RESEARCH



GPT meets PubMed: a novel approach to literature review using a large language model to crowdsource migraine medication reviews

Elyse Mackenzie¹, Roger Cheng¹ and Pengfei Zhang^{2,3*}

Abstract

Objective To evaluate the potential of two large language models (LLMs), GPT-4 (OpenAI) and PaLM2 (Google), in automating migraine literature analysis by conducting sentiment analysis of migraine medications in clinical trial abstracts.

Background Migraine affects over one billion individuals worldwide, significantly impacting their quality of life. A vast amount of scientific literature on novel migraine therapeutics continues to emerge, but an efficient method by which to perform ongoing analysis and integration of this information poses a challenge.

Methods "Sentiment analysis" is a data science technique used to ascertain whether a text has positive, negative, or neutral emotional tone. Migraine medication names were extracted from lists of licensed biological products from the FDA, and relevant abstracts were identified using the MeSH term "migraine disorders" on PubMed and filtered for clinical trials. Standardized prompts were provided to the APIs of both GPT-4 and PaLM2 to request an article sentiment as to the efficacy of each medication found in the abstract text. The resulting sentiment outputs were classified using both a binary and a distribution-based model to determine the efficacy of a given medication.

Results In both the binary and distribution-based models, the most favorable migraine medications identified by GPT-4 and PaLM2 aligned with evidence-based guidelines for migraine treatment.

Conclusions LLMs have potential as complementary tools in migraine literature analysis. Despite some inconsistencies in output and methodological limitations, the results highlight the utility of LLMs in enhancing the efficiency of literature review through sentiment analysis.

Keywords Headaches, Large language model, Artificial intelligence, Literature review, Migraine

*Correspondence:

Pengfei Zhang

phil.p.zhang@gmail.com

¹ Department of Neurology, Rutgers Robert Wood Johnson Medical

School, New Jersey, USA

² Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA, USA

³ Harvard Medical School, Boston, MA, USA

Background

Over one billion individuals worldwide suffer from migraine, with the disruptive and often debilitating symptoms considerably impacting their quality of life [1]. Despite the advent of novel treatments over the past decade, efficient analysis of this expanding body of scientific literature remains a challenge [2]. However, large language models (LLMs) like OpenAI's GPT family (popularly known by their interface *ChatGPT*) and Google's



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Gemini (formerly *Bard*), have the potential to revolutionize literature analysis by automating the synthesis and summary of research findings [3, 4]. Broadly speaking, LLMs are predictive models that generate human-like text in response to a prompt based on prior training on vast datasets of existing text. This gives them an unique ability to simulate the way a human might read, analyze, and interpret an article, possibly enabling insights into topics like migraine therapeutics, but much more rapidly and at larger scale [5–7].

Sentiment analysis, a computational process that discerns and quantifies the subjective tone of text, can offer valuable insights to scientific literature review by identifying predominant attitudes and perspectives about a topic [8]. This is done by identifying keywords and phrases that express positive or negative sentiment then quantifying the sentiment. For example, phrases like "successful" or "curative" would be associated with positive sentiment, while phrases "unsuccessful" or "harmful" may be associated with negative sentiment. With recent advances in computational power, machine learning tools can further enhance the efficiency of sentiment analysis by processing and analyzing vast amounts of text data at scale. Myszewski et al. successfully applied sentiment analysis to literature review using a novel sentiment classification model for clinical trial abstracts. Their model, built on adversarial learning and the BioBERT language processing model (an early example of an LLM trained on biomedical text), achieved a 91.3% accuracy in sentiment classification compared to assessments by expert human raters. [9] While GPT and Gemini are general purpose LLMs, they are several generations advanced and vastly more capable. This pilot study, employing both binary and distribution-based models, aims to assess the ability of two LLMs, specifically the GPT-4 model of GPT and the PaLM2 model of Gemini, to identify migraine medications with the most positive sentiment from PubMed clinical trial abstracts. Successful analysis could suggest broader applications of LLMs in highlighting promising therapies in headache medicine.

Methods

Comprehensive lists of pharmacologic and biologic medications were extracted from the FDA's "Orange Book" and "Purple Book," respectively, including all brand and generic medication names [10, 11]. The two resultant lists were combined, duplicates were removed, and words were further screened through a publicly available database of 466,550 English words [12]. Medication combinations were considered separately from their individual counterparts (i.e., sumatriptan/naproxen was considered a separate compound from sumatriptan or naproxen alone). To identify relevant article abstracts associated with clinical trials for migraine disorders, we conducted a PubMed search using the MeSH (Medical Subject Headings) term "migraine disorders[mh]." MeSH search was utilized over keyword search due to its more structured and consistent approach to finding relevant literature. These results were then filtered to only include clinical trials. Once the PMIDs for these articles were identified, available article abstracts were downloaded using Pub-Med's application programming interface (API).

Binary, cumulative summation model

The text from each extracted abstract was entered into the corresponding APIs for both GPT and Gemini with the following prompt: *Read the following abstract and identify all of the medications. For each medication, determine whether it is effective, ineffective, or neutral for the treatment of migraine. Assign a value of 1 for an effective medication, -1 for an ineffective medication, and 0 for a neutral medication. Output the result in the format* \"(drug, value)\". If no drug is found, output the word \"none\". Do not output any explanations. Here is the *abstract:*"

While this process was scripted and performed via the API, we have provided the equivalent process as screenshots as if it were done manually (Figs. 1 and 2).

Despite instructions in the prompt, as can be seen, the responses from these processes contained formatting variations. Therefore, outputs were parsed via custom-designed (non-AI/ML) code. Exceptions and nonconforming responses were identified by this code and manually corrected.

After parsing was completed, unique pharmacological/biological agents listed in the database were identified and paired with the number of associated "1"s, "0"s, or "-1"s. A cumulative summation of scores (1, 0, -1) was obtained for each pharmacological/biological agent. Results were ranked from highest to lowest score. A mean score was also calculated to identify agents that may have had a bias towards high scores simply due to frequency of study. A final manual screening was conducted to remove any remaining nonsensical words or non-drug related entries.

Distribution-based (non-binary) model

A higher score on the binary/cumulative model maybe influenced by the number of publications of a medication rather than its efficacy. To ameliorate this, we constructed a distribution-based model taking inspiration from Guo et al.'s sentiment analysis model for social media platforms. Here, we aimed to measure the mean, median, and standard deviation of sentiment analysis with output as a *range* from -1 and 1 rather than

Read the following abstract and identify all of the medications. For each medication, determine whether it is effective, ineffective, or neutral for the treatment of migraine. Assign a value of 1 for an effective medication, -1 for an ineffective medication, and 0 for a neutral medication. Output the result in the format \"(drug, value)\". If no drug is found, output the word \"none\". Do not output any explanations. Here is the abstract: B C 6 7 Based on the provided abstract, the identified medication for the treatment of acute migraine is: 1. Diphenhydramine (when administered as adjuvant therapy with metoclopramide) Now, let's determine its effectiveness based on the given information: Diphenhydramine: Ineffective (Value: -1) The study found that intravenous diphenhydramine, when administered as adjuvant therapy with metoclopramide, does not result in greater rates of sustained headache relief compared to placebo+metoclopramide. Output: scss Copy code (Diphenhydramine, -1)

Fig. 1 GPT screenshot (PMID: 2,632,052)

a binary positive/negative score [13]. The following prompt was provided to Gemini and GPT:

"Read the following abstract and identify all of the medications. For each medication, determine whether it is effective, ineffective, or neutral for the treatment of migraine. Assign a value between -1to 1 where 1 is for an effective medication, -1 for an ineffective medication, and 0 for a neutral medication. Output the result in the format \"(drug, value)\" If no drug is found, output the word \"none\". Do not output any explanations. Here is the abstract:" Like the binary approach, results were manually refined to correct for output errors. Recognizing that each medication's average sentiment score could be skewed by the number of studies it appeared in, medications mentioned in only one study were excluded. Instead of a single cumulative score, we aimed to derive a sentiment distribution for each medication based on the mean, median, standard deviation, and citation frequency. Analysis was prioritized by focusing on medications mentioned in the highest number of PubMed articles, selecting those with articles greater or equal to the median article count. This approach aimed to highlight the most extensively researched medications. These medications were then ranked by average sentiment score. Mean scores above Read the following abstract and identify all of the medications. For each medication, determine whether it is effective, ineffective, or neutral for the treatment of migraine. Assign a value of 1 for an effective medication, -1 for an ineffective medication, and 0 for a neutral medication. Output the result in the format \"(drug, value)\". If no drug is found, output the word \"none\". Do not output any explanations. Here is the abstract:

Sure, here is the output of the analysis:

(diphenhydramine, -1)
(metoclopramide, 0)

The medications identified in the abstract are diphenhydramine and metoclopramide. The study found that diphenhydramine, when administered as adjuvant therapy with metoclopramide, does not improve migraine outcomes. Therefore, the value for diphenhydramine is -1, indicating that it is ineffective for the treatment of migraine. Metoclopramide was not found to be effective or ineffective, so the value for metoclopramide is 0.

Here is the explanation for the values:

- -1: Ineffective medication. The study found that the medication did not improve migraine outcomes.
- 0: Neutral medication. The study did not find that the medication improved or worsened migraine outcomes.
- 1: Effective medication. The study found that the medication improved migraine outcomes.

I hope this is helpful! Let me know if you have any other questions.

Fig. 2 Gemini results

0.5 represented medications that were both widely studied and positively regarded based on sentiment analysis.

Implementation

Custom Python code was written to download PMIDs as well as their corresponding abstracts. The Python Optical Character Recognition (OCR) library was used to convert the FDA Orange and Purple Book PDF format to text when required for extractions. GPT and Gemini APIs were accessed through Python libraries. Custom code written in the Haskell programming language was used to parse and clean our database in various stages of the project.

Results

Data identification and abstract acquisition for 2700 articles was completed on January 8, 2023. For the binary part of the project, Gemini API evaluation was completed on July 21, 2023 and GPT evaluations were completed on July 19th, 2023 and August 15th, 2023. Non-binary evaluations were completed on Gemini and GPT APIs on September 29, 2023 and October 5th, 2023, respectively.

Binary results

After excluding nonsensical entries, the ten most favorable pharmaceutical/biological agents as determined by Gemini were sumatriptan, topiramate, rizatriptan, almotriptan, erenumab, zolmitriptan, galcanezumab, frovatriptan, fremanezumab, and lasmiditan (Table 1). For GPT, the ten most favorable pharmaceutical/biological agents were sumatriptan, topiramate, rizatriptan, zolmitriptan, erenumab, galcanezumab, almotriptan, metoclopramide, frovatriptan, and fremanezumab (Table 1). Of note, all of these medications, with the exception of metoclopramide, are FDA approved for migraine.

Distribution-based results

In the Gemini dataset, a total of 71 drugs were identified after manual verification. The median number of articles for each medication was six, therefore medications with more than five PubMed articles were included (n=41). Among this list, 33 medications contained mean sentiment scores greater than 0.5 (Table 2) The ten most favorable medications by mean score were fremanezumab, eptinezumab, ubrogepant, rimegepant, zonisamide, erenumab,

Table 1 B	inary results by	summation of	f sentiment (1	, 0, -1)
-----------	------------------	--------------	----------------	----------

,			
Medication	Cumulative score	Total articles	Mean score
A: Gemini Rankings (Top 20)		
sumatriptan	169.5 ^a	218	0.778
topiramate	73	93	0.785
rizatriptan	63.5 ^a	68	0.934
erenumab	47	48	0.979
almotriptan	47	48	0.979
zolmitriptan	46	51	0.902
galcanezumab	44	45	0.978
frovatriptan	31	32	0.969
fremanezumab	31	31	1.000
lasmiditan	29	30	0.967
onabotulinum- toxina	28	30	0.933
eptinezumab	28	29	0.966
sodium-valproate	21	25	0.840
eletriptan	20	24	0.833
naratriptan	16	19	0.842
ubrogepant	15	15	1.000
naproxen-sodium	15	15	1.000
metoprolol	13	20	0.650
prochlorperazine	12	13	0.923
metoclopramide	12	22	0.545
B: GPT Rankings (Top	20)		
sumatriptan	224	300	0.747
topiramate	84	115	0.730
rizatriptan	78	84	0.929
zolmitriptan	58	70	0.829
erenumab	49	51	0.961
galcanezumab	48	50	0.960
almotriptan	47	51	0.922
metoclopramide	38	66	0.576
frovatriptan	35	35	1.000
fremanezumab	32	32	1.000
onabotulinum- toxina	30	34	0.882
eptinezumab	29	30	0.967
sodium-valproate	28	33	0.848
lasmiditan	28	29	0.966
eletriptan	28	31	0.903
prochlorperazine	20	22	0.909
metoprolol	18	24	0.750
naratriptan	17	19	0.895
naproxen-sodium	17	20	0.850
ubrogepant	13	15	0.867
- ·			

 $^{\rm a}$ Score results of 0.5 occurred due to the LLM not following directions to limit output to -1,0, or 1

galcanezumab, bupivacaine, and levetiracetam. Of note, all of these medications, with the exception of bupivacaine, zonisamide, and levetiracetam are FDA approved for migraine.

Table 2	Distribution-based results by mean sentiment (1,	0, -1)

Medication	Mean sentiment	St.Dev	Total articles
A: Gemini Rankings			
fremanezumab	1.000	0.000	30
eptinezumab	1.000	0.000	29
ubrogepant	1.000	0.000	15
rimegepant	1.000	0.000	7
zonisamide	1.000	0.000	6
erenumab	0.964	0.161	49
galcanezumab	0.957	0.206	46
bupivacaine	0.929	0.267	14
levetiracetam	0.929	0.189	7
naproxensumatriptan	0.917	0.257	18
fluoxetine	0.917	0.204	6
timolol	0.917	0.204	6
almotriptan	0.899	0.356	47
divalproex	0.889	0.333	9
lasmiditan	0.883	0.284	30
venlafaxine	0.875	0.354	8
diclofenac	0.867	0.297	15
rizatriptan	0.857	0.343	77
zolmitriptan	0.845	0.387	63
atogepant	0.833	0.408	6
prochlorperazine	0.826	0 355	16
eletrintan	0.818	0.442	22
onabotulinumtoxina	0.813	0.392	32
naproven	0.800	0.352	25
frovatrintan	0.791	0.346	35
topiramate	0.783	0.210	Q1
sumatrintan	0.773	0.109	242
metoprolol	0.773	0.000	272
valoroate	0.726	0.510	30
naratriotan	0.720	0.513	22
ashapontin	0.723	0.488	7
kataprofen	0.714	0.400	/
mataclapramida	0.052	0.010	10
Record Readings	0.561	0.556	20
from an or umab	1.000	0.000	22
nancovancumatriatan	1.000	0.000	52
rimensumatriptan	1.000	0.000	21
nmegepant	1.000	0.000	7
alogepant	1.000	0.000	0
gaicanezumap	0.980	0.141	50
eptinezumab	0.967	0.183	30
frequetriater	0.962	0.194	52
frovatriptan	0.958	0.184	36
iasmiditan	0.953	0.195	3Z
aiciotenac	0.944	0.236	18
ubrogepant	0.938	0.250	16
levetiracetam	0.929	0.189	/
rızatriptan	0.919	0.307	93
almotriptan	0.907	0.293	54

Medication	Mean sentiment	St.Dev	Total articles
eletriptan	0.889	0.398	36
bupivacaine	0.875	0.342	16
timolol	0.875	0.354	8
venlafaxine	0.875	0.354	8
zolmitriptan	0.867	0.373	79
onabotulinumtoxina	0.857	0.430	35
gabapentin	0.857	0.378	7
valproic	0.857	0.378	7
nadolol	0.833	0.408	6
prochlorperazine	0.820	0.379	25
metoprolol	0.813	0.385	24
naproxen	0.809	0.390	34
valproate	0.800	0.524	55
naratriptan	0.761	0.601	23
sumatriptan	0.758	0.556	325
ketoprofen	0.750	0.577	16
fluoxetine	0.750	0.463	8
topiramate	0.737	0.576	118
clopidogrellopid	0.714	0.488	7
zonisamide	0.714	0.756	7
ketorolac	0.700	0.414	15
divalproex	0.643	0.745	14
verapamil	0.583	0.669	12
metoclopramide	0.528	0.573	71

For GPT, a total of 90 drugs were identified after manual verification. Similarly to Gemini, the median number of articles for each medication was six, therefore medications with more than five PubMed articles were included (n=46). Among this list, 38 medications contained mean sentiment scores greater than 0.5 (Table 2) The ten most favorable medications by mean score were fremanezumab, naproxen/sumatriptan, rimegepant, atogepant, galcanezumab, eptinezumab, erenumab, frovatriptan, and lasmiditan. Of note, all of these medications are FDA approved for migraine.

Post-hoc analysis and manual scoring

As part of a post-hoc analysis, we manually scored a randomly selected sample of 100 abstracts and compared it to both the GPT and Gemini results.

In the binary model for GPT, 61 abstracts matched manual scoring, 21 were scored differently, and 18 contained non-significant variations (detailed later). In the binary model for Gemini, 53 abstracts matched manual scoring, 37 were scored differently, and 10 had non-significant variations.

The non-significant mismatches were semantic in nature. For instance, algorithm output for PMID 12230594

was (MIG-99", 1), ("Tanacetum parthenium", 1), ("feverfew", 1) whereas the human scorer denoted the article as "feverfew, 1". Both are correct, since feverfew is known as MIG-99 commercially, and Tanacetum parthnium scientifically. In PMID 19846269, the human scored (steroids, -1), whereas the algorithm was more precise in scoring ("dexamethasone", -1), ("prednisone", -1). Finally, a source of non-significant mismatch comes from inclusion of a nonpharmacological intervention in the output rather than writing "none" as the prompt suggests.

In the distributive model, unlike in the binary model, any deviation beyond 0.1 was considered a significant mismatch. For example, for PMID 17988947, (topiramate, 1) versus (topiramate, 0.76) were considered significantly different. For GPT, 58 abstracts matched manual scoring, 25 mismatched, and 17 resulted in nonsignificant variations. For Gemini, 52 abstracts matched manual scoring, 37 mismatched, and 11 had non-significant variations.

In summary, when considering a human as the gold standard, outputs differed significantly from GPT in 21% of articles using the binary method and 25% using the distributive method. Compared to Gemini, there was a difference of 37% using both the binary method as well as the distributive method.

Discussion

To our knowledge, our study is the first to investigate the application of large language models to literature search and migraine. Across binary and non-binary methods, medications receiving the most "positive" sentiment align with evidence-based choices for abortive and preventative migraine treatment as presented in the AHS Guide-lines [14]. For example, triptans, topiramate, and CGRP monoclonal antibody treatments consistently ranked among the top ten medications chosen by both GPT and Gemini. These findings suggest the potential of large language models as complementary tools in real-time identification of favorable migraine medications in primary literature analysis.

Discrepancies between binary and distribution-based models stem from their distinct methodologies. The binary model uses discrete scores (-1, 0, 1), favoring simplicity but overemphasizing frequently studied medications. The distribution-based model uses a range of scores (-1 to 1) summarized by mean and median, which mitigates the bias of publication frequency by emphasizing sentiment trends. For instance, sumatriptan's lower mean scores may reflect its use as a comparator in newer drug studies. These differences highlight the need to select models aligned with specific research goals. Combining both approaches could offer a more balanced sentiment analysis.

The study encountered several model-related and methodological limitations. First, both GPT and Gemini have well-documented drawbacks, including improper categorization of medications, production of falsified emerging migraine therapies, and inaccurate source citation [5, 15]. Further, despite providing explicit and uniform prompts, both LLMs displayed inconsistencies in adherence. For example, in the binary model, Gemini's assignment of a 0.5 score for sumatriptan in PMID 8783475 deviated significantly from our prompt. These inconsistencies required the development of custom code to parse the LLM outputs and manually correct formatting variations. There were also significant differences (between 21-37%) between human and machine scoring, though most articles were rated concordantly by both. Finally, while the top ten most positively rated medications by GPT and Gemini generally aligned with evidence-based guidelines, their outputs were not identical, reflecting the inconsistent intra- and interplatform reproducibility of LLMs. Further complicating the issue of consistency, the way LLMs generate output does not guarantee the same result with each execution, so a system may not agree with its own conclusion from a different run.

Future research should address the limitations identified in this pilot study. For example, while we could request rationales for sentiment scores, for this pilot study we have decided that such reliance on validation for each outcome would undermine the purposed autonomy of LLM analysis. In future studies, however, it may be helpful to ask LLM to describe and justify its claim of effectiveness as an output. Furthermore, not all clinical trials are the same – multicenter trials have higher level of evidence than single center studies, and a potential future direction is to incorporate the level of evidence for clinical trials and adjust the sentiment scores accordingly. A more rigorous method is also needed to assess sentiment data from large language models at scale. Our initial summative method using scores from -1 to 1 might be overly simplistic, while our second approach excludes smaller datasets, potentially leading to less reliable and biased evaluations of the current migraine medication landscape.

It is interesting to note that in the distributional model as well as per the mean score in the binary model, sumatriptan and topiramate, the two most evidencebased and therefore most studied medications are not the ones with the highest sentiment. We hypothesize that this discordance has less to do with our model than with the nature of clinical trials where the question of "evidence based" (i.e. volume of articles) is pitted against the question of "effectiveness/tolerability" (i.e. sentiment). Indeed, "evidence-based" does not need to be equivalent to "effective"—because sumatriptan and topiramate remain the most evidence-based medications for abortive/prevention in migraine, they are used most frequently in comparison trials to establish efficacy for newer medications, often with the explicit intention of proving superiority for the newer medication either in efficacy or tolerability. Since this sort of method is done at scale for the goal of proving superiority of newer drug over older more canonical drugs, it is inevitable that the older drug will have higher "volume" of articles, but some of those may cast them in an inferior light over newer drugs. In other words, while the volume of articles which contains these medications are high, newer classes of medications with fewer studies may have more skew towards positive sentiment.

Despite these limitations, the ability of both GPT and Gemini to consistently identify medications that align with established clinical guidelines underscores both the use case for sentiment analysis in medication research and the potential utility of LLMs to aid in primary literature review.

The clinical utility of sentiment analysis is derived from its ability to synthesize and prioritize findings from extensive medical literature. For clinicians, these results can serve as an initial filter to identify promising therapies warranting further investigation. For example, medications such as fremanezumab and galcanezumab, which consistently scored highly in our analysis, align with evidence-based guidelines for migraine treatment and may guide decision-making for patients requiring targeted therapies. For researchers, sentiment trends can help identify gaps in the literature or assess the broader reception of therapeutic innovations. Positive sentiment scores for medications, as observed in this study, may reflect their demonstrated efficacy and tolerability in clinical trials, but could also be influenced by publication bias or authors' writing styles favoring newer treatments. However, as sentiment analysis mimics the way in which humans might interpret a piece of text, without conscious bias mitigation, a human reader may be influenced in the same manner. As such, while sentiment analysis does not replace traditional literature review methods, but its application offers a complementary tool for streamlining decision-making processes in both clinical and research settings.

Overall, we believe that the continued refinement and integration of LLMs into clinical practice may play a promising supporting role in for healthcare professionals navigating and synthesizing vast amounts of literature. In turn, this could contribute to enhanced decision-making in medication management. As machine learning inevitably advances, understanding its capabilities and appropriate applications may lead the way to breakthroughs in headache medicine research and beyond.

Glossary

API	Application Programming Interface
PMID	PubMed Identifier
LLM	Large language model
OCR	Python Optical Character Recognition
GPT	Generative Pre-trained Transformer
ICHD3	International classification of headache disorders, 3rd edition
PaLM	Pathways Language Model

Acknowledgements

We would like to acknowledge Cymbeline LLC for providing a portion of the research resources and Rutgers University for funding of OpenAPI fees and article processing fees.

Authors' contributions

PZ, RC, and EM contributed to design of the study. PZ implemented the study. All authors contributed in drafting, editing, and approving the final manuscript.

Funding

Cymbeline LLC provided a portion of the research resources. Rutgers University provided funding of OpenAPI fees and article processing fees.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate $\ensuremath{\mathsf{N/A}}\xspace.$

Consent for publication

N/A.

Competing interests

PZ: He has received honorarium from Alder Biopharmaceuticals, Board Vitals, and Fieve Clinical Research. He collaborates with Headache Science Incorporated without receiving financial support. He had ownership interest in Cymbeline LLC. He is a consultant for Acument LLC.

Received: 26 August 2024 Accepted: 5 February 2025 Published online: 19 February 2025

References

- Amiri P, Kazeminasab S, Nejadghaderi SA, et al. Migraine: A Review on Its History, Global Epidemiology, Risk Factors, and Comorbidities. Front Neurol. 2022;12: 800605. https://doi.org/10.3389/fneur.2021.800605.
- Do TP, Guo S, Ashina M. Therapeutic novelties in migraine: new drugs, new hope? J Headache Pain. 2019;20(1):37. https://doi.org/10.1186/ s10194-019-0974-3.
- 3. Try Bard, an Al experiment by Google. https://bard.google.com. Accessed 28 July 2023.
- Introducing ChatGPT. https://openai.com/blog/chatgpt. Accessed 28 July 2023.
- Peng KP, May A. Crossing the Rubicon? The future impact of artificial intelligence on headache medicine. Cephalalgia. 2023;43(4):033310242311573. https://doi.org/10.1177/03331024231157379.
- Cohen F. The role of artificial intelligence in headache medicine: Potential and peril. Headache. 2023;63(5):694–6. https://doi.org/10.1111/head.14495.
- Romano MF, Shih LC, Paschalidis IC, Au R, Kolachalama VB. Large Language Models in Neurology Research and Future Practice. Neurology. 2023;101(23):1058–67. https://doi.org/10.1212/WNL.000000000207967.
- Zunic A, Corcoran P, Spasic I. Sentiment Analysis in Health and Well-Being: Systematic Review. JMIR Med Inform. 2020;8(1):e16023. https://doi.org/ 10.2196/16023.

- Myszewski JJ, Klossowski E, Meyer P, Bevil K, Klesius L, Schroeder KM. Validating GAN-BioBERT: A Methodology for Assessing Reporting Trends in Clinical Trials. Front Digit Health. 2022;4. https://doi.org/10.3389/fdgth. 2022.878369. Accessed 29 July 2023.
- Approved Drug Products with Therapeutic Equivalence Evaluations | Orange Book. FDA. Published online January 12, 2024. https://www.fda. gov/drugs/drug-approvals-and-databases/approved-drug-productstherapeutic-equivalence-evaluations-orange-book. Accessed 7 Feb 2024.
- Purple Book: Lists of Licensed Biological Products with Reference Product Exclusivity and Biosimilarity or Interchangeability Evaluations. FDA. Published online August 3, 2020. https://www.fda.gov/drugs/therapeutic-biolo gics-applications-bla/purple-book-lists-licensed-biological-products-refer ence-product-exclusivity-and-biosimilarity-or. Accessed 7 Feb 2024.
- MohamedBechirMejri/enwords. GitHub. https://github.com/Moham edBechirMejri/enwords/blob/main/src/words.json. Accessed 7 Feb 2024.
- Guo Y, Rajwal S, Lakamana S, et al. Generalizable Natural Language Processing Framework for Migraine Reporting from Social Media. AMIA Jt Summits Transl Sci Proc. 2023;2023:261–70.
- 14. Ailani J, Burch RC, Robbins MS, the Board of Directors of the American Headache Society. The American Headache Society Consensus Statement Update on integrating new migraine treatments into clinical practice. Headache. 2021;61(7):1021–39. https://doi.org/10.1111/head.14153.
- King MR. Can Bard, Google's Experimental Chatbot Based on the LaMDA Large Language Model, Help to Analyze the Gender and Racial Diversity of Authors in Your Cited Scientific References? Cel Mol Bioeng. 2023;16(2):175–9. https://doi.org/10.1007/s12195-023-00761-3.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.