis was used to describe the numerical data. Data conforming to a normal distribution were expressed as the mean ± standard deviation and a t-test was applied. Data with a p-value below 0.05 were considered significant.

Result

The average reading level for Open AI was 10.02 ± 2.24, while for Gemini it was 10.42 ± 1.58, p=0.259. Gemini's content showed a significantly higher Automated Readability Index (ARI) of 9.89 ± 0.9 compared to Open AI's 8.93 ± 0.78, with a p-value of <0.001. Gemini's content had a lower Fog Scale score of 9.9 ± 1.01 compared to Open AI's 11.3 ± 0.98, with a p-value of <0.001. A lower Fog Scale score indicates easier readability, suggesting Gemini's output is easier to read by this metric. Gemini's Flesch-Kincaid Grade Level was 8.59 ± 0.56, which was higher than Open AI's 7.97 ± 0.12, with a p-value of <0.001. Gemini's Coleman-Liau Index was 13.29 ± 1.27, significantly higher than Open AI's 11.38 ± 1.12, with a p-value of <0.001. Gemini had a higher SMOG index of 8.38 ± 0.78 compared to Open AI's 7.58 ± 1.34, with a p-value of 0.013. Open AI's Forcast readability formula score was 13.02 ± 0.12, compared to Gemini's 11.82 ± 2.04, p=0.006.

Discussion

Mushroom poisoning is a significant concern in emergency medicine due to the potential for severe health outcomes, ranging from mild gastrointestinal disturbances to life-threatening organ failure. The complexity of mushroom poisoning arises from the diversity of toxic species, the variability in symptoms, and the challenges in accurate identification. Emergency departments play a crucial role in managing these cases, often requiring collaboration with poison centers and mycologists to ensure accurate diagnosis and treatment [20].The following sections explore the key aspects of mushroom poisoning and its implications for emergency medicine. Mushroom poisoning often presents with gastrointestinal symptoms such as nausea, vomiting, and diarrhea, which are reported in a majority of cases [21].Severe cases can lead to organ failure, particularly liver damage, as seen with Amanita phalloides intoxication [22].Symptoms can vary significantly depending on the mushroom species ingested, with some causing neurological effects like hallucinations and decreased consciousness, as observed in Pantherina syndrome (*Yıldırım et al., 2016).* While mushroom poisoning remains a critical issue in emergency medicine, the development of rapid diagnostic tools and effective treatment protocols can significantly improve patient outcomes [23].

The comparison between ChatGPT and Gemini reveals distinct strengths and weaknesses across various domains, including translation, business management, healthcare, and cybersecurity. Both models are advanced AI systems with unique capabilities, but they differ significantly in their performance and application suitability [24]. This analysis explores their comparative advantages and limitations, providing insights into their optimal use cases. In the realm of classical Arabic poetry translation, ChatGPT outperforms Gemini in thematic clarity, creativity, and prosody, as assessed by literature experts. Gemini lags significantly in prosody, indicating a gap in handling complex poetic structures [25]. Gemini excels in real-time market analysis, strategic planning, and data-driven decision-making due to its integration with Google's vast data resources. It is particularly effective in processing extensive datasets for insights and optimization [26]. ChatGPT, on the other hand, is noted for its qualitative insights, customer feedback analysis, and creative content generation, making it valuable for customer interactions and marketing efforts. Its adaptability and conversational skills enhance customer experiences [27]. In the medical field, particularly in glaucoma surgical planning, ChatGPT shows a higher consistency with expert opinions compared to Gemini. ChatGPT's performance is notably better in challenging cases, where it aligns more closely with specialist recommendations [28]. In our study ARI, Fog Scale score, Flesch-Kincaid Grade Level Coleman-Liau Index, SMOG index is significantly different in groups.

Readability indices are the tools used to evaluate how easy or difficult a text is to read. They have significant implications for both human and machine comprehension. These indices are especially important in the field of artificial intelligence (AI), where they can be used to improve the performance of machine learning models in understanding and processing tex [29]. The integration of readability indices with AI is key to achieving more effective information retrieval, text classification and educational applications. This answer definitively explores the relationship between readability indices and AI, focusing squarely on their application, limitations, and potential improvements [30]. The ARI is a tool that automates the process of determining text readability by calculating average word and sentence lengths. This index is an essential tool in technical fields like the Air Force, where it simplifies the evaluation of training materials [31]. Different readability indices can yield varying results for the same text, leading to inconsistencies in text classification. This disparity demands the development of more unified and accurate models [32]. Traditional readability indices are often misleading because they overlook cognitive aspects such as short-term memory capacity. These aspects can significantly impact reading difficulty. The GU index successfully addresses this by incorporating cognitive psychology and linguistic theories [33]. While simpler texts are generally easier for machines to understand, the correlation between readability levels and machine comprehension performance is not as strong as expected. It is clear that current AI systems do not fully replicate human reading capabilities [16-20]. Creating more precise readability models by integrating various indices and considering additional linguistic and cognitive factors is the only way to improve text evaluation accuracy [15-19]. AI and Machine Learning Integration: Leveraging AI is the key to enhancing readability indices, improving predictions of text complexity and enhancing machine comprehension systems. This integration undoubtedly leads to more sophisticated models that account for a wider range of textual features [20-25]. New readability models must be validated and their applications across different languages and contexts explored. It is vital to examine the impact of readability on both human and machine comprehension [24,25].

In summary, while some readability metrics like Average Reading Level, Flesch Reading Ease, and Linsear Write Formula showed no significant difference between ChatGPT and Gemini, other indices such as Automated Readability Index, Fog Scale, Flesch-Kincaid Grade Level, Coleman-Liau Index, SMOG index, and Forcast readability formula revealed statistically significant differences, often indicating that Gemini's output tends to be more complex or, in some cases, easier depending on the specific metric.

Table 1.

Comparison of CHATGPT vs GEMINI in terms of readability scores.

	Open AI (n=20)	GEMINI (n=20)	p-Value
Average Reading Level Consensus	10,02±2,24	10,42±1,58	0,259
Automated Readability Index	8,93±0,78	9,89±0,9	<0,001
Flesch Reading Ease	52±5,65	52±4,76	0,500
Fog Scale	11,3±0,98	9,9±1,01	<0,001
Fesch-Kincaid Grade Level	7,97±0,12	8,59±0,56	<0,001
Coleman–Liau Index	11,38±1,12	13,29±1,27	<0,001
SMOG index	7,58±1,34	8,38±0,78	0,013
Linsear Write Formula	76±6,42	73±5,4	0,059
Forcast readability formula	13,02±0,12	11,82±2,04	0,006

Literatura:

1. *Tuğcan M.O., Akpınar A.A.* Mushroom poisoning: An updated review. Turk J Emerg Med. 2025 Jan 2;25(1):10-16. doi: 10.4103/tjem.tjem_129_24. PMID: 39882097; PMCID: PMC11774429.

2. *Tran H.H., Juergens A.L.* Mushroom Toxicity. 2023 Aug 7. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan–. PMID: 30725796.

3. *Li H., Zhang Y., Zhang H., Zhou J., Li Z., Yin Y., He Q., Jiang S., Zhang Y., Yuan Y., Lang N., Cheng B., Zhong J., Yuan M., Liu Z., Sun C.* Mushroom Poisoning Outbreaks - China, 2024. China CDC Wkly. 2025 May 9;7(19):645-649. doi: 10.46234/ccdcw2025.106. PMID: 40376178; PMCID: PMC12075492.

4. *Romanek K., Haberl B., Eyer F.* Verdacht auf Pilzvergiftung [Emergency checklist: Suspected mushroom poisoning]. MMW Fortschr Med. 2025 Aug;167(13):52-53. German. doi: 10.1007/s15006-025-5045-x. PMID: 40775162.

5. *Luo J., Chen L.* Status and development of spent mushroom substrate recycling: A review. J Air Waste Manag Assoc. 2024 Dec;74(12):843-860. doi: 10.1080/10962247.2024.2410447. Epub 2024 Oct 22. PMID: 39348219.

6. *Cottam R., Vounckx R.* Intelligence: Natural, artificial, or what? Biosystems. 2024 Dec;246:105343. doi: 10.1016/j.biosystems.2024.105343. Epub 2024 Sep 26. PMID: 39341545.

7. *Mukherjee J., Sharma R., Dutta P., Bhunia B.* Artificial intelligence in healthcare: a mastery. Biotechnol Genet Eng Rev. 2024 Nov;40(3):1659-1708. doi: 10.1080/02648725.2023.2196476. Epub 2023 Apr 4. PMID: 37013913.

8. *Judge C.S., Krewer F., O'Donnell M.J., Kiely L., Sexton D., Taylor G.W., Skorburg J.A, Tripp B.* Multimodal Artificial Intelligence in Medicine. Kidney360. 2024 Nov 1;5(11):1771-1779. doi: 10.34067/KID.0000000000000556.

Epub 2024 Aug 21. PMID: 39167446; PMCID: PMC12282626.

9. *Ng Yin Ling C., Zhu X., Ang M.* Artificial intelligence in myopia in children: current trends and future directions. Curr Opin Ophthalmol. 2024 Nov 1;35(6):463-471. doi: 10.1097/ICU.0000000000001086. Epub 2024 Aug 26. PMID: 39259652.

10. *Sasidharan Pillai S., Ashraf A.P.* Artificial Intelligence in Pediatric Endocrinology. Adv Pediatr. 2025 Aug;72(1):185-195. doi: 10.1016/j.yapd.2024.12.003. Epub 2025 Feb 4. PMID: 40582750.

11. *Incerti Parenti S., Bartolucci M.L., Biondi E., Maglioni A., Corazza G., Gracco A., Alessandri-Bonetti G.* Online Patient Education in Obstructive Sleep Apnea: ChatGPT versus Google Search. Healthcare (Basel). 2024 Sep 5;12(17):1781. doi: 10.3390/healthcare12171781. PMID: 39273804; PMCID: PMC11394980.

12. *Dihan Q., Chauhan M.Z., Eleiwa T.K., Brown A.D., Hassan A.K., Khodeiry M.M., Elsheikh R.H, Oke I., Nihalani B.R., VanderVeen D.K., Sallam A.B., Elhusseiny A.M.* Large language models: a new frontier in paediatric cataract patient education. Br J Ophthalmol. 2024 Sep 20;108(10):1470-1476. doi: 10.1136/bjo-2024-325252. PMID: 39174290.

13. *Senbaykal Yigit E., Taskirdi I., Haci I.A., Tuncel T.* Artificial intelligence performance in pediatric asthma. J Asthma. 2025 Aug 1:1-7. doi: 10.1080/02770903.2025.2531500. Epub ahead of print. PMID: 40643318.

14. *Yurdakurban E., Topsakal K.G., Duran G.S.* A comparative analysis of AI-based chatbots: Assessing data quality in orthognathic surgery related patient information. J Stomatol Oral Maxillofac Surg. 2024 Oct;125(5):101757. doi: 10.1016/j.jormas.2023.101757. Epub 2023 Dec 28. PMID: 38157937.

15. *O'Halloran C.P., Agarwal A., Hawcutt D.B., Oni L., Moss J.* Readability of paediatric participant information leaflets in research studies. Pediatr Res. 2025 Feb 22. doi: 10.1038/s41390-025-03943-z. Epub ahead of print. PMID: 39984643.

16. *Erkan A., Koc A., Barali D., Satir A., Zengin S., Kilic M., Dundar G., Guzelsoy M.* Can Patients With Urogenital Cancer Rely on Artificial Intelligence Chatbots for Treatment Decisions? Clin Genitourin Cancer. 2024 Dec;22(6):102206. doi: 10.1016/j.clgc.2024.102206. Epub 2024 Aug 14. PMID: 39236508.

17. *Webb H., Toppi J.* The readability of online laryngectomy patient information: how do Australia and New Zealand compare? N Z Med J. 2025 Mar 14;138(1611):93-101. doi: 10.26635/6965.6853. PMID: 40080764.

18. *Olszewski R., Watros K., Mańczak M., Owoc J., Jeziorski K., Brzeziński J.* Assessing the response quality and readability of chatbots in cardiovascular health, oncology, and psoriasis: A comparative study. Int J Med Inform. 2024 Oct;190:105562. doi: 10.1016/j.ijmedinf.2024.105562. Epub 2024 Jul 19. PMID: 39059084.

19. *Rashid S., Siddiqa A., Agama F.T., Idrees N., Alharthi M.S.* Poisson random measure noise-induced coherence in epidemiological priors informed deep neural networks to identify the intensity of virus dynamics. Sci Rep. 2025 May 17;15(1):17150. doi: 10.1038/s41598-025-94086-y. PMID: 40382439; PMCID: PMC12085636.

20. *Roy S., Saleem H.* Mushroom Poisoning and Acute Liver Injury: A Case-Based Review. Cureus. 2024 Dec 14;16(12):e75706. doi: 10.7759/cureus.75706. PMID: 39677988; PMCID: PMC11646078.

21. *Wu S., Wang Q., Duan H., Xiong J., Lai T., Wang C., Liang J., Jia H.* Mushroom poisoning deaths and prevention practices in Hunan, 2014-2023. PLoS One. 2025 Jun 16;20(6):e0326107. doi: 10.1371/journal.pone.0326107. PMID: 40522929; PMCID: PMC12169554.

22. *Yuan Y., Zhang Y., Zhou J., Lang N., Jiang S., Li H., He Q., Zhang Y., Cheng B., Zhong J., Liang J., Yin Y., Liang Y., Zhang H., Sun C.* Detection and Identification in Reported Mushroom Poisoning Incidents - China, 2012-2023. China CDC Wkly. 2024 Dec 20;6(51):1360-1364. doi: 10.46234/ccdcw2024.270. PMID: 39802087; PMCID: PMC11724132.

23. *Marcaccini G., Seth I., Xie Y., Susini P., Pozzi M., Cuomo R., Rozen W.M.* Breaking Bones, Breaking Barriers: ChatGPT, DeepSeek, and Gemini in Hand Fracture Management. J Clin Med. 2025 Mar 14;14(6):1983. doi: 10.3390/jcm14061983. PMID: 40142791; PMCID: PMC11942733.

24. *Pirkle S., Yang J., Blumberg T.J.* Do ChatGPT and Gemini Provide Appropriate Recommendations for Pediatric Orthopaedic Conditions? J Pediatr Orthop. 2025 Jan 1;45(1):e66-e71. doi: 10.1097/BPO.0000000000002797. Epub 2024 Aug 22. PMID: 39171426.

25. *Singh R., Hamouda M., Chamberlin J.H., Tóth A., Munford J., Silbergleit M., Baruah D., Burt J.R., Kabakus I.M.* ChatGPT vs. Gemini: Comparative accuracy and efficiency in Lung-RADS score assignment from radiology reports. Clin Imaging. 2025 May;121:110455. doi: 10.1016/j.clinimag.2025.110455. Epub 2025 Mar 13. PMID: 40090067.

26. *Silbergleit M., Tóth A., Chamberlin J.H., Hamouda M., Baruah D., Derrick S., Schoepf U.J., Burt J.R., Kabakus I.M.* ChatGPT vs Gemini: Comparative Accuracy and Efficiency in CAD-RADS Score Assignment from Radiology Reports. J Imaging Inform Med. 2025 Aug;38(4):2303-2311. doi: 10.1007/s10278-024-01328-y. Epub 2024 Nov 11. PMID: 39528887; PMCID: PMC12343400.

27. *Güneş Y.C., Cesur T., Çamur E., Günbey Karabekmez L.* Evaluating text and visual diagnostic capabilities of large language models on questions related to the Breast Imaging Reporting and Data System Atlas 5th edition. Diagn Interv Radiol. 2025 Mar 3;31(2):111-129. doi: 10.4274/dir.2024.242876. Epub 2024 Sep 9. PMID: 39248152; PMCID: PMC11880873.

28. *Lee K.L., Kessler D.A., Caglic I., Kuo Y.H., Shaida N., Barrett T.* Assessing the performance of ChatGPT and Bard/Gemini against radiologists for Prostate Imaging-Reporting and Data System classification based on prostate multiparametric MRI text reports. Br J Radiol. 2025 Mar 1;98(1167):368-374. doi: 10.1093/bjr/tqae236. PMID: 39535870; PMCID: PMC11840166.

29. *Wang L., Li J,, Zhuang B., Huang S., Fang M., Wang C., Li W., Zhang M., Gong S.* Accuracy of Large Language Models When Answering Clinical Research Questions: Systematic Review and Network Meta-Analysis.

J Med Internet Res. 2025 Apr 30;27:e64486. doi: 10.2196/64486. PMID: 40305085; PMCID: PMC12079073.

30. *Tippareddy C., Jiang S., Bera K., Ramaiya N.* Radiology Reading Room for the Future: Harnessing the Power of Large Language Models Like ChatGPT. Curr Probl Diagn Radiol. 2025 Jul-Aug;54(4):408-413. doi: 10.1067/j.cpradiol.2023.08.018. Epub 2023 Aug 30. PMID: 37758604.

31. *Fujimoto M., Kuroda H., Katayama T., Yamaguchi A., Katagiri N., Kagawa K., Tsukimoto S., Nakano A., Imaizumi U., Sato-Boku A., Kishimoto N., Itamiya T., Kido K., Sanuki T.* Evaluating Large Language Models in Dental Anesthesiology: A Comparative Analysis of ChatGPT-4, Claude 3 Opus, and Gemini 1.0 on the Japanese Dental Society of Anesthesiology Board Certification Exam. Cureus. 2024 Sep 27;16(9):e70302. doi: 10.7759/cureus.70302. PMID: 39469383; PMCID: PMC11513210.

32. *Wu J., Wang Z., Qin Y.* Performance of DeepSeek-R1 and ChatGPT-4o on the Chinese National Medical Licensing Examination: A Comparative Study. J Med Syst. 2025 Jun 3;49(1):74. doi: 10.1007/s10916-025-02213-z. PMID: 40459679.

33. Tippareddy C., Jiang S., Bera K., Ramaiya N. Radiology Reading Room for the Future: Harnessing the Power of Large Language Models Like ChatGPT. Curr Probl Diagn Radiol. 2025 Jul-Aug;54(4):408-413. doi: 10.1067/j.cpradiol.2023.08.018. Epub 2023 Aug 30. PMID: 37758604.

*** Corresponding Author:**
Cihan Bedel, MD, Health Science University Antalya Training And Research Hospital
Postal code: Kazım Karabekir Street postal zip code: 07100, Muratpaşa, Antalya, Turkey
E-mail: cihanbedel@hotmail.com
Phone +905075641254, Fax: +902422494487