

Research Article

Patient Experiences in the Cochlear Implant Reddit Community: Comparing Human and Large Language Model Categorization

Daniel R. S. Habib,^a  Kiran Depala,^a Jack Lin,^a Samuel Le,^a Natalie McFall,^a Shiv S. Dewan,^a Justin Huang,^a Michael W. S. Habib,^b Anthony E. Bishay,^a Konrad Siebor,^c Gizem Babaoglu,^d  Naweed I. Chowdhury,^d and Aaron C. Moberly^d 

^aSchool of Medicine, Vanderbilt University, Nashville, TN ^bDepartment of Computer Science, Duke University, Durham, NC ^cAmazon Web Services, Herndon, VA ^dDepartment of Otolaryngology—Head & Neck Surgery, Vanderbilt University Medical Center, Nashville, TN

ARTICLE INFO

Article History:

Received August 23, 2025

Revision received November 4, 2025

Accepted December 5, 2025

Editor-in-Chief: Ryan W. McCreery

Editor: Douglas P. Sladen

https://doi.org/10.1044/2025_AJA-25-00216

ABSTRACT

Purpose: Although some work has leveraged automated analyses of online communities to gain cochlear implant (CI) patient insights, there remains a gap in comparing human versus automated analysis of the nuanced, real-world experiences patients share outside clinical settings. This study characterizes experiences within the r/Cochlearimplants Reddit community and compares human to large language model (LLM) performance in annotating posts.

Method: Using reflexive thematic analysis, 996 publicly available r/Cochlearimplants posts (October 2024–June 2025) were manually coded and consolidated into themes. Three LLMs—OpenAI o3, Gemini 2.5 Pro, and Claude Sonnet 4—were prompted with the posts and human-generated codebook to perform post categorization. Model performance was evaluated against human coding using Cohen's kappa, percent agreement, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and time.

Results: Five themes emerged. Community engagement and support ($N = 944$, 94.8%) frequently involved eliciting advice ($N = 721$, 72.4%), seeking shared experiences ($N = 249$, 25.0%), and sharing negative experiences ($N = 247$, 24.8%). Other themes included the medical/surgical journey ($N = 463$, 46.5%), device/technical issues ($N = 343$, 34.4%), daily life/adjustments ($N = 236$, 23.7%), and media/outreach (7.2%, $N = 72$). OpenAI o3 and Gemini 2.5 Pro achieved the highest interrater reliability with human annotators ($\kappa = .35$ and $\kappa = .34$, respectively). OpenAI o3 had higher sensitivity (46.7%) but lower specificity (90.4%) than Gemini 2.5 Pro, which had the highest specificity (93.4%) but lower sensitivity (38.0%). Claude Sonnet 4 showed the lowest agreement ($\kappa = .25$) and PPV (30.9%). Compared to human annotation requiring 52 hr across all annotators, each LLM required less than 20 min.

Conclusions: Reddit posts revealed rich discourse across CI topics. LLMs demonstrated fair agreement with human coders and can quickly aid in large-scale qualitative analysis. Although careful model selection and human expertise remain essential for accurate interpretation, LLM annotation shows potential for real-time monitoring of patient concerns to inform counseling, rehabilitation strategies, and iterative device design.

Supplemental Material: <https://doi.org/10.23641/asha.31362847>

Correspondence to Daniel R. S. Habib: daniel.r.habib@vanderbilt.edu.

Disclosure: Daniel R. S. Habib founded KairoSys Applied Intelligence LLC, which is developing software to provide insights from digital data. All other authors have declared that no competing financial or nonfinancial interests existed at the time of publication.

Cochlear implants (CIs) are a well-established treatment for individuals with severe to profound hearing loss who receive limited benefit from hearing aids (Deep et al., 2019). Despite the proven benefits of CIs in improved quality of life and speech and language outcomes (McRackan

et al., 2018; Rasmussen et al., 2022), uptake remains low—estimated at less than 10% of eligible candidates in the United States—due to barriers that persist across age groups, geographic regions, and socioeconomic status (Neukam et al., 2024; Sorkin, 2013). Much of the existing literature focuses on quantifiable outcomes such as speech perception, device performance, and quality of life scores (Gaylor et al., 2013), while fewer studies explore patients' lived experiences with CIs. Although surveys and interviews are the predominant tools for capturing patient perspectives, they often include fewer than 100 participants and are limited by sampling bias, closed-ended question formats, and recall error (Hallberg & Ringdahl, 2004; Rapport et al., 2020; Rembar et al., 2009). These constraints risk overlooking nuanced concerns or the full range of day-to-day experiences that shape patient satisfaction, expectations, and decision making around implantation.

In recent years, social media platforms have emerged as alternative spaces where patients express health-related concerns, experiences, and decisions (Chen & Wang, 2021). Reddit, a large anonymous platform organized into thematic communities known as subreddits, has gained recognition for its role in health discourse (Proferes et al., 2021). Automated language models on Reddit data can surface patient perspectives, recurring themes, and everyday trends that traditional surveys and interviews may miss (Williams et al., 2023), and recent studies using computational models show good concordance with human coding while substantially reducing analysis time (Deiner et al., 2024; Prescott et al., 2024). Reddit-based analyses on otolaryngology topics can yield novel insights not typically disclosed in clinician-led settings (Manchaiah, Deshpande, et al., 2021). The primary online Reddit community for CIs is the *r/Cochlearimplants* subreddit, which serves as a forum where users share stories, ask questions, and provide peer support on topics ranging from CI device selection to emotional coping. Prior work analyzed this subreddit using BERTopic (Williams et al., 2023), an unsupervised clustering method that groups posts based on language patterns.

For clarity, artificial intelligence (AI) refers to systems that detect patterns and support decision making, generative AI (GenAI) encompasses models that create new text or insights from learned data, and large language models (LLMs) are transformer-based architectures that generate and summarize text at scale. Although rooted in computational linguistics and more commonly applied within the social sciences, these methods are only beginning to emerge in health research, which has traditionally relied on established epidemiological and qualitative approaches. As accessibility and performance improve, the rapid evolution and democratization of AI offer new opportunities to analyze patient experiences in ways previously unattainable. When applied to Reddit posts about CIs, LLMs can uncover patient perspectives and themes often missed by traditional

surveys or interviews. These insights may support health care providers in better anticipating patient concerns, refining counseling protocols, and identifying recurrent device issues that warrant manufacturer or provider action. To date, no study has applied a supervised approach using a predefined codebook, where a computational model assigns Reddit posts to relevant human-defined themes for direct comparison with manual coding.

This study aims to evaluate the performance of AI in the context of qualitative analysis of Reddit content related to CIs. Specifically, we compare differences in time required and percent agreement between human-coded and AI-driven classification of *r/Cochlearimplants* Reddit posts. The primary objective is to evaluate the performance of GenAI compared with human annotators in automating topic classification of Reddit posts, while the secondary objective is to present the key themes and categories reflecting CI users' experiences. Through this comparison, we seek to clarify the potential benefits and limitations of AI as a scalable complement to human qualitative research in hearing health.

Method

Data Collection

We extracted the 996 most recent posts from the *r/Cochlearimplants* subreddit (a Reddit subforum) with the corresponding dates ranging from October 22, 2024, to June 8, 2025 (Reddit, n.d.). The subreddit was created on March 1, 2015 and has about 7,300 subscribers. Data included the post date, title, body, link, and number of upvotes (i.e., likes) and comments. Shared anonymously by a Reddit user, each post contains free text consisting of a title and main body. No data were missing.

Thematic Analysis

Following post extraction, we applied reflexive thematic analysis (RTA): a qualitative method ideal for identifying patterns in rich free-text data (Braun & Clarke, 2021). This approach enabled us to systematically explore how individuals discuss CI-related experiences, questions, and concerns in an anonymous online setting. The analysis began with familiarization, during which researchers read and reread both the title and main body text of each post to develop a deep understanding of the content. Using an open coding approach, the team identified key concepts across posts. Six researchers independently coded an initial set of 20 posts from which they collaboratively developed a shared codebook (see Supplemental Material S1). Following a second round of annotating 20 posts, the codebook was developed inductively from the data and refined collaboratively

through discussion. Discrepancies were discussed as a group and resolved by consensus, a similar process as described previously (Profita et al., 2016). For the remaining posts, the team divided into three pairs, with each pair independently coding a subset of posts and discrepancies being resolved by consensus. Final codes were categorized into themes, and representative quotes were tabulated by theme with some wording adjustments to maintain user anonymity. The research team had a background in otolaryngology and medical research, whose prior clinical experience and familiarity with CI patient concerns informed code development, interpretation of posts, and thematic categorization, while efforts were made to minimize bias through independent coding. Coder bias was mitigated through regular consensus meetings while data saturation was monitored during the iterative coding process. The team logged the amount of time required to complete the annotation task. Krippendorff's alpha coefficient was calculated to assess interrater reliability between human annotators. This study was conducted in accordance with the Standards for Reporting Qualitative Research (SRQR) guidelines (see Supplemental Material S2).

Comparison to LLMs

Generative AI (GenAI) describes models that can create new content based on what they have learned, and an LLM is a type of GenAI model involving transformer-based systems that generate and summarize human text/language at scale (Shah et al., 2023). Three LLMs (OpenAI o3, Gemini 2.5 Pro, and Claude Sonnet 4) were provided the codebook established during the annotation process, the text content of each Reddit post (without any metadata such as date of post, username, or upvotes), and an identical standardized prompt to annotate each Reddit post according to the codebook. No prompt-engineering techniques (e.g., few-shot examples or temperature adjustments) were used to ensure parity across models, and the full prompt text is provided in Supplemental Material S3. Sensitivity, specificity, positive predictive value (PPV; i.e., precision), negative predictive value (NPV), interrater reliability (percent agreement), and Cohen's kappa were calculated pairwise between each set of AI model annotations and the final consensus set of human annotations. Agreement was interpreted according to values of 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect (Landis & Koch, 1977). All statistical analyses were performed using R statistical software Version 4.3.3 within the RStudio environment.

Results

Among the 996 r/Cochlearimplants posts, the median (interquartile range [IQR]) number of characters in the main

text body was 411 (226–704), the median (IQR) number of upvotes was three (two to six), and the median (IQR) number of comments was nine (five to 16). The median duration for the annotation task per person was 8.75 hr (IQR: 8.5–9.2 hr), including time spent discussing new themes and developing the codebook together. The total time spent across all annotators was 51.9 hr. The Krippendorff's alpha coefficient across human annotators was 0.97. RTA identified 24 distinct codes that were collapsed into five themes. Themes, their respective codes, and representative quotes are shown in Table 1. From most to least common, the themes were community engagement and support ($N = 944$, 94.8%), the medical and surgical journey ($N = 463$, 46.5%), device and technical issues ($N = 343$, 34.4%), daily life and adjustments ($N = 236$, 23.7%), and media and outreach ($N = 72$, 7.2%). The five most prevalent codes from most to least common include eliciting a response/advice ($N = 721$, 72.4%), general CI and medical questions ($N = 331$, 33.3%), general discussion ($N = 257$, 25.8%), seeking a shared experience ($N = 249$, 25.0%), and sharing a negative personal experience ($N = 247$, 24.8%). Four of these five most common codes fall within the community engagement and support theme, with the outlier being the second most prevalent code, general CI and medical questions, which resides within the medical and surgical journey theme.

Evaluating each theme further, posts within the community and engagement theme that sought answers or advice ($N = 721$, 72.4%) largely coincided with users seeking community ($N = 249$, 25.0%) or sharing personal experiences, with negative experiences ($N = 247$, 24.8%) being shared more often than positive ones ($N = 74$, 7.4%). The most popular topic in the medical and surgical journey theme involved medical inquiries and concerns regarding CIs ($N = 331$, 33.3%) and the surgical procedures ($N = 130$, 13.1%). Users less frequently shared their postsurgical experiences ($N = 92$, 9.2%) or their progress via objective results ($N = 30$, 3.0%). Device and technical issue posts most commonly included discussion around CI device connectivity and pairing ($N = 127$, 12.8%), selecting a device and brand ($N = 123$, 12.4%), and device troubleshooting ($N = 102$, 10.2%). Supplementary accessories and personalization ($N = 55$, 5.2%) as well as lost devices ($N = 3$, 0.3%) proved to be rarer points of discussion within the theme. Daily life and adjustment posts largely focused on general use complications ($N = 97$, 9.7%) and activity adjustments/restrictions ($N = 96$, 9.6%) associated with CIs. Posts about concerns with day-to-day use were more prevalent than posts that expressed financial concerns ($N = 55$, 5.2%). Within the theme of media and outreach, discussion posts with picture attachments were the most common ($N = 37$, 3.7%). Posts including video attachments, advertisements/promotions, or

Table 1. Count, percentage, and representative quotes by codes.

Theme/code	Count (%)	Representative quote
Community engagement and support	944 (94.8)	
Eliciting a response/advice	721 (72.4)	Can anybody that knows more about these topics offer some knowledge or advice?
General discussion	257 (25.8)	It's just been announced that the next gen of CI will soon be available, meaning this could be a big upgrade.
Seeking shared experience	249 (25.0)	Does anyone else with CIs feel this way? I'm curious if it's just me.
Sharing a negative personal experience	247 (24.8)	My implant continues to become infected—I get fevers, pain, and throw up each time.
Supporting another person	148 (14.9)	I'm reaching out to this community for guidance and support about CIs for my child.
Sharing a positive personal experience	74 (7.4)	This is the one medical procedure that has had the greatest impact on my life. My experience with music is exponentially better and I can hear from both ears for the first time in my life.
Offering tips/advice	28 (2.8)	If you have an issue with loud environments, follow these steps to enhance your experience . . .
Ethical/societal concerns	21 (2.1)	Do you have any thoughts about CIs as eugenics due to an erasure of Deaf culture?
Medical and surgical journey	463 (46.5)	
General CI and medical questions	331 (33.3)	How do CIs impact your speech?
Concerns/questions about surgery	130 (13.1)	Is there anything you wished you would have understood more about the surgery?
Postsurgical complications/symptoms	92 (9.2)	My tinnitus has increased following surgery, but I was expecting that.
Objective test results	30 (3.0)	Here is the official report of my audiogram.
Device and technical issues	343 (34.4)	
Device pairing/connectivity	127 (12.8)	Do smartwatches connect to nucleus sound processors?
Brand/device selection	123 (12.4)	Please help me choose one of the 3 CI providers.
Device troubleshooting	102 (10.2)	My wireless charging is not reliable, any tips?
Device accessories/personalization	55 (5.5)	I designed an ear cover to wear during my wrestling matches to act as a pad for the implanted part.
Lost device	3 (0.3)	I have misplaced my CI in my very cluttered apartment; can I find it using a magnet?
Daily life and adjustments	236 (23.7)	
General use complications	97 (9.7)	I've noticed dry skin and irritation on my scalp and ears only on the side of my CI.
Activity adjustments/restrictions	96 (9.6)	I've never had any issues prior to implantation, but can I go in a metal detector with my CI?
Financial concerns	55 (5.5)	I'd appreciate any advice as I'm pursuing insurance but out-of-pocket costs are so much.
Media and outreach	72 (7.2)	
Pictures	37 (3.7)	How can I get rid of this image— continues to show up in my app on iOS [picture attached].
Advertisements/promotions	13 (1.3)	I post content as a CI user and would love to share my social media accounts with others to check out my stuff.
Research/surveys	13 (1.3)	We're looking for adults to fill out our survey to understand the experiences of CI users.
Videos	13 (1.3)	I am excited and wanted to share progress with this video [video attached].

Note. Quotes were slightly modified to maintain anonymity while preserving original meaning. CI = cochlear implant.

research/surveys had a less frequent but even distribution (all $N = 13$, 1.3%).

Among the three LLMs evaluated, OpenAI's o3 achieved the highest Cohen's κ of .35 (see Table 2). It also demonstrated the best overall sensitivity (46.7%) while

maintaining high specificity (90.4%) and the highest NPV (92.1%). Google's Gemini 2.5 Pro performed similarly in terms of reliability ($\kappa = .34$) and achieved the highest overall agreement (86.3%), specificity (93.4%), and PPV (45.9%). However, it lagged in sensitivity (38.0%), suggesting a tendency toward underlabeling. Anthropic's Claude Sonnet 4

Table 2. Artificial intelligence (AI) model post annotation performance metrics compared to human annotations.

AI model	Cohen's kappa	Percent agreement (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
OpenAI o3	0.35	84.8	46.7	90.4	41.4	92.1
Gemini 2.5 Pro	0.34	86.3	38	93.4	45.9	91.1
Claude Sonnet 4	0.25	80.3	43.7	85.7	30.9	91.2

Note. PPV = positive predictive value; NPV = negative predictive value.

exhibited lower Cohen's kappa ($\kappa = .25$), percent agreement (80.3%), and PPV (30.9%). The annotation task required less than 20 min for each LLM.

Discussion

This study is unique in analyzing the largest sample of recent r/Cochlearimplants posts, being the first to manually annotate them, and being the only one to compare human and AI CI post annotations. Multilabel thematic analysis highlighted the prevalence of posts focused on community engagement and support, with frequently co-occurring sub-themes of eliciting advice, sharing personal experiences, and seeking connection. Other less common themes included the medical and surgical journey, device-related questions and troubleshooting, daily life and adjustments, and media and outreach. OpenAI o3 (Cohen's $\kappa = .35$) and Gemini 2.5 Pro ($\kappa = .34$) demonstrated the highest human-machine reliability, with trade-offs between sensitivity, specificity, and predictive values. Claude Sonnet 4 demonstrated overall lower reliability ($\kappa = .25$) and precision. Our analysis highlights the richness of patient-generated online discussion while demonstrating the potential of AI tools in analyzing large online communities to gain medical device insights.

r/Cochlearimplants Insights

Rich textual data from large forums with thousands of users, for example, Facebook (Almugathwi et al., 2020; Choudhury et al., 2017; Profita et al., 2018, 2016; Saxena et al., 2015; Zabielska & Żelazowska-Sobczyk, 2019) and X (formerly Twitter; Choudhury et al., 2017; Feier et al., 2023; Saxena et al., 2015), and common review platforms, such as Amazon (Bennett et al., 2021; Heselton et al., 2022; Manchaiah et al., 2019; Manchaiah, Swanepoel, et al., 2021), have proven valuable in understanding patient experiences. Because each post could be assigned to multiple codes, a more comprehensive understanding of the various experiences expressed by users was possible. Unsurprisingly, the most common theme is “community engagement and support,” since Reddit has risen in popularity as an outlet for people to discuss various health conditions (Chan et al., 2025), especially if there is a desire to obtain multiple quick responses while remaining anonymous (Chen & Wang,

2021; Xie et al., 2022) or if there are barriers to accessing care (Pleasure et al., 2022). Given that this theme also includes a sharing of experiences, patients may seek an outlet to discuss their CI experiences outside of a medical setting to process their situation or feel validated by other CI users. The “medical and surgical journey” and “device and technical issues” themes—while common—were far less prevalent, indicating that, while members seek practical and medical guidance, these needs are secondary to peer connection in the context of an online environment. This idea aligns with prior studies of social media posts involving CI users on other sites, for example, Facebook, Twitter, and YouTube (Feng et al., 2023; Saxena et al., 2015), and other medical topics, for example, cancer and surgery (Braun et al., 2019; Koball et al., 2025). Clinically, this can be seen as a positive signal since patients may feel adequately informed and supported by their medical team (Li, Zhou, et al., 2024), deciding to consult Reddit for fixing minor device issues and quelling implantation fears. Similarly, daily life and adjustments were even less frequently discussed. This could be due to multiple factors, including the nonurgent nature of these concerns and the ability to discuss solutions with the medical team, as indicated by our representative quotes. Finally, media and outreach were barely discussed. One of the main benefits of using a platform such as Reddit is remaining anonymous (Chen & Wang, 2021), so photos and videos may risk compromising anonymity. Additionally, advertisements and research may focus on more niche topics that have less applicability to a long-text, conversational support community such as Reddit. Although Reddit can serve as a valid crowdsourcing platform for research, surveys may be limited by data validity concerns (Goodwin et al., 2023).

Human Versus LLM Post Categorization

Most prior work on LLM annotation has focused on single-label tasks or qualitative comparisons, leaving a gap in knowledge about complex multilabel classification problems such as patient forum analysis. Wulcan et al. (2025) reported near-perfect reproducibility for GPT-4o in veterinary clinical information extraction (Cohen's $\kappa = .98$), exceeding human-human interrater reproducibility ($\kappa = .80$). However, their tasks involved predefined structured extraction using one company's models, not comparing multiple

companies' models on open-ended forum coding. Other studies have compared LLMs using interrater reliability metrics (e.g., Cohen's kappa) for the U.S. Medical Licensing Examination (Mishra et al., 2025) and pharmacology questions (Salman et al., 2025), which have objective answers. There has also been some work using LLMs to analyze qualitative patient interviews (Li, Fernandez, et al., 2024). Nevertheless, none have compared multiple models on a multilabel annotation task of Reddit posts.

In this study, we compared leading LLMs on CI forum posts. Unlike the high human interrater reliability in this study (Krippendorff's α of .97) and in prior work for a post annotation task about CI customization (Profita et al., 2016), there was a fair level of agreement between human and AI categorization in this study. Although lower than human-human agreement, this performance is still useful for exploratory analyses and early topic detection; AI-assisted annotation allows researchers to prioritize content for human review, reduce manual coding time, and detect emerging topics while substantially reducing the time and effort required for large-scale thematization.

Notably, model-specific biases were observed. All three AI models in this study tended toward underlabeling, particularly for more vague codes such as "general CI and medical questions" and "general discussion" compared to more specific categories such as "financial concerns" and "objective test results." However, Claude Sonnet 4's lower precision (PPV = 30.9%) likely reflects a higher tendency toward overgeneralization and assigning incorrect but plausible labels ("hallucination"), consistent with its design for interpretive reasoning tasks. Although OpenAI o3 and Gemini 2.5 Pro demonstrated comparable interrater reliability ($\kappa = .34-.35$), OpenAI o3 offered a better sensitivity-specificity balance, indicating a favorable trade-off between identifying relevant content and minimizing overlabeling.

Overall, these findings suggest that LLMs can serve as a practical tool for efficient, large-scale annotation of patient-generated content while maintaining reasonable accuracy for early topic detection, but high-stakes decision making would require supplementation with additional context or human-validated responses to improve reliability. Although prior work using BERTopic achieved a higher Cohen's kappa ($\kappa = .62$) between machine and human coding (Williams et al., 2023), likely due to domain-specific tuning, its unsupervised clustering can produce topics that lack clinical relevance, require substantial manual interpretation, and vary with parameter settings. Such variability limits reproducibility and adaptability to nuanced or evolving coding schemes compared to flexible prompt-based approaches. Performance differences among LLMs may stem not only from prompt design but also from underlying training data and model alignment objectives. By quantifying both reliability and error profiles, our

study reinforces the importance of aligning model selection with specific analytic objectives and the acceptable balance between false positives and false negatives in multilabel classification tasks.

Ethical Considerations

When examining Reddit discussions about CI, researchers enter an ethical space that is not always anticipated by those contributing content (Fiesler et al., 2024; Proferes et al., 2021). Even though posts are publicly visible, many participants may not expect their posts to become data in scholarly work (Stommel & de Rijk, 2021). Safeguarding trust in these communities involves minimizing the likelihood that individuals can be recognized, for example, by paraphrasing excerpts, removing specific contextual details, and avoiding linking usernames or platform activity across sites (Benton et al., 2017; Gliniecka, 2023). Ethical responsibility extends beyond privacy: Reporting patterns of dissatisfaction could alter patient decision making, reinforce stigma, or change community interactions (David & Werner, 2016). Researchers should therefore balance the value of amplifying patient perspectives with the responsibilities of preventing harm and engaging respectfully with digital spaces from which these insights are drawn (Chancellor et al., 2019; Conway, 2014; Gliniecka, 2023; McKee, 2013).

Limitations

This study exhibits some limitations. First, the sample of 996 posts was drawn from the most recent 7 months of activity on the r/Cochlearimplants subreddit and may not be fully representative of the broader subreddit or of CI users who do not participate in online forums. There are also "lurkers" (users who read but do not post) whose perspectives are absent from our data set, limiting insights about patients who may be hesitant to share publicly. While earlier posts are not included, the recency of the data emphasizes current experiences and concerns, particularly those related to relevant technologies and evolving practices. Second, thematic analysis involves an element of subjectivity, and the categorization of posts may have been influenced by individual judgment. To mitigate this, we employed a multistep coding process involving multiple independent reviewers to resolve discrepancies, enhancing consistency and reliability. As with all studies relying on unstructured social media data, concerns may arise regarding the accuracy and completeness of self-reported information. There may be biases in online communities, variability in data quality, and challenges in accurately interpreting nuanced or context-dependent language. However, prior research has demonstrated the value of social media platforms as a rich source of real-world perspectives, particularly for responsibly capturing patient experiences that may be underrepresented in traditional clinical settings (Correia et al., 2020; Giorgi et al., 2022, 2024; Lane et al., 2023; Levanti et al., 2023;

Murray et al., 2024). Another important consideration is the demographic skew of Reddit users, which may influence the content and perspectives represented in our data set. Reddit users are disproportionately young, well educated, and male; they are more likely than the general population to identify as liberal, have higher incomes, and engage with the platform for entertainment, news, and technology-related interests (Barthel et al., 2016; Duarte, 2025; YouGov, 2023). Although no demographic data are available for the r/Cochlearimplants subreddit, these broader patterns suggest that Reddit users may represent a subset of the CI community. Although findings from this study may not fully capture the experiences of older individuals, the findings provide valuable insight into the diverse perspectives of digitally active individuals with a CI. Because our analysis was restricted to English-language posts, the experiences of non-English-speaking users are also likely underrepresented. Recognizing demographic and linguistic constraints is important for contextualizing results and for guiding future work toward complementary data sources that can enhance external validity. Additionally, LLMs are limited by the prompt and context provided. If the codebook had better descriptions with good and bad examples of annotations, the models might have performed better. However, this would stray further from the purpose of establishing if LLMs can be used for this Reddit annotation task with as little human involvement as possible. The comparably low kappa values suggest that while out-of-the-box performance is limited with publicly available LLMs, there may be a role for improved performance with using a specifically trained LLM. Additionally, it is unclear if LLMs perform better inferences than human annotations when codewords are not explicitly stated in social media posts, especially considering many posts likely contain slang terms and typos. Although the human annotations were considered the ground truth in comparison with the AI models, there might have been mistakes in the human annotations that the models had annotated correctly but were considered incorrect.

Implications and Future Directions

This study helps affirm the value of integrating AI-driven analysis with social media narratives to better understand the lived experiences of individuals with CIs. Natural language processing (NLP) tools have been used to successfully identify emerging patient concerns across conditions, such as COVID-19 vaccine hesitancy and mental health stigma, which suggests applicability to audiologic contexts as we did in our study (Al-Garadi et al., 2022; Jang et al., 2022). Within the CI sphere itself, user feedback has been used to understand patient experiences such as perception of music (Gfeller et al., 2019). Frequent references to sound quality distortions or music aversion in online discourse, for example, can guide postimplant auditory training protocols (e.g., more fine-grained spectral resolution training). Additionally, AI

has been used to detect hearing loss-related misinformation and document psychological distress signals in online discourse, which highlights its potential roles in both research and real-time public health monitoring (Choudhury et al., 2017; Saxena et al., 2015). Online user expression of emotional distress may help clinicians recognize psychosocial burdens earlier, supporting integration of peer-support programming or counseling into standard rehabilitation.

The capacity of AI to rapidly process large volumes of data enables efficient analysis of social media posts, including those on Reddit, to identify CI patient needs in real time—an otherwise impractical task for human reviewers. Although core challenges faced by CI users may remain consistent, AI-driven analysis can detect emerging cultural, political, financial, or social shifts within the community. These insights can be translated into more responsive, patient-centered audiologic care informed by evolving online discourse. AI-based social media analysis can be applied to monitor emerging patient concerns, guide the development of educational materials, and inform device design or counseling strategies based on real-world user experiences. As shown in other domains, such as real-time vaccine sentiment tracking using NLP, LLMs may allow for early detection of behaviors such as device nonadherence that can support timely clinical intervention (Huang et al., 2024). Future work should explore expanding AI tools beyond categorized subforums such as Reddit to other platforms such as X, YouTube, or TikTok, for example, by using hashtags. Tapping into patient-generated data across diverse online communities can help inform patient-centered counseling and medical device improvements.

Conclusions

This is the first study to assess interrater reliability between humans and multiple LLMs in a multilabel thematic analysis of discussions within the Reddit CI community. Analysis of this large online data set illustrates the value of online patient discourse in discerning patient concerns and attitudes that may not be expressed in clinical settings, which can inform care. Our findings highlight that current AI models can quickly categorize online content about CIs without fine-tuning or example annotations. Notably, by bringing to light real-time concerns beyond the clinic walls such as emotional distress, this methodology provides a novel adjunct to standard follow-up instruments. The use of LLMs to monitor online discourse can serve as an early warning layer to flag issues that might otherwise evade standard audiometric care. Trade-offs in various performance metrics call for thoughtful consideration of the best model for post annotation tasks. This study validates the future role of AI-assisted analysis in real-time monitoring of patient concerns to guide patient-centered clinical support, education, and medical device improvements.

Ethics Statement

This study does not involve human subjects or identifiable information.

Data Availability Statement

Data are publicly available on Reddit and are available upon request from the corresponding author. The data dictionary (Supplemental Material S1) and R code for statistical analysis (Supplemental Material S4) are provided to facilitate reproducibility.

Acknowledgments

This article is published Open Access under a read and publish agreement between Vanderbilt University and the American Speech-Language-Hearing Association.

References

- Al-Garadi, M. A., Yang, Y.-C., & Sarker, A. (2022). The role of natural language processing during the COVID-19 pandemic: Health applications, opportunities, and challenges. *Healthcare, 10*(11), Article 2270. <https://doi.org/10.3390/healthcare10112270>
- Almugathwi, M., Wearden, A., Green, K., Hill-Feltham, P., & Powell, R. (2020). Online support group users' perceptions and experiences of bone-anchored hearing aids (BAHAs): A qualitative study. *International Journal of Audiology, 59*(11), 850–858. <https://doi.org/10.1080/14992027.2020.1771440>
- Barthel, M., Stocking, G., Holcomb, J., & Mitchell, A. (2016). *Seven-in-ten Reddit users get news on the site*. Pew Research Center. <https://www.pewresearch.org>
- Bennett, R. J., Swanepoel, D. W., Ratinaud, P., Bailey, A., Pennebaker, J. W., & Manchaiah, V. (2021). Hearing aid acquisition and ownership: What can we learn from online consumer reviews? *International Journal of Audiology, 60*(11), 917–926. <https://doi.org/10.1080/14992027.2021.1931487>
- Benton, A., Coppersmith, G., & Dredze, M. (2017). Ethical research protocols for social media health research. *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing, 94–102*. <https://doi.org/10.18653/v1/w17-1612>
- Braun, L. A., Zomorodbakhsch, B., Keinki, C., & Huebner, J. (2019). Information needs, communication and usage of social media by cancer patients and their relatives. *Journal of Cancer Research and Clinical Oncology, 145*(7), 1865–1875. <https://doi.org/10.1007/s00432-019-02929-9>
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE Publications.
- Chan, G. J., Fung, M., Warrington, J., & Nowak, S. A. (2025). Understanding health-related discussions on Reddit: Development of a topic assignment method and exploratory analysis. *JMIR Formative Research, 9*, Article e55309. <https://doi.org/10.2196/55309>
- Chancellor, S., Baumer, E. P. S., & De Choudhury, M. (2019). Who is the “human” in human-centered machine learning: The case of predicting mental health from social media. *Proceedings of the ACM on Human-Computer Interaction, 3*(CSCW), 1–32, Article 147. <https://doi.org/10.1145/3359249>
- Chen, J., & Wang, Y. (2021). Social media use for health purposes: Systematic review. *Journal of Medical Internet Research, 23*(5), Article e17917. <https://doi.org/10.2196/17917>
- Choudhury, M., Dinger, Z., & Fichera, E. (2017). The utilization of social media in the hearing aid community. *American Journal of Audiology, 26*(1), 1–9. https://doi.org/10.1044/2016_AJA-16-0044
- Conway, M. (2014). Ethical issues in using twitter for public health surveillance and research: Developing a taxonomy of ethical concepts from the research literature. *Journal of Medical Internet Research, 16*(12), Article e290. <https://doi.org/10.2196/jmir.3617>
- Correia, R. B., Wood, I. B., Bollen, J., & Rocha, L. M. (2020). Mining social media data for biomedical signals and health-related behavior. *Annual Review of Biomedical Data Science, 3*(1), 433–458. <https://doi.org/10.1146/annurev-biodatasci-030320-040844>
- David, D., & Werner, P. (2016). Stigma regarding hearing loss and hearing aids: A scoping review. *Stigma and Health, 1*(2), 59–71. <https://doi.org/10.1037/sah0000022>
- Deep, N. L., Dowling, E. M., Jethanamest, D., & Carlson, M. L. (2019). Cochlear implantation: An overview. *Journal of Neurological Surgery Part B: Skull Base, 80*(02), 169–177. <https://doi.org/10.1055/s-0038-1669411>
- Deiner, M. S., Honcharov, V., Li, J., Mackey, T. K., Porco, T. C., & Sarkar, U. (2024). Large language models can enable inductive thematic analysis of a social media corpus in a single prompt: Human validation study. *JMIR Infodemiology, 4*, Article e59641. <https://doi.org/10.2196/59641>
- Duarte, F. (2025). *Reddit user age, gender, & demographics (2025)*. Exploding Topics. <https://explodingtopics.com/blog/reddit-users>
- Feier, J. S., Nguyen, K., & Choi, J. S. (2023). Twitter perspectives on cochlear implantation: Sentiment and thematic analysis. *Otolaryngology—Head and Neck Surgery, 169*(3), 642–650. <https://doi.org/10.1002/ohn.292>
- Feng, S. J., Yu, M., Leong, S., & Chern, A. (2023). Exploration and analysis of cochlear implant content on social media. *Cureus, 15*(9), Article e45801. <https://doi.org/10.7759/cureus.45801>
- Fiesler, C., Zimmer, M., Proferes, N., Gilbert, S., & Jones, N. (2024). Remember the human: A systematic review of ethical considerations in Reddit research. *Proceedings of the ACM on Human-Computer Interaction, 8*(GROUP), 1–33, Article 5. <https://doi.org/10.1145/3633070>
- Gaylor, J. M., Raman, G., Chung, M., Lee, J., Rao, M., Lau, J., & Poe, D. S. (2013). Cochlear implantation in adults: A systematic review and meta-analysis. *JAMA Otolaryngology—Head & Neck Surgery, 139*(3), Article 265. <https://doi.org/10.1001/jamaoto.2013.1744>
- Gfeller, K., Driscoll, V., & Schwalje, A. (2019). Adult cochlear implant recipients' perspectives on experiences with music in everyday life: A multifaceted and dynamic phenomenon. *Frontiers in Neuroscience, 13*, Article 1229. <https://doi.org/10.3389/fnins.2019.01229>
- Giorgi, S., Bellew, D., Habib, D. R. S., Sedoc, J., Smitterberg, C., Devoto, A., Himelein-Wachowiak, M., & Curtis, B. (2024). Lived experience matters: Automatic detection of stigma toward people who use substances on social media. *Proceedings of the International AAAI Conference on Web and Social Media, 18*, 474–487. <https://doi.org/10.1609/icwsm.v18i1.31328>
- Giorgi, S., Himelein-wachowiak, M., Habib, D., Ungar, L., & Curtis, B. (2022). Nonsuicidal self-injury and substance use disorders: A shared language of addiction. In *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology* (pp. 177–183). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.clpsych-1.15>

- Gliniecka, M.** (2023). The ethics of publicly available data research: A situated ethics framework for Reddit. *Social Media + Society*, 9(3), Article 20563051231192021. <https://doi.org/10.1177/20563051231192021>
- Goodwin, S. R., Dwyer, M. J., Caliva, S. L., Burrows, C. A., & Raiff, B. R.** (2023). Using Reddit as a recruitment strategy for addiction science research. *Journal of Substance Use and Addiction Treatment*, 148, Article 209011. <https://doi.org/10.1016/j.josat.2023.209011>
- Hallberg, L. R.-M., & Ringdahl, A.** (2004). Living with cochlear implants: Experiences of 17 adult patients in Sweden. *International Journal of Audiology*, 43(2), 115–121. <https://doi.org/10.1080/14992020400050016>
- Heselson, T., Bennett, R. J., Manchaiah, V., & Swanepoel, D. W.** (2022). Online reviews of hearing aid acquisition and use: A qualitative thematic analysis. *American Journal of Audiology*, 31(2), 284–298. https://doi.org/10.1044/2021_AJA-21-00172
- Huang, L.-C., Eiden, A. L., He, L., Annan, A., Wang, S., Wang, J., Manion, F. J., Wang, X., Du, J., & Yao, L.** (2024). Natural language processing-powered real-time monitoring solution for vaccine sentiments and hesitancy on social media: System development and validation. *JMIR Medical Informatics*, 12, Article e57164. <https://doi.org/10.2196/57164>
- Jang, H., Rempel, E., Roe, I., Adu, P., Carenini, G., & Janjua, N. Z.** (2022). Tracking public attitudes toward COVID-19 vaccination on tweets in Canada: Using aspect-based sentiment analysis. *Journal of Medical Internet Research*, 24(3), Article e35016. <https://doi.org/10.2196/35016>
- Koball, A. M., Dougherty, E., Stanley, T. B., Kerver, G. A., Grothe, K. B., Clark, M. M., Ghanem, O. M., Patten, C. A., Phelan, S., & Engel, S. G.** (2025). Social media use among patients before and after metabolic and bariatric surgery: A systematic review. *Surgery for Obesity and Related Diseases*, 21(7), 838–847. <https://doi.org/10.1016/j.soard.2024.12.022>
- Landis, J. R., & Koch, G. G.** (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <http://www.ncbi.nlm.nih.gov/pubmed/843571>, <https://doi.org/10.2307/2529310>
- Lane, J. M., Habib, D., & Curtis, B.** (2023). Linguistic methodologies to surveil the leading causes of mortality: Scoping review of twitter for public health data. *Journal of Medical Internet Research*, 25, Article e39484. <https://doi.org/10.2196/39484>
- Levanti, D., Monastero, R. N., Zamani, M., Eichstaedt, J. C., Giorgi, S., Schwartz, H. A., & Meliker, J. R.** (2023). Depression and anxiety on twitter during the COVID-19 stay-at-home period in 7 major U.S. cities. *AJPM Focus*, 2(1), Article 100062. <https://doi.org/10.1016/j.focus.2022.100062>
- Li, K., Zhou, R., Zheng, W., Zhang, Y., & Qiu, J.** (2024). Knowledge, attitude, and practice toward cochlear implants among deaf patients who received cochlear implants. *Scientific Reports*, 14(1), Article 4451. <https://doi.org/10.1038/s41598-024-55006-8>
- Li, K. D., Fernandez, A. M., Schwartz, R., Rios, N., Carlisle, M. N., Amend, G. M., Patel, H. V., & Breyer, B. N.** (2024). Comparing GPT-4 and human researchers in health care data analysis: Qualitative description study. *Journal of Medical Internet Research*, 26, Article e56500. <https://doi.org/10.2196/56500>
- Manchaiah, V., Amlani, A. M., Bricker, C. M., Whitfield, C. T., & Ratinaud, P.** (2019). Benefits and shortcomings of direct-to-consumer hearing devices: Analysis of large secondary data generated from Amazon customer reviews. *Journal of Speech, Language, and Hearing Research*, 62(5), 1506–1516. https://doi.org/10.1044/2018_JSLHR-H-18-0370
- Manchaiah, V., Deshpande, A. K., Bellon-Harn, M. L., & Boyd, R. L.** (2021). Sudden sensorineural hearing loss: What can we learn from examining Reddit posts? *The Journal of Laryngology & Otolaryngology*, 135(12), 1109–1113. <https://doi.org/10.1017/S0022215121003042>
- Manchaiah, V., Swanepoel, D. W., Bailey, A., Pennebaker, J. W., & Bennett, R. J.** (2021). Hearing aid consumer reviews: A linguistic analysis in relation to benefit and satisfaction ratings. *American Journal of Audiology*, 30(3), 761–768. https://doi.org/10.1044/2021_AJA-21-00061
- McKee, R.** (2013). Ethical issues in using social media for health and health care research. *Health Policy*, 110(2–3), 298–301. <https://doi.org/10.1016/j.healthpol.2013.02.006>
- McRackan, T. R., Bauschard, M., Hatch, J. L., Franko-Tobin, E., Droghini, H. R., Nguyen, S. A., & Dubno, J. R.** (2018). Meta-analysis of quality-of-life improvement after cochlear implantation and associations with speech recognition abilities. *The Laryngoscope*, 128(4), 982–990. <https://doi.org/10.1002/lary.26738>
- Mishra, V., Lurie, Y., & Mark, S.** (2025). Accuracy of LLMs in medical education: Evidence from a concordance test with medical teacher. *BMC Medical Education*, 25(1), Article 443. <https://doi.org/10.1186/s12909-025-07009-w>
- Murray, C., Mitchell, L., Tuke, J., & Mackay, M.** (2024). Revealing patient-reported experiences in healthcare from social media using the design-acquire-process-model-analyse-visualise framework. *Digital Health*, 10. <https://doi.org/10.1177/20552076241251715>
- Neukam, J. D., Kunnath, A. J., Patro, A., Gifford, R. H., Haynes, D. S., Moberly, A. C., & Tamati, T. N.** (2024). Barriers to cochlear implant uptake in adults: A scoping review. *Otology & Neurotology*, 45(10), e679–e686. <https://doi.org/10.1097/MAO.0000000000004340>
- Pleasure, Z. H., Frohirth, L. F., Li, N., & Polis, C. B.** (2022). A content analysis of Reddit users' posts about challenges to contraceptive care-seeking during COVID-19-related restrictions in the United States. *Journal of Health Communication*, 27(10), 746–754. <https://doi.org/10.1080/10810730.2022.2157911>
- Prescott, M. R., Yeager, S., Ham, L., Rivera Saldana, C. D., Serrano, V., Narez, J., Paltin, D., Delgado, J., Moore, D. J., & Montoya, J.** (2024). Comparing the efficacy and efficiency of human and generative AI: Qualitative thematic analyses. *Journal of Medical Internet Research - Artificial Intelligence*, 3, Article e54482. <https://doi.org/10.2196/54482>
- Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M.** (2021). Studying Reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media + Society*, 7(2). <https://doi.org/10.1177/20563051211019004>
- Profita, H. P., Stangl, A., Matuszewska, L., Sky, S., & Kane, S. K.** (2016). Nothing to hide: Aesthetic customization of hearing aids and cochlear implants in an online community. *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, 219–227. <https://doi.org/10.1145/2982142.2982159>
- Profita, H. P., Stangl, A., Matuszewska, L., Sky, S., Kushalnagar, R., & Kane, S. K.** (2018). “Wear it loud”: How and why hearing aid and cochlear implant users customize their devices. *ACM Transactions on Accessible Computing*, 11(3), 1–32. <https://doi.org/10.1145/3214382>
- Rapport, F., Hughes, S. E., Boisvert, I., McMahon, C. M., Braithwaite, J., Faris, M., & Bierbaum, M.** (2020). Adults' cochlear implant journeys through care: A qualitative study. *BMC Health Services Research*, 20(1), Article 457. <https://doi.org/10.1186/s12913-020-05334-y>
- Rasmussen, K. M. B., West, N. C., Bille, M., Sandvej, M. G., & Cayé-Thomassen, P.** (2022). Cochlear implantation improves both speech perception and patient-reported outcomes: A prospective follow-up study of treatment benefits among adult

- cochlear implant recipients. *Journal of Clinical Medicine*, 11(8), Article 2257. <https://doi.org/10.3390/jcm11082257>
- Reddit**. (n.d.). *r/Cochlearimplants*. Retrieved August 23, 2025, from <https://www.reddit.com/r/Cochlearimplants/>
- Rembar, S., Lind, O., Arnesen, H., & Helvik, A.-S.** (2009). Effects of cochlear implants: A qualitative study. *Cochlear Implants International*, 10(4), 179–197. <https://doi.org/10.1179/cim.2009.10.4.179>
- Salman, I. M., Ameer, O. Z., Khanfar, M. A., & Hsieh, Y.-H.** (2025). Artificial intelligence in healthcare education: Evaluating the accuracy of chatGPT, copilot, and google gemini in cardiovascular pharmacology. *Frontiers in Medicine*, 12, Article 1495378. <https://doi.org/10.3389/fmed.2025.1495378>
- Saxena, R. C., Lehmann, A. E., Hight, A. E., Darrow, K., Remenschneider, A., Kozin, E. D., & Lee, D. J.** (2015). Social media utilization in the cochlear implant community. *Journal of the American Academy of Audiology*, 26(2), 197–204. <https://doi.org/10.3766/jaaa.26.2.8>
- Shah, N. H., Entwistle, D., & Pfeffer, M. A.** (2023). Creation and adoption of large language models in medicine. *JAMA*, 330(9), 866–869. <https://doi.org/10.1001/jama.2023.14217>
- Sorkin, D. L.** (2013). Cochlear implantation in the world's largest medical device market: Utilization and awareness of cochlear implants in the United States. *Cochlear Implants International*, 14(Suppl. 1), S12–S54. <https://doi.org/10.1179/1467010013Z.00000000076>
- Stommel, W., & de Rijk, L.** (2021). Ethical approval: None sought. How discourse analysts report ethical issues around publicly available online data. *Research Ethics*, 17(3), 275–297. <https://doi.org/10.1177/1747016120988767>
- Williams, C. Y. K., Li, R. X., Luo, M. Y., & Bance, M.** (2023). Exploring patient experiences and concerns in the online cochlear implant community: A cross-sectional study and validation of automated topic modelling. *Clinical Otolaryngology*, 48(3), 442–450. <https://doi.org/10.1111/coa.14037>
- Wulcan, J. M., Jacques, K. L., Lee, M. A., Kovacs, S. L., Dausend, N., Prince, L. E., Wulcan, J., Marsilio, S., & Keller, S. M.** (2025). Classification performance and reproducibility of GPT-4 omni for information extraction from veterinary electronic health records. *Frontiers in Veterinary Science*, 11, Article 1490030. <https://doi.org/10.3389/fvets.2024.1490030>
- Xie, D. X., Boss, E. F., & Stewart, C. M.** (2022). An exploration of otolaryngology in the Reddit community. *The Laryngoscope*, 132(2), 284–286. <https://doi.org/10.1002/lary.29696>
- YouGov**. (2023). *Reaching redditors report: Inside the hivemind*. <https://yougov.com/reports/47908-us-reaching-redditors-report>
- Zabielska, M., & Żelazowska-Sobczyk, M.** (2019). Hearing aids and cochlear implants—A study of the discourse of patients' English ENT online support forums. *Token*, 9, 143–165. <https://doi.org/10.25951/2960>