

DO CHATBOTS PROVIDE RELIABLE INFORMATION ABOUT MOBILE APPS IN AUDIOLOGY?

Małgorzata Pastucha^{1A-F}

Institute of Physiology and Pathology of Hearing, World Hearing Center, Department of Experimental Audiology, Warsaw/Kajetany, Poland

Corresponding author: Małgorzata Pastucha, Department of Experimental Audiology, Institute of Physiology and Pathology of Hearing, ul. Mokra 17, Kajetany, 05-830 Nadarzyn, Poland; email: m.pastucha@ifps.org.pl

Contributions:
A Study design/planning
B Data collection/entry
C Data analysis/statistics
D Data interpretation
E Preparation of manuscript
F Literature analysis/search
G Funds collection

Abstract

Introduction: In light of the growing interest in utilizing AI for information retrieval, assessing the accuracy and reliability of tools such as chatbots is essential. This study aims to evaluate the efficacy of chatbots in providing accurate information about mobile applications (apps) in the field of audiology.

Material and methods: The responses of the Gemini and ChatGPT chatbots to eight open-ended questions posed in Polish and English were compared. Each answer was assessed for correctness.

Results: Gemini_ENG achieved the highest correctness with a score of 5 points (62.5%), while ChatGPT_PL scored 2 points (25%), and both Gemini_PL and ChatGPT_ENG scored 1 point (12.5%). Chatbots were most effective in recommending apps for older adults, with three of the four chatbots providing accurate recommendations. However, they struggled when asked to recommend apps for non-English speakers, to describe apps, or to provide direct links, with none of them scoring points in these areas.

Conclusions: Chatbots are currently unreliable sources of information about audiology apps. Depending on the language, there is significant variability in response accuracy. A good example is that Gemini_ENG performed far better than Gemini_PL. A major issue for all of them was the frequent fabrication of data, including the creation of nonexistent app names and incorrect links.

Keywords: ChatGPT • Gemini • mobile apps • hearing test • Polish • English

CZY CHATBOTY MOGĄ DOSTARCZAĆ WIARYGODNYCH INFORMACJI O APLIKACJACH MOBILNYCH W AUDIOLOGII?

Streszczenie

Wprowadzenie: W świetle rosnącego zainteresowania wykorzystaniem sztucznej inteligencji do wyszukiwania informacji ocena dokładności i niezawodności narzędzi takich jak chatboty jest niezbędna. Niniejsze badanie ma na celu ocenę skuteczności chatbotów w dostarczaniu dokładnych informacji o aplikacjach mobilnych w dziedzinie audiologii.

Materiał i metody: Porównano odpowiedzi chatbotów Gemini i ChatGPT na osiem pytań otwartych zadanych w języku polskim i angielskim. Każda odpowiedź była oceniana pod kątem poprawności.

Wyniki: Gemini_ENG osiągnął najwyższą poprawność z wynikiem 5 punktów (62,5%), podczas gdy ChatGPT_PL uzyskał 2 punkty (25%), a zarówno Gemini_PL, jak i ChatGPT_ENG uzyskały 1 punkt (12,5%). Chatboty były najbardziej skuteczne w polecaniu aplikacji dla osób starszych, przy czym trzy z czterech chatbotów zapewniały dokładne rekomendacje. Miały one jednak trudności, gdy poproszono je o polecenie aplikacji dla osób nieanglojęzycznych, opisywanie aplikacji lub dostarczanie bezpośrednich linków, przy czym żaden z nich nie zdobył punktów w tych obszarach.

Wnioski: Chatboty są obecnie niewiarygodnym źródłem informacji o aplikacjach audiologicznych. W zależności od języka istnieje znaczna zmienność w dokładności odpowiedzi. Dobrym przykładem jest to, że Gemini_ENG działał znacznie lepiej niż Gemini_PL. Głównym problemem dla wszystkich z nich była częsta fabrykacja danych, w tym tworzenie nieistniejących nazw aplikacji i nieprawidłowych linków.

Słowa kluczowe: ChatGPT • Gemini • aplikacje mobilne • badanie słuchu • polski • angielski

Introduction

Chatbots, advanced conversational tools based on artificial intelligence (AI), are able to conduct natural language dialogues [1]. Following an initial training phase on large data sets, they generate responses to queries based on a wide range of information available on the internet. Due to their potential applications in science and medicine [2], chatbots are gaining widespread interest, with researchers testing them in various health sectors [3–5].

Another example of rapidly developing technology in healthcare is mobile apps [6,7]. In the field of audiology, rapid technology development has led to the marketing of a number of apps for testing hearing [8]. The apps have good sensitivity and specificity, and a number of studies have demonstrated they have potential for screening purposes [9–11]. However, the plethora of apps can make it difficult for a user to choose the most appropriate one. Patients may therefore seek sources of information in this area. One source of information might include a chatbot, which might appeal to a patient who, for various reasons, may find it difficult to access a specialist [12]. However, the extent to which chatbots can provide relevant information about mobile apps in the field of hearing is unknown.

To date, there have been few studies in the field of audiology on the use of chatbots. Existing works have mainly concerned the extent to which AI can handle specialized questions in audiology [13–16]. The answers vary depending on the version of the chatbot used [14] and may change over time [13,15]. In addition, the accuracy of the answers can be affected by the form of the question asked, e.g. whether it is an open-ended question [13] or multiple-choice [15]. Moreover, the responses generated by chatbots can differ depending on the language used in the query, as different databases will be searched [17].

In summary, the efficacy of chatbots in providing correct audiological information is unknown. To date, no research has yet been done to verify the accuracy of information provided by chatbots about mobile apps for testing hearing. The objective of this study is therefore to test whether common chatbots are able to supply reliable information

about mobile apps for testing hearing. Two types of chatbots were selected, and each was asked questions in two languages, English and Polish. The accuracy, correctness, and usefulness of the answers were assessed.

Material and methods

Two types of free chatbots – ChatGPT version 4o, (OpenAI, San Francisco, CA, USA) and Gemini (Google LLC, Mountain View, CA, USA) – were selected for study. Questions were asked in July 2024 in English and Polish, resulting in four versions for analysis: ChatGPT_ENG, ChatGPT_PL, Gemini_ENG, and Gemini_PL.

Eight open-ended questions related to hearing test apps were formulated (**Table 1**). These questions were based on information that could be verified with reference to specific studies [8,18] or by searching the mobile app market (e.g., Google Play Store, App Store). After asking questions of the chatbots, their answers were saved and analyzed for accuracy. It was checked whether the identified apps existed (questions 1–5), whether their descriptions were correct (question 6), and which apps the chatbot recommended (questions 7 and 8).

The responses to questions 1–4 were analyzed in terms of both the number of total answers provided and the number of correct answers. An answer was considered correct if the listed app was available on at least one platform (Google Play Store or App Store) and allowed a hearing test to be done.

For question 5, it was examined whether the chatbots provided a link to a specific app, and if so, whether the link was correct.

For question 6, the evaluation was in terms of:

- app description (accuracy, completeness, incorrect information);
- user rating (quantitative/qualitative; if quantitative, whether it matched the ratings given on the platform where the app is available),
- app availability on platforms (correct indication of availability on one or both platforms).

Table 1. List of questions asked of chatbots

| No. | Questions |
|-----|--|
| 1 | What mobile apps for testing hearing are currently available? |
| 2 | Are these apps available in Poland? |
| 3 | Are there mobile hearing test apps for children? |
| 4 | Are there any mobile hearing test apps specifically designed for children to perform so-called “play audiometry”? |
| 5 | Can you provide a link to these apps? |
| 6 | Can you point to specific examples of hearing testing apps with their description, including features, user ratings, and the platforms on which they are available? |
| 7 | Which mobile hearing test app would you recommend for a person who has manual dexterity difficulties, is unfamiliar with smartphones, or has increased reaction time? Explain your choice. |
| 8 | Can you recommend an app available in Polish for non-English speakers? |

Table 2. Summary of chatbot responses to questions about mobile apps for testing hearing

| Question No. | Chatbot | Number of answers given | Number of correct answers | Correctness of responses |
|--------------|-------------|-------------------------|---------------------------|--------------------------|
| 1 | Gemini_ENG | 3 | 3 | + |
| | Gemini_PL | 3 | 2 | - |
| | ChatGPT_ENG | 8 | 7 | - |
| | ChatGPT_PL | 4 | 4 | + |
| 2 | Gemini_ENG | 2 | 2 | + |
| | Gemini_PL | 2 | 2 | + |
| | ChatGPT_ENG | 5 | 3 | - |
| | ChatGPT_PL | 4 | 3 | - |
| 3 | Gemini_ENG | 1 | 1 | + |
| | Gemini_PL | 0 | 0 | - |
| | ChatGPT_ENG | 4 | 1 | - |
| | ChatGPT_PL | 4 | 1 | - |
| 4 | Gemini_ENG | 1 | 1 | + |
| | Gemini_PL | 0 | 0 | - |
| | ChatGPT_ENG | 3 | 1 | - |
| | ChatGPT_PL | 4 | 0 | - |
| 5 | Gemini_ENG | 0 | 0 | - |
| | Gemini_PL | 6 | 1 | - |
| | ChatGPT_ENG | 2 | 0 | - |
| | ChatGPT_PL | 0 | 0 | - |

For the recommendations (questions 7 and 8), it was checked whether the chatbots recommended existing apps, considered the specified user constraints, and whether their recommendations were based on certain criteria (e.g., features, method of conducting test, availability, customization options like language). The information provided by the chatbots was verified for accuracy.

The full chatbot responses to the questions in **Table 1** can be found in the online supplementary material.

Results

Questions 1–5: searching for mobile apps

Table 2 is a comprehensive summary of the responses provided by all versions of the chatbots to questions 1–5. In the last column, the correctness of the responses is marked with a “+” for correct answers and a “-” for incorrect ones. A minus sign was assigned if the number of correct answers was less than the number of provided answers, or if the chatbot did not provide the name of any app or a direct link, despite the availability of this information.

Detailed responses, including the specific names of the apps identified, can be found in the supplementary material.

ChatGPT_ENG generated the greatest number of responses to questions 1–4, achieving a score of 20. ChatGPT_PL

furnished a total of 16 responses. In contrast, Gemini_ENG provided just 7 responses, while Gemini_PL delivered the fewest, with only 5 in total. Notably, all of the Gemini_ENG answers were correct. Gemini_PL made an error by proposing an app that monitors music listening but cannot test hearing. Despite the fact that ChatGPT_ENG generated the greatest number of responses, its error rate was higher than the other chatbots. ChatGPT_PL demonstrated a lower error rate than ChatGPT_ENG, the only error being with the final question. Curiously, ChatGPT_PL pointed to the same app in response to two questions, giving its name once in Polish and once in English; however, because the app exists in both languages these answers were considered correct.

All chatbots were asked to provide links directing a user to the respective apps. Both Gemini_ENG and ChatGPT_PL reported that they were unable to provide direct links. Gemini_ENG argued that this was for security reasons, while ChatGPT_PL said it was because of not having access to the internet and not being able to view up-to-date sources. Nevertheless, Gemini_ENG included a direct link to a specific app in its response to the first question (see Appendix 1 of the supplementary material). On the positive side, both chatbots, Gemini_ENG and ChatGPT_PL, furnished step-by-step instructions on how to search for a specific app. Gemini_ENG provided examples of existing apps, while ChatGPT_PL gave the name of an app that didn't exist. Strangely, ChatGPT_ENG and Gemini_PL

Table 3. Summary of chatbot responses to questions about app descriptions

| Chatbot | Number of apps provided | Number of correct apps | Platforms* (chatbot answer compared to availability on platform)** | Correct description of app** | Quantitative user rating: concordance of responses** | Concordance of all responses |
|-------------|-------------------------|------------------------|--|------------------------------|--|------------------------------|
| Gemini_ENG | 3 | 3 | 4/3 | 3 | 0 | – |
| Gemini_PL | 3 | 2 | 4/2 | 2 | 1 | – |
| ChatGPT_ENG | 5 | 5 | 10/9 | 4 | 1 | – |
| ChatGPT_PL | 5 | 2 | 3/2 | 2 | 0 | – |

* App Store and/or Google Play Store, ** Concerning existing hearing test apps

provided direct links to apps on both platforms, but the app named by ChatGPT_ENG was not available on either platform, and incorrect links were given. At least Gemini_PL was able to provide one correct link.

Question 6 – description of apps

Table 3 provides a summary of the responses given by all versions of the chatbots to the question about specific examples of hearing test apps. The analysis considered the descriptions, user ratings, and the platforms on which the apps were available. Column 4 shows data on the number of platforms (App Store and Google Play) on which apps were available (reported by chatbots and compared to their actual availability). The “Correct description of app” column shows the number of apps for which the chatbots provided descriptions containing all the correct information. The “Quantitative user rating: concordance of responses” column, on the other hand, shows the number of apps for which the chatbots provided consistent quantitative ratings, corresponding to those posted on platforms such as the App Store and Google Play. If any of these criteria were not met, the response was considered incorrect. In the last column, a plus sign is assigned if correct answers were all given in the other columns; otherwise, a minus sign is assigned.

Detailed responses, including descriptions of specific apps, can be found in the supplementary material.

Both versions of Gemini described 3 mobile apps, and both versions of ChatGPT described 5 apps each. However, Gemini_PL and ChatGPT_PL described one app that was designed for noise assessment rather than hearing testing. In addition, ChatGPT_PL described 2 apps that do not exist.

None of the chatbots correctly answered the question about the availability of the app on specific platforms (considering only existing hearing test apps). In one instance, both versions of ChatGPT and Gemini_ENG incorrectly indicated that the app was available on both platforms, despite the fact that it was, in fact, only accessible on one. Gemini_ENG made this error when describing two apps.

The descriptions given by Gemini_ENG were short, one-sentence descriptions. They were based on a description of the apps available on the platform. Although they were correct, there was a lack of detailed information about the

functions of the app and the types of tests it could perform. Gemini_PL, on the other hand, focused mainly on listing the functions each app offered. The descriptions of ChatGPT_PL were also short, but the information was more accurate: the purpose of the app, the available tests, and the target group were clearly stated. The information was correct, except of course for the non-existent apps. ChatGPT_ENG gave the most comprehensive answers. They consisted of one or two sentences and gave an indication of the features of the app. All descriptions were correct, except for one in which the form of the test was incorrectly stated (pure tone audiometry instead of triplets-in-noise test).

In terms of users’ evaluations of the apps, both versions of ChatGPT and Gemini_PL provided quantitative user ratings. However, only in two cases was the rating awarded consistent with the rating visible on the platform: (1) ChatGPT_ENG, giving the rating of the “Petralex Hearing Aid” app in the App Store; and (2) Gemini_PL, giving ratings for the “Mimi Hearing Test” app in the App Store. It should also be added that ChatGPT_ENG gave only a general number of ratings (e.g. thousands or hundreds), whereas ChatGPT_PL and Gemini_PL gave the exact number of times the app had been rated. In contrast, Gemini_ENG provided only a qualitative assessment, i.e. whether it was rated positively or not. However, the assessment was largely inconclusive, with the chatbot emphasizing that it was only a subjective rating made by users.

Questions 7 and 8 – recommendations

Table 4 provides a summary of the recommendations provided by all versions of the chatbots. A response was deemed accurate (column 4) if the recommended app was for a hearing test. If there was even one false statement in the recommendation rationale (column 5), it was considered an error. In the last column, the correctness of the responses is marked with a plus or minus symbol. A minus sign was also assigned in cases where the chatbot did not make a recommendation, even though there was a corresponding app.

In response to question 7 (limitations of an older person), both versions of Gemini proposed two apps, whereas the two versions of ChatGPT proposed one each. However, one of the apps recommended by Gemini_PL is used for monitoring hearing during music listening and does not

Table 4. Summary of the recommendations given by all versions of chatbots

| Recommendations for | Chatbot | Number of answers given | Number of correct answers | Number of correct justifications provided for the assessment (no false information)* | Correctness of responses |
|-----------------------------|-------------|-------------------------|---------------------------|--|--------------------------|
| Older person | Gemini_ENG | 2 | 2 | 2 | + |
| | Gemini_PL | 2 | 1 | 1 | – |
| | ChatGPT_ENG | 1 | 1 | 1 | + |
| | ChatGPT_PL | 1 | 1 | 1 | + |
| Non-English speaking person | Gemini_ENG | 0 | 0 | 0 | – |
| | Gemini_PL | 3 | 1 | 0 | – |
| | ChatGPT_ENG | 1 | 0 | 0 | – |
| | ChatGPT_PL | 1 | 0 | 0 | – |

* Concerning only existing hearing test apps

Table 5. Summary of the correctness of the chatbots' answers to all questions

| Question No. | Gemini_ENG | Gemini_PL | ChatGPT_ENG | ChatGPT_PL |
|----------------------------|-------------|-------------|-------------|------------|
| 1 | + | – | – | + |
| 2 | + | + | – | – |
| 3 | + | – | – | – |
| 4 | + | – | – | – |
| 5 | – | – | – | – |
| 6 | – | – | – | – |
| 7 | + | – | + | + |
| 8 | – | – | – | – |
| Total points | 5 | 1 | 1 | 2 |
| % correct responses | 62.5 | 12.5 | 12.5 | 25 |

include additional hearing test functionality. All chatbots took into account the indicated limitations of the user and the apps they proposed are currently available on the market. Both versions of Gemini gave only brief explanations of its app choices. In contrast, both ChatGPT versions based their recommendations on detailed information about the app's features, the type of test to be conducted, and availability on different platforms.

In question 8, chatbots were asked to recommend apps for Polish users who do not speak English. Gemini_ENG indicated at the outset that it might be difficult to find such an app. It then recommended the apps previously identified, adding information that they may not be available in Polish. Chatbot also recommended using a translator or choosing an app with a simple interface. In contrast, Gemini_PL recommended 3 apps, but only one of them enables users to undertake a hearing test. It is also noteworthy that the description of this app provided by the chatbot contained incorrect information. Both versions of ChatGPT indicated, and described in detail, a selected app, but this app does not actually exist.

A detailed description of the user limitations and the exact responses of the chatbots is provided the supplementary material.

Table 5 presents a summary of the correctness of the responses provided by the chatbots to all the questions posed. The last two lines shows the number of points obtained and the percentage of correct responses.

The results demonstrate that Gemini_ENG achieved a score of 5 points (62.5%), ChatGPT_PL attained a score of 2 points (25%), and Gemini_PL and ChatGPT_ENG scored 1 point each (12.5%). On this basis, Gemini_ENG clearly outperformed the other chatbots.

Discussion

The objective of this study was to assess the effectiveness of chatbots in providing accurate information about mobile apps in the field of audiology. Two chatbots, Gemini and ChatGPT, were employed for the analysis of responses in both Polish and English.

The results indicate that AI cannot currently be considered a reliable source of information about mobile apps in audiology. Significant variability was observed in the correctness of responses depending on the language of the query, the type of chatbot, and the context of the question.

Among the tested chatbots, Gemini_ENG proved to be the most reliable, providing correct answers to more than half the questions posed. It outperformed other chatbots in terms of the correctness of responses, even compared to its Polish counterpart, Gemini_PL, which performed poorly. This disparity probably stems from the broader and more comprehensive training database available in English. The richer English dataset allows for better verification and validation of information, resulting in higher correctness. Similar observations were made by Jędrzejczak et al. [14], who recommend asking questions in English. Likewise, another study [19] found that ChatGPT 3.5 was unable to provide references for queries in Italian and Spanish because these versions don't include references in these languages. By way of contrast, ChatGPT (version 4o) did not exhibit significant language-related discrepancies in this study, suggesting a more consistent performance across languages.

There were noteworthy differences in the ways in which Gemini and ChatGPT provided answers. While Gemini focused on providing concise, relevant information, ChatGPT's responses were more detailed. Descriptions of apps were longer and explanations were more comprehensive. Similar observations have been reported in studies conducted by others [14]. In contrast, in questions about hypertension, Gemini gave more elaborate answers than ChatGPT [20]. This indicates that the length of the responses may depend on the specific domain of the questions.

Most chatbots demonstrated higher proficiency in recommending apps tailored to the needs of older people, with three of the four chatbots providing accurate responses. The exception was Gemini_PL, which failed in this area,

References

1. Yang HS, Wang F, Greenblatt MB, Huang SX, Zhang Y. AI chatbots in clinical laboratory medicine: foundations and trends. *Clin Chem*, 2023; 69(11): 1238–46. <https://doi.org/10.1093/clinchem/hvad106>
2. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*, 2023; 29(8): 1930–40. <https://doi.org/10.1038/s41591-023-02448-8>
3. Biswas S, Logan NS, Davies LN, Sheppard AL, Wolffsohn JS. Assessing the utility of ChatGPT as an artificial intelligence-based large language model for information to answer questions on myopia. *Ophthalmic Physiol Opt*, 2023; 43(6): 1562–70. <https://doi.org/10.1111/opo.13207>
4. Eggmann F, Weiger R, Zitzmann NU, Blatz MB. Implications of large language models such as ChatGPT for dental medicine. *J Esthet Restor Dent*, 2023; 35(7): 1098–102. <https://doi.org/10.1111/jerd.13046>
5. Valentini M, Szkandera J, Smolle MA, Scheipl S, Leithner A, Andreou D. Artificial intelligence large language model ChatGPT: is it a trustworthy and reliable source of information for sarcoma patients? *Front Public Health*, 2024; 12: 1303319. <https://doi.org/10.3389/fpubh.2024.1303319>
6. Okolo CA, Babawarun O, Arowoogun JO, Adeniyi AO, Chidi R. The role of mobile health applications in improving patient engagement and health outcomes: a critical review. *Int J Sci Res Arch*, 2024; 11(1): 2566–74. <https://doi.org/10.30574/ijrsra.2024.11.1.0334>
7. Galetsi P, Katsaliaki K, Kumar S. Exploring benefits and ethical challenges in the rise of mHealth (mobile healthcare) technology for the common good: an analysis of mobile applications for health specialists. *Technovation*, 2023; 121: 102598. <https://doi.org/10.1016/j.technovation.2022.102598>
8. Almufarrij I, Dillon H, Dawes P, Moore DR, Yeung W, Charalambous A-P, et al. Web- and app-based tools for remote hearing assessment: a scoping review. *Int J Audiol*, 2023; 62(8): 699–712. <https://doi.org/10.1080/14992027.2022.2075798>

probably due to restricted access to Polish-language sources. However, it was the sole chatbot to accurately identify an app suitable for non-English speakers, despite including erroneous information in the description.

In general, chatbots performed better on questions where general information was sought. The number of incorrect or unanswered answers increased for questions which included specific details, such as names, numbers, or direct links. These findings align with those of other studies, which indicate that ChatGPT has inferior performance on more specialized questions, such as those pertaining to sarcoma treatment, compared to general questions [5]. Furthermore, ChatGPT_ENG was less inclined to concede ignorance, frequently providing fictitious data [13,19].

Conclusions

In the light of these findings, it can be concluded that, at the time of writing, chatbots cannot be considered reliable sources of information for mobile apps in the field of audiology. Among the chatbots tested, Gemini_ENG exhibited the highest level of correctness. However, the responses of the chatbots were variable and depended on the linguistic context in which the queries were posed. Notably, the Polish version of Gemini demonstrated a serious deficiency in accuracy and correctness compared to its English counterpart. A particularly concerning issue affecting all chatbots was the widespread and egregious error of supplying fabricated data, including the names of nonexistent apps and providing erroneous links to them.

Acknowledgements

The author thanks Prof. W. Wiktor Jędrzejczak for stimulating discussions on earlier versions of the manuscript.

Supplementary material

Supplementary material is available at <https://www.journalofhearingscience.com/>

9. Masalski M, Grysiński T, Kręcicki T. Hearing tests based on biologically calibrated mobile devices: comparison with pure-tone audiometry. *JMIR Mhealth Uhealth*, 2018; 6(1): e10. <https://doi.org/10.2196/mhealth.7800>
10. Potgieter J-M, Swanepoel DW, Smits C. Evaluating a smartphone digits-in-noise test as part of the audiometric test battery. *S Afr J Commun Disord*, 2018; 65(1): 574. <https://doi.org/10.4102/sajcd.v65i1.574>
11. Dillon H, Mee C, Moreno JC, Seymour J. Hearing tests are just child's play: the sound scouts game for children entering school. *Int J Audiol*, 2018; 57(7): 529–37. <https://doi.org/10.1080/14992027.2018.1463464>
12. Zuhair V, Babar A, Ali R, et al. Exploring the Impact of artificial intelligence on global health and enhancing healthcare in developing nations. *J Prim Care Community Health*, 2024; 15: 21501319241245847. <https://doi.org/10.1177/21501319241245847>
13. Jedrzejczak WW, Skarzynski PH, Raj-Koziak D, Sanfins MD, Hatzopoulos S, Kochanek K. ChatGPT for tinnitus information and support: response accuracy and retest after three and six months. *Brain Sci*, 2024; 14(5): 465. <https://doi.org/10.3390/brainsci14050465>
14. Jedrzejczak WW, Kochanek K. Comparison of the audiological knowledge of three chatbots: ChatGPT, Bing Chat, and Bard. *Audiol Neurotol*, 2024; 1–7. <https://doi.org/10.1159/000538983>
15. Kochanek K, Skarzynski H, Jedrzejczak WW. Accuracy and repeatability of ChatGPT based on a set of multiple-choice questions on objective tests of hearing. *Cureus*, 2024; 16(5): e59857. <https://doi.org/10.7759/cureus.59857>
16. Wang S, Mo C, Chen Y, Dai X, Wang H, Shen X. Exploring the performance of ChatGPT-4 in the Taiwan Audiologist Qualification Examination: preliminary observational study highlighting the potential of AI chatbots in hearing care. *JMIR Med Educ*, 2024; 10: e55595. <https://doi.org/10.2196/55595>
17. Chayka A, Sukhov A. Comparing chatbots trained in different languages. *Commun ACM*, 2023; 66(12): 6–7. <https://doi.org/10.1145/3626835>
18. Pastucha M, Kochanek K, Piłka E, Jedrzejczak WW. [Review of mobile apps for self-testing of hearing in Polish language]. *Now Audiofonol*, 2024; 13(1): 21–31 [in Polish]. <https://doi.org/10.17431/na/173553>
19. McGowan A, Gui Y, Dobbs M, Shuster S, Cotter M, Selloni A, et al. ChatGPT and Bard exhibit spontaneous citation fabrication during psychiatry literature search. *Psychiatry Res*, 2023; 326: 115334. <https://doi.org/10.1016/j.psychres.2023.115334>
20. Lee TJ, Campbell DJ, Patel S, Hossain A, Radfar N, Siddiqui E, et al. Unlocking health literacy: the ultimate guide to hypertension education from ChatGPT versus Google Gemini. *Cureus*, 2024; 16(5): e59898. <https://doi.org/10.7759/cureus.59898>