National Journal of Multidisciplinary Research and Development
www.newresearchjournal.com
ISSN: 2455-9040
Received: 08-08-2025, Accepted: 07-09-2025, Published: 22-09-2025
Volume 10, Issue 3, 2025, Page No. 82-85

A linguistic analysis of human and ai content creation in news media: A corpus-based study of mental health news coverage

Kiran[1], Umesh Arya[2]
[1] Research Scholar, Department of Mass Communication, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India
[2] Professor, Department of Mass Communication, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India

Abstract
This study investigates the linguistic and structural differences between human-written and AI-generated content of news articles on mental health, using The Times of India articles and AI-generated content from ChatGPT and Gemini AI. Through corpus analysis using AntConc software, the study examines keywords, collocational behaviour, sentiment, modality, and framing strategies across three distinct corpora. Using a corpus-based approach and software such as AntConc, the study examines 130 human-written articles from TOI and 130 articles written by AI between 1 January 2025 and 30 April 2025. The findings reveal that TOI articles exhibit a richer emotional vocabulary, with frequent use of terms related to personal experience and stigma, while ChatGPT and Gemini content lean towards formal, analytical language, with Gemini focusing on policy and structural framing. Sentiment analysis shows that human-written articles are more emotionally engaged, employing stronger, assertive language, whereas AI-generated articles show more neutrality and caution, particularly in terms of modality.
Furthermore, the construction of mental health problems varies, with The Times of India prioritising individual narratives, ChatGPT taking an educational frame to promote awareness and stigma reduction, and Gemini content taking an institutional and policy-focused approach. These results highlight the limitations of AI in replicating the emotional complexity, empathy, and subtle narrative structures inherent in human journalism, especially in sensitive topics like mental health.

Keywords: Corpus linguistics, antconc, mental health, ai generated news, lexical patterns, and human versus ai content

Introduction

The advent of artificial intelligence (AI) in content generation has changed the face of journalism in the ever-changing media ecosystem. AI-generated content has advanced to the point where it can now replicate human writing in terms of structure, style, and coherence due to the development of Large Language Models (LLMs) like Google's Gemini and OpenAI's ChatGPT (Brown *et al*., 2020 [5]; OpenAI, 2023). Though such technologies are brought about efficiency and scalability, they also raise extremely serious issues of language authenticity, journalistic integrity, and the delicate handling of matters like mental health.

Media reporting of mental illness is responsible, as words and phrases for reporting mental illness can influence public opinion, stigma, and policy (Thornicroft *et al*., 2013) [25]. Human journalists are responsible for conforming to ethical standards and editorial guidelines that ensure responsible and sensitive reporting of mental illness (Samra *et al*., 2020) [22]. On the other hand, AI text can lack this context sensitivity, as in being drawn from patterns derived from a massive corpus of training data that could be biased, uninformed, or old (Bender *et al*., 2021) [4]. It is thus necessary to compare language features of human-written and AI-written news reporting on mental illness so that implications for representation, framing, and reader interpretation can be valued.

Previous research has explored linguistic and stylistic properties in traditional and online news (Bell, 1991 [3]; Cotter, 2010) [7], and recent automated media and reading of news have also been in the limelight (Graefe, 2016) [13]. But there have been practically no studies of systematically examining contrasting variation in language between robot-generated and human-generated accounts of mental illness, especially corpus-based studies. Corpus linguistics can facilitate empirical research on vast corpora of text and is therefore best equipped to identify patterns in lexical choice, sentiment, modality, and discourse framing (Baker, 2006 [2]; McEnery & Hardie, 2012) [18].

This research seeks to bridge that gap by performing a corpus-based linguistic analysis of human-authored mental health news stories from Indian and global newspapers (e.g., The Times of India, The Guardian) and AI-authored content on the same subject using LLMs such as ChatGPT and Gemini. Through an examination of mental health reporting, the study explores how AI and human writers deal with language sensitivity, subjectivity, and framing differently or similarly.

The results of this research will add to media linguistics and AI ethics by providing information on the linguistic appropriateness and credibility of AI-produced journalism, especially on delicate social topics. Since AI increasingly influences newsrooms around the world, such variations become crucial for researchers, media practitioners, and policymakers interested in ethical communication and mental illness reporting.

Review of Literature

The convergence of language, media, artificial intelligence, and mental health has emerged as a key field of study for both linguistics and media studies. As AI-powered content creation tools like ChatGPT, Gemini, and other Large Language Models (LLMS) are increasingly involved in

content production, it is imperative to know how these technologies differ linguistically from human-created content, particularly in the delicate field of reporting mental health.

Previously, media portrayals of mental health have been subjected to critical scrutiny regarding their influence on the general public's attitudes and inducing stigmatisation. Researchers such as Wahl (1995) [26] and, more recently, Coverdale *et al*. (2021) [8], note the influence of journalists' use of language to sustain stereotypes or to foster empathy and understanding. Reporting on mental illness is heavily reliant on narrative framing, word choice, and metaphor, all of which impact public comprehension (Seale, 2004 [23]; Goulden *et al*., 2011) [12]. In Indian news media, Rao and Varshney (2022) [21] research identified that, although coverage of mental health has grown, it remains superficial in terms of content and sensationalised or stereotyped in regards to mental illness, which suggests that there is a greater demand for responsible linguistic framing.

Alongside this evolution, corpus linguistics became a prominent method for researching journalistic language trends. Scholars like Baker (2006) [2] and Partington *et al*. (2013) [19] have shown the value of corpus-based approaches in exposing news reporting discursive patterns. Lexical units, modality, posture, collocations, and emotion may all be measured and examined through corpus analysis, which helps compare output produced by AI and humans. For instance, to examine bias and emotional tone—two factors that are crucial to tales about mental health—Clarke and Grieve (2017) [6] used corpus methods in their study of evaluative language within media texts.

Researchers are beginning to investigate the linguistic stability and effects of LLM-generated content as AI becomes more common in media. Large corpora of online text are used to train LLMS, and while they can mimic human speech, they can also replicate biases and false information in their training data (Bender *et al*., 2021) [4]. According to studies by Floridi and Chiriatti (2020) [11] and Kasneci *et al*. (2023) [15], LLMs may lack critical judgment and be context-insensitive, two traits that are crucial for reporting on mental health issues. Furthermore, AI-generated language will likely compromise the empathy and subtlety found in human-written mental health reporting in favour of neutrality or standardised presentation (Pavlik, 2023) [20].

Further recent comparative studies have appeared to test AI against human content more overtly. Diakopoulos (2019) [9], for instance, studied algorithmic accountability in newsrooms and concluded that AI can aid journalism but cannot replace editorial judgment. Marconi and Siegman (2017) [16] explored the use of automated journalism within companies such as Bloomberg and The Washington Post and described the struggle between speed and depth. In India, empirical research is limited, though exploratory research by Thapliyal (2023) [24] indicates Indian media's AI adoption is uncertain and under-researched.

Furthermore, linguistic characteristics such as sentiment, modality, and hedging in human as opposed to AI text have been examined in narrower areas such as financial news and politics, but have not yet been considered in reporting on mental illness. Hämäläinen *et al*. (2022) [14] demonstrated that humans prefer writing more factual and declarative sentence types, while AI systems prefer using hedges, metaphors, and affective words, particularly in delicate issues.

The lacunae in existing research necessitate the existence of a corpus-based, comparative linguistic analysis of AI-generated and human-generated news reports within the mental health genre. All studies have researched media linguistics or computer-generated text in isolation, and in no prior instance have the two been placed into a social-critical intersection as here. This research fills that gap by contrasting human and machine-generated language use within articles on mental health, taking into account tone differences, structure, empathy, and framing—necessary attributes not only for journalism but also for ethical communication of mental health.

Objective

1. To investigate and compare lexical patterning, keyword distribution, and collocation tendency in AI-generated and human-written news reports on mental illness with corpus study tools like AntConc.
2. To examine variations in linguistic parameters, sentiment, modality, and framing strategy between human-written and AI-generated mental health news through concordance and keyword-in-context (KWIC) analysis on AntConc.

Theoritical Framework

1. Framing Theory

Framing Theory, as defined by Entman (1993) [10], informs us that the way information is framed—or simply "framed" in the media—has some bearing on the manner audiences receive and make sense of the given information. In mental health news, the utilisation of certain words, tone, and emphasis can help diminish stigma or nurture stereotypes. The text applies Framing Theory to consider the divergence in the depiction by AI writers compared to human writers of mental illness, both on the lexical front, emotional, and in narration.

2. Corpus Linguistics Theory

The basis for the idea that language trends can be methodically addressed through collections of texts known as corpora is attributed to corpus linguistics theory, as articulated by McEnery and Hardie (2012) [18]. The approach supports the quantitative and repeatable analysis of variables, including keyword frequencies, collocations, concordance lines, and semantic trends, using software programs like AntConc. The theory gives us the ability to compare AI and human work empirically using linguistic evidence that is objective.

Methodology

1. Research Design

Through linguistic style comparison, this study compares human-generated versus AI-generated mental health news texts using a comparative, corpus-based, qualitative-quantitative research technique. In this instance, corpus linguistics techniques and resources, such as AntConc software (Anthony, 2023) [1], are employed to compare the lexical and discourse-level characteristics of texts created by humans and artificial intelligence methodically. Finding usage variations in language, tone, and framing that could affect public awareness of mental diseases is its goal.

2. Sample Size

The study analysed a total of 130 news articles focused on mental health, collected between January 1, 2025, and April

30, 2025. This sample included human-written articles from The Times of India and AI-generated articles created using ChatGPT and Gemini AI.

3. Data Collection

Data was collected from two primary sources:

1. Human-Written News Articles

A total of 130 news articles on mental health were sourced from The Times of India (India), reputable mainstream news platforms. The Times of India was chosen because it reaches over 13.5 million readers, positioning it as one of the most prominent and widely read English-language newspapers in India. Articles were selected based on relevance to mental health issues (e.g., depression, anxiety, suicide, therapy) and published during the period from 1 January 2025 to 30 April 2025. Articles were accessed via the websites' search tools using keywords such as mental health. A search was conducted on the "Google" search engine with the use of "mental health" keywords. An advanced search operator method utilising URL data collection search, like "allinurl", can successfully restrict relevant results.

2. AI-Generated News Articles

Using the same titles or headlines (duplicated from human-written news reports), 130 news reports were created using two Large Language Models: ChatGPT (OpenAI) and Gemini AI (Google DeepMind). The two models were instructed to create reports of approximately 500-800 words, journalistic tone and a fact-based tone, to simulate real-world AI content creation.

Result

The linguistic analysis using AntConc 4.2.0 provided clear distinctions between human-written articles (from The Times of India) and AI-generated counterparts (ChatGPT and Gemini AI). The results are organised according to the two main objectives.

1. Lexical Patterns, Keyword Frequency, and Collocational Behaviour

Keyword Frequency in Human-written articles showed a more emotional and socially based vocabulary, with terms like support, therapy, trauma, and stigma occurring more frequently. Symptoms, interventions, conditions, and data were frequently used in AI-generated content, which also used more formal and organised language. Awareness, burden, family, and treatment were common collocations of the phrase "mental health" in human publications, but initiatives, models, frameworks, and statistics were more frequently used in AI-generated papers (particularly Gemini).

Keyword Frequency

Keyword frequency was slightly higher in the Times of India articles using emotionally laden words such as struggle (86), stigma (74), family (71), and depression (92), indicating a personal and social perspective of mental health. ChatGPT articles leaned towards mental health education and care, and used words such as mental illness (89), treatment (78), stress (64), and self-care (58). Conversely, Gemini AI leaned more towards analytical and policy terms, with the typical terms like policy (84), framework (76), outcomes (67), and clinical (59) having a formalised and institutionalised pattern of approach towards mental health language.

Collocational Behaviour

In collocational contexts, mental health in Times of India news headlines was likely to co-occur with words like issues, struggles, burden, and therapy, consistent with the subjective and multifaceted nature of the issue. ChatGPT text, by contrast, utilised to combine mental health with words like campaigns, importance, and strategies, and emphasised awareness and action. Simultaneously, Gemini AI was leaning towards associating mental health with programs, access, and initiatives, hinting at a focus on structural and institutional responses to the state of mental health.

Word Lists and Token Variety

The TOI corpus had a higher type-token ratio (0.094), indicating lexical richness.

AI corpora showed slightly repetitive structures, with ChatGPT at 0.083 and Gemini at 0.081.

2. Sentiment, Modality, and Framing (via KWIC Analysis)

The Times of India's reporting was highly emotive and typically personal experiences or direct statements from people with mental illness. ChatGPT content was neutral-to-positive in tone with supportive and empathetic construction, whereas Gemini AI content was data-driven and analytical and expressed little emotion. Concerning modality, computer-generated articles used the modal verbs such as may, could, and might often, which expressed a tentative or speculative tone. In contrast, human-written articles were more forceful and decisive. In developing strategies, TOI prioritized narrative frames based on personal and family experiences, ChatGPT used an educational frame for the aim of raising awareness and lowering stigma, and Gemini targeted structural and policy frames with focus on institutional reaction and systemic issues.

Sentiment Analysis (via Concordance Lines)

Sentiment analysis on the articles showed widespread variation in emotive involvement and sentiment between the three sources. The Times of India articles employed rich emotive language like "heartbreaking stories," "fighting stigma," and "overcoming trauma," which indicated high emotive involvement with the readers. Conversely, ChatGPT's text contained a positive-neutral tone in sentences such as "mental health is important," "raising awareness," and "encouraging dialogue," showing an informative and supportive tone. On the other hand, Gemini AI's text leaned towards a formal, policy-like tone in sentences such as "data supports the need for reform" and "clinical interventions are needed," with a factual and analytical tone.

Modality

The news stories in the Times of India used mostly stronger and more aggressive verbs such as is, will, and must, conveying the message of certainty and straightforwardness. The AI-written content depicted a more cautious approach. Modal verbs such as could, might and can were used by ChatGPT at moderate levels to exhibit some degree of ambiguity. A more uncertain and hypothetical tone was evident in Gemini, which mostly drew on hypothetical vocabulary and ordinary language expressions such as likely, implies, and predicted.

Framing Strategies

The framing methods employed in the articles differed largely between the three sources. The Times of India used mostly a narrative or story framing style, personalising mental health conditions by emphasising individual stories and case studies. Such framing emphasised individual struggles and overcoming, rendering the issue more tangible to readers. By comparison, ChatGPT used a policy and institutional framing, focusing on increasing awareness and combating the stigma surrounding mental health. Its language was intended to inform and foster open conversation. In contrast, Gemini AI used a policy and institutional framing, with an emphasis on system issues, healthcare services, and data-driven facts. This strategy frequently emphasised the necessity for policy change and enhanced infrastructure in the mental health sector.

Comparison Table: Human vs AI-Generated Content on Mental Health News

Feature / Metric	Times of India (Human)	ChatGPT (AI)	Gemini (AI)
Total Articles	130	65	65
Corpus Size (Words)	76418	39209	37856
Top Keywords	stigma, trauma, therapy	awareness, wellness, care	policy, outcomes, access
Tone / Sentiment	Empathetic, personal	Positive-neutral, inclusive	Neutral, analytical
Modality Usage	Assertive (low modal)	Moderate (can, might)	High (likely, anticipated)
Framing Strategy	Narrative, human experience	Supportive, informative	Structural, policy-driven
Lexical Richness (TTR)	0.094	0.083	0.081
Collocates of 'mental health'	stigma, burden, youth	awareness, practice, needs	intervention, outcomes

Conclusion

This research investigated the linguistic and structural contrasts between human-written and AI-written news stories on mental health. The study examined, conducted with AntConc on a corpus of 130 articles from The Times of India and 130 AI-written articles (divided between ChatGPT and Gemini AI), identified striking differences in lexical patterns, sentiment, modality, and framing strategies.

Lexical and Collocational Variations in Human-authored articles from The Times of India show high emotional lexis and a high use of personal accounts, frequently adopting words such as stigma, family and trauma. However, both Gemini and ChatGPT utilised more formal, analytical lexis, with ChatGPT content prioritising awareness and Gemini prioritising policies and structures. Sentiment analysis indicated that TOI articles expressed higher emotional involvement through stories or narratives of struggle and recovery. AI-written articles, particularly Gemini, took a more neutral, analytical tone, using a lot of speculative language with words such as suggests and likely. ChatGPT walked the line between empathy and neutrality but did not have the same level of emotional depth as human journalism.

Framing Strategies: TOI news framed mental health concerns in a personal, experiential way, with emphasis on individual narratives. ChatGPT used an instructional and helpful tone while Gemini took a structural and policy-oriented framing approach with a focus on institutional reactions and reform requirements.

Lexical Richness and Sentence Structure: TOI had more lexical variety and longer, more complex sentences, while ChatGPT and Gemini used simpler structures. This indicates that AI-written content, although factually correct and coherent, is less emotionally engaging and nuanced than human-written articles.

Summing up, even though AI-sourced news has factually informed and properly framed content, it does not reflect the richness in depth, emotive engagement, and socio-cultural embedding exhibited by human reporting. These conclusions indicate the restrictions faced by AI when mimicking human reporting's richness of emotions as well as complex narratives in specific areas like mental health.

Reference

1. Anthony L. AntConc (Version 4.2.0) [Computer Software]. Waseda University, 2023. https://www.laurenceanthony.net/software
2. Baker P. Using Corpora in Discourse Analysis. Continuum, 2006.
3. Bell A. The language of news media. Blackwell,1991:84-85.
4. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency,2021:610-623.
5. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, *et al*. Language models are few-shot learners. Advances in neural information processing systems,2020:33:1877-1901.
6. Clarke I, Grieve J. Stylistic variation on the Donald Trump Twitter account: A linguistic analysis of tweets posted between 2009 and 2018. PLOS ONE,2017:14(9):e0222062.
7. Cotter C. News talk: Investigating the language of journalism. Cambridge University Press, 2010.
8. Coverdale J, Nairn R, Claasen D. The association between media portrayals and public attitudes toward people with mental illness: A meta-analytic review. Psychiatric Services,2021:72(7):715-722.
9. Diakopoulos N. Automating the News: How Algorithms Are Rewriting the Media. Harvard University Press, 2019.
10. Entman RM. Framing: Toward clarification of a fractured paradigm. Journal of Communication,1993:43(4):51-58.
11. Floridi L, Chiriatti M. GPT-3: Its nature, scope, limits, and consequences. Minds and Machines,2020:30:681-694.
12. Goulden R, Corker E, Evans-Lacko S, Rose D, Thornicroft G. Newspaper coverage of mental illness in the UK, 1992-2008. BMC Public Health,2011:11(1):1-8.
13. Graefe A. Guide to Automated Journalism. Tow Centre for Digital Journalism, Columbia University, 2016.
14. Hämäläinen M, Alnajjar K, Rueter J. Human vs. AI-generated text: Linguistic patterns and perceptual evaluation. Language Resources and Evaluation,2022:56:1395-1418.

15. Kasneci E, Sessler K, Bannert M. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education and Research. Frontiers in Artificial Intelligence,2023:6:2023.
16. Marconi F, Siegman A. The Future of Augmented Journalism: A Guide for Newsrooms in the Age of Smart Machines. Associated Press & Polis, 2017.
17. McEnery T, Hardie A. Corpus linguistics: Method, theory and practice. Cambridge University Press, 2011.
18. McEnery T, Hardie A. Corpus Linguistics: Method, Theory and Practice. Cambridge University Press, 2012.
19. Partington A, Duguid A, Taylor C. Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (CADS). John Benjamins, 2013.
20. Pavlik J. Generative AI and Journalism: Ethical Considerations for Newsrooms. Digital Journalism,2023:11(2):243-258.
21. Rao N, Varshney N. Mental health and Indian news media: Framing, frequency, and representation. Media Watch,2022:13(3):478-490.
22. Samra R, Hankir A, Fullerton D. Media Portrayal of Mental Illness and Its Implications for Public Attitudes and Stigma. The Psychiatric Bulletin,2020:44(3):89-92.
23. Seale C. Media and Health. Sage, 2004.
24. Thapliyal S. AI in Indian Newsrooms: Trends, Transparency, and Ethical Dilemmas. Indian Journal of Communication Research,2023:15(1):25-39.
25. Thornicroft G, Mehta N, Clement S, *et al*. Evidence for Effective Interventions to Reduce Mental-Health-Related Stigma and Discrimination. The Lancet,2013:387(10023):1123-1132.
26. Wahl OF. Media Madness: Public Images of Mental Illness. Rutgers University Press, 1995.