Received: 12 May 2025 / Accepted: 19 August 2025 / Published online: 28 August 2025

DOI 10.34689/SH.2025.27.4.014

UDC 616.831-008.918:004.8

EVALUATION OF READABILITY INDICES OF CHATGPT-4 AND GOOGLE GEMINI IN PATIENT EDUCATION ABOUT INTRACRANIAL HEMORRHAGES

Umut Ogün Mutlucan[1], https://orcid.org/0000-0002-5052-7244

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

Ökkeş Zortuk[3], https://orcid.org/0000-0001-6776-2702

Fatih Selvi[2], https://orcid.org/0000-0002-9701-9714

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

Abstract

Aim. Subarachnoid haemorrhage (SAH) and intracranial aneurysms are critical neurological conditions with significant implications for patient morbidity and mortality. The intersection of readability indices and artificial intelligence (AI) is an emerging field that aims to improve the accessibility and understanding of written material in different areas. Readability indices, such as the Automated Readability Index (ARI) and the Flesch–Kincaid grade level, provide quantitative measures of text complexity that are crucial for tailoring content to specific audiences . Therefore, in this study, we aimed to examine the answers given to questions asked by patients with intracranial haemorrhage using readability indices.

Materials and Methods. In this study, questions directly posed by patients and their relatives concerning subarachnoid haemorrhage and intracranial aneurysms were compiled. The collated questions were then divided into subcategories, including definition, diagnosis, treatment options, surgical procedures, complications, and impact on daily life. Flesch Reading Ease (FRE) Formula, Fog Scale (Gunning FOG Formula), SMOG Index, Automated Readability Index (ARI),

Coleman-Liau Index, Linsear Write Formula, Dale-Chall Readability Score, Spache Readability Formula. AI technologies were compared across groups.

Results. The results indicate that for most readability indices, there is no statistically significant difference between the two models, with one notable exception; Coleman-Liau Readability Index: Gemini: 10.59 ± 0.98 vs. ChatGPT: 11.80 ± 1.64; p-Value: 0.014 The only exception is the Coleman-Liau Readability Index, where a statistically significant difference was found, with ChatGPT showing a slightly higher score, implying potentially greater complexity according to that specific measure.

Conclusion. Our article provides valuable quantitative data on the readability of texts from ChatGPT and Gemini, its scope is narrow. A more comprehensive study would ideally include qualitative assessments, a broader range of text types, and detailed information on the methodology and model versions to provide a more holistic understanding of the models' performance.

Keywords: *Subarachnoid haemorrhage, intracranial aneurysms, Coleman-Liau Readability.*

For citation:

Mutlucan U.O., Bedel C., Zortuk Ö., Selvi F. Evaluation of readability indices of ChatGPT-4 and Google Gemini in patient education about intracranial hemorrhages // *Nauka i Zdravookhranenie* [Science & Healthcare]. 2025. Vol.27 (4), pp. 107-112. *doi 10.34689/SH.2025.27.4.014*

Резюме

ОЦЕНКА ПОКАЗАТЕЛЕЙ ЧИТАЕМОСТИ CHATGPT-4 И GOOGLE GEMINI В ОБУЧЕНИИ ПАЦИЕНТОВ ПО ВОПРОСАМ ВНУТРИЧЕРЕПНЫХ КРОВОИЗЛИЯНИЙ

Умут Огюн Мутлукан[1], https://orcid.org/0000-0002-5052-7244

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

Оккеш Зортук[3], https://orcid.org/0000-0001-6776-2702

Фатих Сельви[2], https://orcid.org/0000-0002-9701-9714

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

Актуальность. Субарахноидальное кровоизлияние (САК) и внутричерепные аневризмы являются критическими неврологическими состояниями, имеющими значительные последствия для заболеваемости и смертности пациентов. Пересечение индексов читаемости и искусственного интеллекта (ИИ) является новой областью, направленной на улучшение доступности и понимания письменных материалов в различных областях. Индексы читаемости, такие как автоматический индекс читаемости (ARI) и уровень сложности текста по шкале Флеша-Кинкейда, предоставляют количественные показатели сложности текста, которые имеют решающее значение для адаптации контента к конкретной аудитории. Поэтому в данном исследовании мы поставили **цель** проанализировать ответы на вопросы, заданные пациентами с внутричерепным кровоизлиянием, с использованием индексов читаемости.

Материалы и методы. В данном исследовании были собраны вопросы, непосредственно заданные пациентами и их родственниками, касающиеся субарахноидального кровоизлияния и внутричерепных аневризм. Затем собранные вопросы были разделены на подкатегории, включая определение, диагностику, варианты лечения, хирургические процедуры, осложнения и влияние на повседневную жизнь. Формула читабельности Флеша (FRE), шкала Fog (формула Gunning FOG), индекс SMOG, автоматический индекс читабельности (ARI), индекс Коулмана-Ляу, формула Линсеара, шкала читабельности Дейла-Чалла, формула читабельности Спаче. Технологии искусственного интеллекта были сравнены между группами.

Результаты. Результаты показывают, что для большинства индексов читаемости нет статистически значимой разницы между двумя моделями, за одним заметным исключением; Индекс читаемости Коулмана-Лиау: Gemini: 10,59 ± 0,98 против ChatGPT: 11,80 ± 1,64; p-значение: 0,014 Единственным исключением является индекс читаемости Коулмана-Лиау, где была обнаружена статистически значимая разница: ChatGPT показал немного более высокий балл, что означает потенциально большую уровень сложности по этому конкретному показателю.

Заключение. Наша статья предоставляет ценные количественные данные о читаемости текстов ChatGPT и Gemini, однако ее область применения ограничен. Более комплексное исследование в идеале должно включать качественную оценку, более широкий спектр типов текстов и подробную информацию о методологии и версиях моделей, чтобы обеспечить более целостное понимание производительности моделей.

Ключевые слова: *субарахноидальное кровоизлияние, внутричерепные аневризмы, индекс читаемости Коулмана-Ляу.*

Для цитирования:

Мутлукан У.О., Бедел Дж., Зортук О., Сельви Ф. Оценка показателей читаемости CHATGPT-4 и Google Gemini в обучении пациентов по вопросам внутричерепных кровоизлияний // Наука и Здравоохранение. 2025. Vol.27 (4), С.107-112. *doi 10.34689/SH.2025.27.4.014*

Түйіндеме

ПАЦИЕНТТЕРДІҢ БАС СҮЙЕК ІШІНЕ ҚАН ҚҰЙЫЛУ БОЙЫНША БІЛІМ АЛУЫНДА CHATGPT-4 ЖӘНЕ GOOGLE GEMINI ОҚУ КӨРСЕТКІШТЕРІН БАҒАЛАУ

Умут Огюн Мутлукан[1], https://orcid.org/0000-0002-5052-7244

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

Оккеш Зортук[3], https://orcid.org/0000-0001-6776-2702

Фатих Сельви[2], https://orcid.org/0000-0002-9701-9714

[1] Анталия медицина ғылымдары университетінің нейрохирургия бөлімі, білім беру және ғылыми-зерттеу ауруханасы, Анталия, Түркия;

[2] Анталия медициналық ғылымдар университетінің жедел медициналық көмек бөлімі, білім беру және ғылыми-зерттеу ауруханасы, Анталия, Түркия;

[3] Жедел медициналық көмек бөлімі, Бандырма университетінің ауруханасы, Балыкесир, Түркия.

Мақсаты. Субарахноидты қан кету (SAH) және бас сүйек іші аневризмалары пациенттердің аурушаңдығы мен өліміне айтарлықтай әсер ететін маңызды неврологиялық жағдайлар болып табылады. Оқу және жасанды интеллект (AI) индекстерінің тоғысуы әртүрлі салалардағы жазбаша материалдардың қолжетімділігі мен түсінігін жақсартуға бағытталған жаңа сала болып табылады. Автоматты оқылым индексі (ARI) және Флэш-Кинкейд шкаласы бойынша мәтіннің күрделілік деңгейі сияқты оқылым индекстері мазмұнды белгілі бір аудиторияға бейімдеу үшін маңызды мәтіннің күрделілігінің сандық көрсеткіштерін ұсынады. Сондықтан, осы зерттеуде біз бас сүйек ішіне қан құйылумен ауыратын науқастардың оқылым индекстерін қолдана отырып қойған сұрақтарына жауаптарды талдауды мақсат етіп қойдық.

Материалдар мен әдістер. бұл зерттеуде пациенттер мен олардың туыстары субарахноидты қан кетуге және бас сүйек іші аневризмасына қатысты тікелей қойған сұрақтары жинақталды. Содан кейін жинақталған сұрақтар анықтау, диагностика, емдеу нұсқалары, хирургиялық процедуралар, асқынулар және күнделікті өмірге әсер ету сияқты кіші санаттарға бөлінді. Флештің оқылым формуласы (FRE), Fog шкаласы (Gunning FOG формуласы), SMOG индексі, Автоматты оқылым индексі (ARI), Коулман-Ляу индексі, Линсеар формуласы, Дейл-Чаллдың оқылым шкаласы, Спаче оқылым формуласы. Жасанды интеллект технологиялары топтар арасында салыстырылды

Нәтижелер. Нәтижелер көптеген оқылым индекстері үшін екі модель арасында статистикалық маңызды айырмашылық жоқ екенін көрсетеді, тек бір ерекше жағдайды қоспағанда; Коулман-Лиаудың оқылу индексі: Gemini: 10,59 ± 0,98 керісінше ChatGPT: 11,80 ± 1,64; p-мәні: 0,014 жалғыз ерекшелік-Коулман-Лиаудың оқылу индексі, онда статистикалық маңызды айырмашылық: ChatGPT сәл жоғары балл көрсетті, бұл нақты көрсеткіш бойынша күрделіліктің ықтималдылығының жоғары деңгейін білдіреді.

Қорытынды. Біздің мақалада ChatGPT және Gemini мәтіндерінің оқылуы туралы құнды сандық мәліметтер келтірілген, бірақ оның қолдану аясы шектеулі. Неғұрлым жан-жақты зерттеу модельдердің өнімділігі туралы біртұтас түсінік беру үшін сапалы бағалауды, мәтін түрлерінің кең спектрін және модельдердің әдістемесі мен нұсқалары туралы егжей-тегжейлі ақпаратты қамту керек.

***Түйін сөздер**: субарахноидты қан кету, бас сүйек іші аневризмалары, Коулман-Ляу оқу индексі.*

Дәйексөз үшін:

Мутлукан У.О., Бедел Дж., Зортук О., Сельви Ф. Пациенттердің бас сүйек ішіне қан құйылу бойынша білім алуында CHATGPT-4 және Google Gemini оқу көрсеткіштерін бағалау // Ғылым және Денсаулық сақтау. 2025. Vol.27 (4), Б. 107-112. *doi 10.34689/SH.2025.27.4.014*

Introduction

Subarachnoid haemorrhage (SAH) and intracranial aneurysms are critical neurological conditions with significant implications for patient morbidity and mortality. SAH is primarily caused by the rupture of an aneurysm in the brain, leading to bleeding in the subarachnoid space [1]. This condition is associated with high mortality and disability rates, particularly in younger populations. Managing and treating SAH and intracranial aneurysms involves making complex clinical decisions and often requires a multidisciplinary approach [2]. Although SAH accounts for only a small percentage of all strokes, it has a disproportionate impact due to its high morbidity and mortality rates, particularly in younger individuals, with an average age of onset of 50–55 years [3].

Artificial intelligence (AI) has emerged as a transformative tool in the detection and management of these conditions. Early detection and intervention are critical for these conditions due to their high morbidity and mortality rates [4,5]. AI has been integrated into clinical workflows to support decision-making in the management of SAH. This includes predicting complications, optimising treatment strategies and reducing the time to intervention, all of which are crucial for improving survival rates [6]. The intersection of readability indices and artificial intelligence (AI) is an emerging field that aims to improve the accessibility and understanding of written material in different areas. Readability indices, such as the Automated Readability Index (ARI) and the Flesch–Kincaid grade level, provide quantitative measures of text complexity that are crucial for tailoring content to specific audiences [7].

Therefore, in this study, we aimed to examine the answers given to questions asked by patients with intracranial haemorrhage using readability indices.

Metods. The present study did not require ethics committee approval as it did not involve any human or animal models.

Data Collection and Preparation

In this study, questions directly posed by patients and their relatives concerning subarachnoid haemorrhage and intracranial aneurysms were compiled. The questions presented herein were collated from a variety of online forums, neurosurgical websites, and the websites of relevant associations, hospitals and organisations.

The collated questions were then divided into subcategories, including definition, diagnosis, treatment options, surgical procedures, complications, and impact on daily life. From each category, the twenty most frequently asked questions were selected for analysis. Prior to analysis, these selected questions were reviewed and corrected for grammatical and semantic accuracy.

Response Generation by AI Models

The prepared questions were submitted to the ChatGPT-5 and Gemini-2.5 Pro artificial intelligence models to generate responses. Prior to the formulation of the inquiries, both AI models were instructed to formulate their responses in a language appropriate for a general audience with no prior medical knowledge on the subject. The responses generated by the models were recorded.

Readability Analysis and Statistical Evaluation

The readability levels of the responses generated by the AI models were measured using the nine different standard formulas listed below:

- ***Flesch Reading Ease (FRE) Formula:*** This formula calculates readability on a 100-point scale using sentence length and word syllables, where a higher score means the text is easier to understand [8].
- ***Flesch-Kincaid Grade Level (FKGL):*** This formula estimates the U.S. grade level required to comprehend a text by analyzing the average sentence length and syllables per word [9].
- ***Fog Scale (Gunning FOG Formula):*** The Gunning FOG Index estimates the years of formal education a person needs to understand a piece of writing upon first reading [10].
- ***SMOG Index:*** The SMOG Index determines the required years of education to understand a text by counting the number of words with three or more syllables [11].
- ***Automated Readability Index (ARI):*** The Automated Readability Index calculates the U.S. grade level of a text using character, word, and sentence counts, making it ideal for computer processing [12].
- ***Coleman-Liau Index:*** This computer-friendly formula estimates the U.S. grade level by measuring the average number of letters and sentences per 100 words [13].
- ***Linsear Write Formula:*** The Linsear Write Formula calculates a U.S. grade level, primarily for technical documents, by assigning different weights to "easy" and "hard" words within sentences [14].

• ***Dale-Chall Readability Score:*** This score determines a text's grade level based on its percentage of "difficult" words, which are defined as words not found on a specific list of 3,000 common words [15].

• ***Spache Readability Formula:*** The Spache Readability Formula is specifically designed to assess the reading level of texts for young children in early elementary grades 1-3 [16]

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

The readability of neurosurgical educational materials is a critical factor in ensuring effective patient education. A plethora of extant resources have been found to be incongruent with the recommended readability levels. This has the potential to impede patient comprehension and engagement. This issue assumes particular pertinence within the domain of neurosurgery, wherein the communication of intricate medical information necessitates effective accessibility. The integration of readability indices and advanced technologies, such as AI, has the potential to significantly enhance the quality of neurosurgical education materials.

Statistical Analysis

In our study, we analysed data from artificial intelligence models using SPSS version 27 (IBM Corp., USA). The data were classified 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 numerical data. Data conforming to a normal distribution were expressed as the mean ± standard deviation and a t-test was applied. Data that did not conform to a normal distribution were expressed as the median, minimum and maximum, and non-parametric tests were applied to them. Data with a p-value below 0.05 were considered significant.

Results

Our study presents a comparison of readability scores between ChatGPT and Gemini, utilizing various readability formulas. The results indicate that for most readability indices, there is no statistically significant difference between the two models, with one notable exception ARI (Automated Readability Index) Gemini: 10.05 ± 1.16 vs ChatGPT: 10.43 ± 1.81, p-Value: 0.327; Flesch Reading Ease: Gemini: 56.81 ± 7.13 vs. ChatGPT: 50.31 ± 14.04, p-Value: 0.114 Fog Scale (Gunning FOG Formula) Gemini: 0.76 ± 1.10 vs. ChatGPT: 10.95 ± 2.32, p-Value: 0.947; Flesch-Kincaid Grade Level Geini: 9.36 ± 1.16 vs. ChatGPT: 9.59 ± 1.68, p-Value: 0.718; Coleman-Liau Readability Index: Gemini: 10.59 ± 0.98 vs. ChatGPT: 11.80 ± 1.64; p-Value: 0.014 In summary, the results indicate that for most readability metrics, there is no significant difference in the readability of texts generated by Gemini and ChatGPT. The only exception is the Coleman-Liau Readability Index, where a statistically significant difference was found, with ChatGPT showing a slightly higher score, implying potentially greater complexity according to that specific measure.

Discussion

It is evident that the levels of readability are elevated. Research has indicated that a significant proportion of neurosurgical educational materials are written at a level that exceeds the recommended reading grade level for the general public [17]. For instance, the American Society of Neuroradiology's resources are written at an average grade level of 13.9, far above the range of 3 to 7 recommended by the National Institutes of Health and the American Medical Association [18]. Moreover, online patient education materials concerning neurointerventional procedures frequently exhibit a comprehension level commensurate with that of 10th to 11th-grade students, with some necessitating college-level cognitive abilities [19]. Complexity of Content: The complexity of neurosurgical procedures and conditions contributes to the high readability scores. For instance, the American Association of Neurological Surgeons (AANS) and other prominent organisations have materials that are frequently at a postgraduate readability level [20]. This inherent complexity can act as an impediment to the effective dissemination of patient education information, as many patients may lack the requisite literacy skills to comprehend the material in its entirety [21].

AI-Generated Summaries: The utilisation of artificial intelligence (AI), exemplified by ChatGPT, has demonstrated potential in the generation of more accessible summaries of neurosurgical literature. Research has indicated that AI-generated summaries exhibit considerably diminished readability scores in comparison to original abstracts, thereby rendering them more accessible to non-specialists. To illustrate this point, one may consider the finding that GPT-4-generated summaries have a Flesch-Kincaid reading grade of 7.70, in comparison to 12.55 for original abstracts [22]. This finding indicates that artificial intelligence (AI) has the potential to enhance the clarity and accessibility of neurosurgical educational materials [23]. The following indices of readability are employed: A variety of readability indices, including the Flesch-Kincaid Grade Level, Gunning Fog Index, and Simple Measure of Gobbledygook (SMOG) index, are employed to evaluate the readability of educational materials. These indices provide a quantitative measure of the reading difficulty, which can guide the development of more accessible content .The utilization of these indices by organisations facilitates the alignment of materials with recommended readability levels [24].

The readability indices frequently employed in the field are predicated on mathematical formulae that incorporate linguistic features such as word and sentence length. For instance, the Automated Readability Index (ARI) employs the calculation of average word length and sentence length to derive readability scores, which demonstrate a high degree of correlation with other indices [25]. The Flesch Reading Ease (FRE) scale is a widely utilized metric for assessing the readability of texts. Higher scores on this scale are indicative of texts that are more easily readable. The utilization of this approach has been demonstrated to be effective in a variety of contexts, including the design of prompts for oral proficiency assessments [26].

It is evident that certain indices, such as the one proposed by *Matricciani,* take into account cognitive aspects such as short-term memory capacity, an aspect which is often overlooked by traditional indices [27]. In the context of educational settings, readability indices have been shown to facilitate the alignment of written materials with students' reading abilities, thereby ensuring that the materials do not present an excessively facile or overly challenging level of

difficulty. This is of critical importance for the development of language and comprehension [28]. In the domain of health communication, the term "readability indices" is employed to denote the instruments utilized for the assessment of the clarity of online health-related information. For instance, during the course of the pandemic, it was demonstrated that a significant proportion of online resources were found to exceed the recommended sixth-grade reading level, which may have hindered public understanding [16-20]. Notwithstanding their practical applications, readability indices are subject to certain constraints. It is evident that these models frequently neglect to consider the semantic and contextual subtleties inherent within a text, a factor which has the potential to compromise comprehension [29]. It is important to note that the indices may not fully capture the complexity of texts in different languages or cultural contexts. To illustrate, a system developed for the purpose of Italian readability assessment offers a more intuitive understanding by comparing text difficulty to educational levels [11-14]. In summary, the results indicate that for most readability metrics, there is no significant difference in the readability of texts generated by Gemini and ChatGPT. The only exception is the Coleman-Liau Readability Index, where a statistically significant difference was found, with ChatGPT showing a slightly higher score, implying potentially greater complexity according to that specific measure. The provided document primarily focuses on presenting the comparative readability scores between ChatGPT and Gemini. It does not explicitly detail the limitations of the study. However, based on the information presented, several potential limitations can be inferred:

The study relies entirely on quantitative readability metrics. A qualitative assessment, involving human evaluators, could provide deeper insights into how understandable, engaging, or natural the texts generated by each model truly are, which might not be fully captured by formulaic scores. While the table indicates n=40 for both Gemini and ChatGPT, the paper does not specify how these samples were generated (e.g., what prompts were used, what topics were covered, or the length of the texts). A limited or homogenous sample of generated texts might not fully represent the models' capabilities across a wide range of applications or user queries. M Readability formulas, by their nature, are statistical approximations based on word and sentence characteristics (e.g., syllable count, word length, sentence length). They do not account for semantic complexity, context, or prior knowledge of the reader, which can significantly impact actual comprehension. Therefore, relying solely on these formulas might not provide a complete picture of true readability. The specific versions of ChatGPT and Gemini used for the study are not mentioned. Large language models are constantly evolving, and results obtained from older versions might not be representative of their current capabilities.

In conclusion, while the paper provides valuable quantitative data on the readability of texts from ChatGPT and Gemini, its scope is narrow. A more comprehensive study would ideally include qualitative assessments, a broader range of text types, and detailed information on the methodology and model versions to provide a more holistic understanding of the models' performance.

Table 1.

Comparison of CHATGPT vs GEMINI in terms of readability scores.

	GEMINI	CHAT –GPT	p-Value
ARI	10.05±1.16	10.43±1.81	0.327
Flesch Reading Ease	56.81±7.13	50.31±14.04	0.114
Fog Scale (Gunning FOG Formula)	10.76 ±1.10	10.95±2.32	0.947
Flesch-Kincaid Grade Level	9.36±1.16	9.59±1.68	0.718
Coleman-Liau Readability Index	10.59±0.98	11.80± 1.64	**0.014**
The SMOG Index	9.11±0.99	8.81±1.63	0.547
Linsear Write Formula	71.68±3.32	7016±3.36	0.446
Dale-Chall Readability Score	8.28±0.58	8.24±0.84	0.925
Spache Readability Formula	6.37±0.46	6.37±0.49	0.738

References:

1. *Muehlschlegel S.* Subarachnoid Hemorrhage. Continuum (Minneap Minn). 2018 Dec;24(6):1623-1657. doi: 10.1212/CON.0000000000000679. PMID: 30516599.

2. *Claassen J., Park S.* Spontaneous subarachnoid haemorrhage. Lancet. 2022 Sep 10;400(10355):846-862. doi: 10.1016/S0140-6736(22)00938-2. Epub 2022 Aug 16. PMID: 35985353; PMCID: PMC9987649.

3. *Abraham M.K., Chang W.W.* Subarachnoid Hemorrhage. Emerg Med Clin North Am. 2016 Nov;34(4):901-916. doi: 10.1016/j.emc.2016.06.011. PMID: 27741994.

4. *Zhou M., Pan Y., Zhang Y., et al.* Evaluating AI-generated patient education materials for spinal surgeries: Comparative analysis of readability and DISCERN quality across ChatGPT and deepseek models. Int J Med Inform. 2025 Jun;198:105871. doi: 10.1016/j.ijmedinf.2025.105871. Epub 2025 Mar 13. PMID: 40107040.

5. *Büker M., Mercan G.* Readability, accuracy and appropriateness and quality of AI chatbot responses as a patient information source on root canal retreatment: A comparative assessment. Int J Med Inform. 2025 Sep;201:105948. doi: 10.1016/j.ijmedinf.2025.105948. Epub 2025 Apr 25. PMID: 40288015.

6. *Zaleski A.L., Berkowsky R., Craig KJT, Pescatello LS.* Comprehensiveness, Accuracy, and Readability of Exercise Recommendations Provided by an AI-Based Chatbot: Mixed Methods Study. JMIR Med Educ. 2024 Jan 11;10:e51308. doi: 10.2196/51308. PMID: 38206661; PMCID: PMC10811574.

7. *Yalla G.R., Hyman N., Hock L.E., Zhang Q., Shukla A.G., Kolomeyer N.N.* Performance of Artificial Intelligence Chatbots on Glaucoma Questions Adapted From Patient Brochures. Cureus. 2024 Mar 23;16(3):e56766. doi: 10.7759/cureus.56766. PMID: 38650824; PMCID: PMC11034394.

8. *Gibson D., Jackson S., Shanmugasundaram R., Seth I., Siu A., et al.* Evaluating the Efficacy of ChatGPT as a Patient Education Tool in Prostate Cancer: Multimetric Assessment. J Med Internet Res. 2024 Aug 14;26:e55939. doi: 10.2196/55939.

9. *Wong K., Levi J.R.* Partial Tonsillectomy. Ann Otol Rhinol Laryngol. 2017 Mar;126(3):192-198. doi: 10.1177/0003489416681583.

10. *Zheng J., Yu H.* Readability Formulas and User Perceptions of Electronic Health Records Difficulty: A Corpus Study. J Med Internet Res. 2017 Mar 2;19(3):e59. doi: 10.2196/jmir.6962.

11. *White A., Danis M.* Enhancing patient-centered communication and collaboration by using the electronic health record in the examination room. J Am Med Assoc. 2013 Jun 12;309(22):2327–8. doi: 10.1001/jama.2013.6030.

12. *Delbanco T., Walker J., Bell S.K., Darer J.D., Elmore J.G., Farag N., Feldman H.J., et al.* Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med. 2012 Oct 2;157(7):461–70. doi: 10.7326/0003-4819-157-7-201210020-00002.

13. *Wiljer D., Bogomilsky S., Catton P., Murray C., Stewart J., Minden M.* Getting results for hematology patients through access to the electronic health record. Can Oncol Nurs J. 2006;16(3):154–64. doi: 10.5737/1181912x163154158.

14. *Tang P.C., Lansky D.* The missing link: bridging the patient-provider health information gap. Health Aff (Millwood) 2005;24(5):1290–5. doi: 10.1377/hlthaff.24.5.1290.

15. *Keselman A., Slaughter L., Smith C.A., Kim H., Divita G., Browne A., Tsai C., Zeng-Treitler Q.* Towards consumer-friendly PHRs: patients' experience with reviewing their health records. AMIA Annu Symp Proc. 2007:399–403.
http://europepmc.org/abstract/MED/18693866 .

16. .*Keselman A., Smith C.A.* A classification of errors in lay comprehension of medical documents. J Biomed Inform. 2012 Dec;45(6):1151–63. doi: 10.1016/j.jbi.2012.07.012.

17. *Pyper C., Amery J., Watson M., Crook C.* Patients' experiences when accessing their on-line electronic patient records in primary care. Br J Gen Pract. 2004 Jan;54(498):38–43.

18. *Ader E.R., Allam M.M., Harris T., Suchdev N., Loke, Y.K., & Barlas R.S.* (2025). Thrombolysis for aneurysmal subarachnoid haemorrhage. The Cochrane Library. https://doi.org/10.1002/14651858.cd013748.pub2

19. *Fukuda H., Hyohdoh Y., et al.* (2025). Risk factors of short-term poor functional outcomes and long-term durability of ruptured large or giant intracranial aneurysms. Journal of Neurosurgery. https://doi.org/10.3171/2024.8.jns24894

20. *Yang BSK, Blackburn S.L., Lorenzi P.L., Choi H.A., Gusdon A.M.* Metabolomic and lipidomic pathways in aneurysmal subarachnoid hemorrhage. Neurotherapeutics. 2025 Jan;22(1):e00504. doi: 10.1016/j.neurot.2024.e00504. Epub 2024 Dec 19. PMID: 39701893; PMCID: PMC11840353.

21. *Irwin S.C., Lennon D.T., Stanley C.P., Sheridan G.A., Walsh J.C.* Ankle conFUSION: The quality and readability of information on the internet relating to ankle arthrodesis. Surgeon. 2021 Dec;19(6):e507-e511. doi: 10.1016/j.surge.2020.12.001. Epub 2021 Jan 13. PMID: 33451875.

22. *Odeh M., Oqal M., AlDroubi H., Al-Omari B.* Assessing the competency of pharmacists in writing effective curriculum vitae for job applications: a cross-sectional study and readability index evaluation. BMC Med Educ. 2023 Nov 20;23(1):884. doi: 10.1186/s12909-023-04870-5. PMID: 37985997; PMCID: PMC10662548.

23. *Shah R., Mahajan J., Oydanich M., Khouri A.S.* A Comprehensive Evaluation of the Quality, Readability, and Technical Quality of Online Information on Glaucoma. Ophthalmol Glaucoma. 2023 Jan-Feb;6(1):93-99. doi: 10.1016/j.ogla.2022.07.007. Epub 2022 Aug 6. PMID: 35940574.

24. *Guerra G.A., Grove S., Le J., Hofmann H.L., Shah I., Bhagavatula S., et al.* Artificial intelligence as a modality to enhance the readability of neurosurgical literature for patients. J Neurosurg. 2024 Nov 8;142(4):1189-1195. doi: 10.3171/2024.6.JNS24617. PMID: 39504543.

25. *Hershenhouse J.S., Mokhtar D., Eppler M.B., Rodler S., Storino Ramacciotti L., et al.* Accuracy, readability, and understandability of large language models for prostate cancer information to the public. Prostate Cancer Prostatic Dis. 2025 Jun;28(2):394-399. doi: 10.1038/s41391-024-00826-y. Epub 2024 May 14. PMID: 38744934; PMCID: PMC12106072.

26. *Villani V., Nguyen H.T., Shanmugarajah K.* Evaluating Quality and Readability of AI-generated Information on Living Kidney Donation. Transplant Direct. 2024 Dec 10;11(1):e1740. doi: 10.1097/TXD.0000000000001740. PMID: 39668891; PMCID: PMC11634323.

27. *Papuc M., Scheffler P.* The impact of internet resources and artificial intelligence on information on myringotomy tubes. Eur Arch Otorhinolaryngol. 2025 Apr;282(4):2149-2153. doi: 10.1007/s00405-024-09148-0. Epub 2024 Dec 12. PMID: 39668220.

28. *Yang, S.K., Blackburn S., Lorenzi P.L., Choi H.A., & Gusdon, A.M.* (2024). Metabolomic and lipidomic pathways in aneurysmal subarachnoid hemorrhage. Neurotherapeutics.
https://doi.org/10.1016/j.neurot.2024.e00504.

29. *Mehryar S., Yazdanpanah V., Tong J.* AI and climate resilience governance. iScience. 2024 Apr 26;27(6):109812. doi: 10.1016/j.isci.2024.109812. PMID: 38784017; PMCID: PMC11112607.

Corresponding Author:
Bedel Cihan. - MD, Health Science University Antalya Training and Research Hospital, Kazım Karabekir.
Address: 07100, Muratpaşa, Antalya, Turkey.
E-mail: cihanbedel@hotmail.com
Phone: +905075641254,
Fax: +902422494487