Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Emotional Differences Between Human- and AI-Generated Texts

Vinh Truong

truongnguyenxuanvinh@gmail.com

RMIT University https://orcid.org/0000-0003-4247-766X

Research Article

Keywords: emotion, affect, sentiment, AI, human

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11686057/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

AI is often conceptualized primarily as a cognitive tool, with its outputs assumed to convey information rather than emotional expression. This study challenges that assumption by systematically comparing the emotional characteristics of human- and AI-generated text across 25 languages using five emotional measures: sentiment polarity, sentiment categories, Valence-Arousal-Dominance (VAD), basic emotion classification, and fine-grained emotion classification. The results reveal a systematic difference in the emotional profiles of AI- and human-generated text across languages. Compared with human writing, AI-generated text generally exhibits higher valence, arousal, and dominance, while expressing fewer negative emotions. These findings extend our understanding of the affective characteristics of generative AI and highlight important implications for applications in customer service, marketing, personalized communication, and AI-generated content detection.

I. INTRODUCTION

Artificial Intelligence (AI) has traditionally been framed as a cognitive tool: a system that retrieves, summarizes, and reorganizes information on demand, with the language it produces often treated as a by-product of statistical pattern matching rather than as a meaningful vehicle for affect (Moreno-Ortiz & García-Gámez, 2026; Sui, 2025). This framing has shaped the study of AI-generated text, as much of the early natural language processing research on large language model (LLM) outputs focused primarily on factual accuracy, coherence, and fluency, treating emotional content as incidental rather than as a property worthy of systematic investigation in its own right (Foo et al., 2025). As generative AI systems increasingly move beyond single-turn question answering into sustained, human-facing roles—including customer service, education, marketing, companionship, and everyday writing assistance—this framing becomes increasingly inadequate. The emotional tone conveyed by an AI system, rather than merely the informational content it produces, can materially shape how users perceive, trust, and engage with AI-generated communication.

Research increasingly compares the emotional characteristics of AI- and human-generated text, but key gaps remain. Most studies use a single measure, such as sentiment polarity, which cannot distinguish calm positivity from emotional neutrality; this requires complementary methods such as VAD and categorical emotion analysis (Buscemi et al., 2025). Many studies also focus on one dataset or domain, limiting evidence of whether results generalise across academic, casual, business, and instructional writing (Bai et al., 2022; Bardol, 2025). Finally, the evidence is predominantly English-based, despite LLMs being used globally and trained on languages with unequal representation.

This study addresses these three gaps through a large-scale, systematic comparison of the emotional properties of human- and AI-generated text across multiple domains and languages, including both high- and low-resource linguistic settings. Rather than relying on a single sentiment measure, the study adopts a triangulated approach that combines multilingual sentiment scoring, rule-based sentiment categorization, Valence-Arousal-Dominance dimensional scoring, basic categorical emotion

classification (Ekman, 1984), and fine-grained categorical emotion classification based on GoEmotions (Demszky et al., 2020). The use of complementary measurement frameworks enables convergent evidence to be examined across different conceptualizations of emotion, reducing reliance on the characteristics or limitations of any single measurement approach.

Accordingly, this study addresses the following research question: **How does the emotional profile of AI-generated text differ from that of human-generated text across languages and domains?** By examining these differences across multiple emotional dimensions and categorical frameworks, the study seeks to establish whether AI-generated language exhibits a systematic emotional signature that distinguishes it from human writing.

II. LITERATURE REVIEW

Computational stylometric research shows that AI-generated text exhibits systematic tonal differences from human-authored writing (Kishnani, 2025). Psycholinguistic feature-extraction studies comparing the two suggest that AI-generated content tends to be more emotionally uniform, stereotyped, and less experientially grounded, whereas human writing typically embeds emotion within personal experiences, reflections, and contextual narratives rather than expressing it directly or formulaically (Opara, 2025). These findings suggest that emotional characteristics may represent an important dimension of the broader stylistic distinction between AI-generated and human-authored language.

Cross-linguistic and multi-model evidence indicates that such differences are not confined to a particular language, dataset, or generative model. A Portuguese-language comparison using LIWC and SAGE lexical analysis across GPT-4o, Mistral Large, and Llama 3.3 found that AI-generated text was consistently more formal, structured, positive, and motivational, while human-authored text showed greater variation in length and contained more negative emotion and personal references (Rodrigues et al., 2026). This suggests that AI-generated and human-authored texts differ systematically across multiple linguistic and psycholinguistic dimensions, providing a basis for examining whether similar differences extend to emotional characteristics specifically.

A growing subset of this literature has focused directly on sentiment. Several independent studies report that AI-generated text tends to be more emotionally neutral than human-written text (Liao et al., 2023), although evidence on the balance of positive versus negative sentiment among non-neutral text is less consistent—suggesting the distinction may be more robust at the level of overall neutrality than simple polarity. Detection-oriented research reinforces this, showing that sentiment and emotion-related features can serve as effective classifiers between AI-generated and human-written text (Mindner et al., 2023), with LLM outputs displaying greater stability in sentiment distributions under semantic-preserving transformations than the more variable patterns seen in human writing (Li et al., 2026).

Nevertheless, the literature does not establish a single, universally consistent emotional direction for AI-generated text, as reported differences in neutrality, positivity, and negativity vary across studies, models, and contexts. This matters because sentiment polarity captures only one dimension of affective

expression: two texts may share similar overall sentiment while differing substantially in valence, arousal, dominance, or the distribution of discrete emotions. A comprehensive assessment therefore requires complementary approaches capturing both continuous affective dimensions and discrete emotional states, rather than reliance on sentiment alone.

One plausible explanation for these systematic emotional differences lies in post-training alignment rather than core language-modelling objectives. Reinforcement learning from human feedback (RLHF) has been associated with softened, more agreeable responses to harsh or emotionally negative content (Bardol, 2025). If such processes contribute to greater neutrality, reduced negativity, or increased positivity, the magnitude and direction of emotional differences may vary with each model's alignment characteristics.

Direct comparisons between AI and human commentary provide further evidence of divergence. A study comparing human tweets about generative AI with ChatGPT's responses on the same topics found the AI responses more cohesive and systematically more positive (Pyreddy & Zaman, 2025). Conversely, research on LLM responses to mental-health queries found significant variation in emotional patterns across models and topics (VarastehNezhad et al., 2025), suggesting that while the AI-human emotional gap may be a broader phenomenon, it remains sensitive to the generating model and contextual domain.

Taken together, existing research establishes measurable linguistic, psycholinguistic, and sentiment-related differences between AI-generated and human-authored text, but does not yet clarify whether these extend consistently across multiple representations of emotion, given that most studies examine single models, languages, or domains with mixed directional findings. This gap motivates an approach combining sentiment analysis (broad polarity), VAD (continuous affective dimensions), and categorical emotion classification (discrete states such as joy, sadness, anger, fear, and surprise), where convergence across frameworks would offer stronger evidence of a systematic emotional distinction than any single measure alone. The present study addresses this gap through a systematic, multi-generator, multi-domain comparison using these three complementary frameworks, testing whether “AI-generated text exhibits a systematically different emotional profile from human-authored text across sentiment polarity, VAD dimensions, and categorical emotions.”

III. METHODOLOGY

A. Data Collection

To construct a comprehensive and linguistically diverse dataset, this study conducted an extensive search across public data repositories, including Kaggle and Zenodo, to identify high-quality datasets containing both AI-generated and human-authored text across as many languages and domains as possible, rather than relying on a single existing corpus. This multi-source strategy aimed to increase linguistic diversity, reduce dependence on the characteristics of any individual dataset, and provide a more robust basis for comparing the emotional profiles of AI-generated and human-authored text. Three

inclusion criteria guided screening: datasets had to contain text from more than one AI generator, rather than treating AI output as a single homogeneous category; they had to offer multilingual coverage spanning languages with differing levels of representation in typical LLM training data, avoiding a narrow focus on high-resource languages such as English, French, and German; and human- and AI-generated texts within each domain had to be collected under sufficiently comparable conditions to minimise confounding between authorship, language, topic, and collection procedures.

Following this screening process, four sources were selected for the final corpus: Multitude (Macko et al., 2023), Multisocial (Macko et al., 2025), ViDetect (Tran et al., 2024), and the M4 Benchmark (Wang et al., 2024). These were chosen because they collectively offered extensive multilingual coverage, contained both human- and AI-authored text, and provided relatively reliable language and authorship labels—criteria that few other candidate datasets satisfied simultaneously. Drawing on multiple independent sources, rather than a single corpus, allows the observed emotional differences to be examined across heterogeneous data, distinguishing this study from research based on one dataset alone.

The resulting raw corpus spans 25 languages, extending beyond the small set of high-resource languages that dominates much existing research on AI-generated text. Alongside widely represented languages such as English, Spanish, Arabic, and Chinese, it includes lower-resource languages such as Irish, Scottish Gaelic, Slovak, Slovenian, and Estonian, deliberately addressing the underrepresentation of smaller linguistic communities in existing AI-text resources. In total, the corpus contains 874,325 text samples, comprising 181,870 human-authored and 692,455 AI-generated texts. English is the largest contributor (190,992 samples), followed by Spanish (61,115), Portuguese (55,522), Dutch (50,949), German (47,187), and Russian (46,407), with substantial representation of Indonesian, Urdu, Vietnamese, and languages from the Baltic and Balkan regions. This scale and diversity provide a broader empirical foundation for investigating emotional differences than would be possible using a small number of high-resource languages alone.

Because the raw corpus is substantially larger for some languages and authorship categories than others, a balancing procedure was applied before the main analysis, limiting observations for each language-by-authorship combination to a maximum of 1,000 samples. Where sufficient data were available, up to 1,000 human-authored and 1,000 AI-generated texts were retained per language. This procedure was designed to prevent languages or data sources with larger original corpora from disproportionately influencing the analysis, providing a more equitable basis for cross-linguistic comparison.

B. Data Analysis

Consistent with the study's three outcome categories, this study employs distinct instruments to measure the sentiment, dimensional emotion, and categorical emotion properties of each text, rather than relying on a single classifier. This multi-instrument design follows the triangulation logic described

above: each instrument captures a different facet of emotional expression, and convergence across instruments is treated as stronger evidence than any single model's output in isolation.

Sentiment is measured using two complementary models: tabularisai/multilingual-sentiment-analysis, a Transformer-based multilingual classifier producing a five-class polarity judgment mapped to a signed numeric scale, chosen for its multilingual training to address the English-only limitation common in prior AI-emotion research; and VADER, a rule-based lexicon-and-heuristic analyser classifying text into positive, neutral, or negative categories via a compound polarity score, included as a methodologically distinct, non-Transformer cross-check, though its output for non-English text is treated as indicative rather than fully validated.

Table 1: Emotion Detection Models

No	Task	Model name
1	Text to sentiment scores	tabularisai/multilingual-sentiment-analysis
2	Text to sentiment categories	vaderSentiment
3	Text to VAD scores	hplisiecki/word2affect_english
4	Text to basic emotion categories	bhadresh-savani/bert-base-uncased-emotion
5	Text to complex emotion categories	fyaronskiy/ModernBERT-large-english-go-emotions

Dimensional emotion is measured using hplisiecki/word2affect_english, a Transformer-based regression model predicting valence, arousal, and dominance at the word level; since it was trained for word-level rather than document-level prediction, this study scores each unique word in the sampled corpus and averages word-level predictions per document, mirroring classical lexicon-based VAD aggregation but replacing static lexicon lookup with model-predicted values—an English-only design that produces the non-Latin-script scoring limitation discussed later, most visible in the Urdu subsample. Categorical emotion is measured at two levels of granularity: basic emotion categories via bhadresh-savani/bert-base-uncased-emotion, a BERT-based classifier trained on Ekman's six-class taxonomy, and advanced/complex emotion categories via fyaronskiy/ModernBERT-large-english-go-emotions, a ModernBERT-based classifier trained on the GoEmotions taxonomy, offering substantially finer-grained resolution across 28 emotion categories.

IV. RESULTS

A series of univariate tests of between-subjects effects was conducted to identify the individual affective dimensions contributing to the overall multivariate effect of authorship (human vs. AI). As reported in Table 1, all seven dependent variables—sentiment score, sentiment category, valence, arousal, dominance, basic emotion category, and complex emotion category—differed significantly between human- and AI-generated texts, with all effects significant at *p* < .001 except arousal, which remained

significant at *p* = .001. These findings indicate that the overall multivariate effect of authorship is not attributable to only one or two affective measures, but reflects systematic variation across almost the entire set of emotional dimensions examined.

Table 1
Univariate Test Results

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Human/AI	Sentiment S	8.444	1	8.444	144.133	.000
	Sentiment C	101.867	1	101.867	139.732	.000
	Valence	2.588	1	2.588	289.898	.000
	Arousal	.047	1	.047	10.668	.001
	Dominance	2.907	1	2.907	545.198	.000
	Basic Emotion	1024.689	1	1024.689	211.947	.000
	Complex Emotion	10559.873	1	10559.873	387.524	.000

The sentiment measures showed substantial and similarly sized effects—sentiment score, *F*(1) = 144.13, *p* < .001, and sentiment category, *F*(1) = 139.73, *p* < .001—indicating that the human–AI distinction holds whether sentiment is represented continuously or categorically. Valence produced a considerably larger effect, *F*(1) = 289.90, *p* < .001, suggesting that this VAD-based measure captures authorship-related differences in affective orientation more distinctly than sentiment alone.

Within the VAD dimensions, dominance showed the largest effect among the continuous measures, *F*(1) = 545.20, *p* < .001, indicating that perceived strength, control, or assertiveness differentiates human- and AI-generated text more strongly than valence or arousal. Arousal, by contrast, showed the smallest effect, *F*(1) = 10.67, *p* = .001; although statistically significant, this suggests emotional activation is relatively similar between the two groups compared with the divergence seen in valence and dominance. The categorical emotion measures also showed pronounced differences: basic emotion category yielded *F*(1) = 211.95, *p* < .001, while complex emotion category produced the largest F-statistic and Type III sum of squares in the analysis, *F*(1) = 387.52, *p* < .001. This suggests that fine-grained categorical classification is particularly sensitive to authorship differences, capturing shifts in specific emotion types rather than only overall polarity or intensity. Taken together, the univariate results show that the emotional profiles of AI-generated and human-authored text differ consistently across multiple affective representations, with the strongest divergence observed for complex emotion and dominance and the weakest for arousal.

Figure 1 provides a visual comparison of sentiment and VAD scores between human- and AI-generated texts. The compound sentiment score shows a clear directional difference: human-authored text has an average score of approximately − 0.022, whereas AI-generated text has an average of approximately

0.006. Although both values are close to the neutral point of the − 1 to + 1 scale, they fall on opposite sides of zero, indicating that human-authored text is mildly negative on average while AI-generated text is slightly positive—consistent with the broader statistical finding that authorship is associated with differences in emotional polarity. The VAD results reinforce this distinction: AI-generated text has a mean valence of 0.515, compared with approximately 0.500 for human-authored text. Although modest in magnitude, this difference is directionally consistent with the sentiment results, and the convergence of the two measures suggests that the greater positive orientation of AI-generated text is detectable using both sentiment classification and continuous affective norms.

Arousal shows the greatest similarity between the two groups, with human-generated text averaging approximately 0.428 and AI-generated text approximately 0.430. This near-overlap is consistent with the comparatively small univariate effect for arousal: although statistically significant, the substantive difference appears limited, suggesting that the principal emotional distinction between the two sources lies not in overall activation or intensity but in the direction of affective tone and perceived strength of expression. Dominance presents the clearest difference among the three VAD dimensions, with AI-generated text averaging approximately 0.550 compared with approximately 0.535 for human-generated text—corresponding to the largest F-statistic among the continuous affective measures. Taken together, the sentiment and VAD results indicate that AI-generated text is characterised by somewhat more positive valence and greater perceived dominance, while human- and AI-generated texts exhibit very similar levels of arousal.

Figure 2 provides a more detailed examination of the 28 GoEmotions categories, showing the relative proportion of human- and AI-generated texts associated with each emotion. The distribution reveals a systematic pattern rather than a random separation between the two authorship groups. Human-authored text shows the largest relative representation in amusement (approximately 57%), followed by anger (54%), neutral (53%), surprise (51%), disapproval (51%), annoyance (50%), and disgust (50%). Several of these represent reactive or evaluative responses toward an object, event, or situation, and their greater representation in human-authored text is consistent with the lower average compound sentiment observed for that group. In contrast, AI-generated text shows its strongest relative representation in caring (approximately 64%), remorse (63%), optimism (62%), admiration (62%), and joy (62%), with curiosity, love and desire also skewing AI (57–58%).

The middle of the distribution—comprising realization, confusion, grief, embarrassment, disappointment, fear, pride, relief, gratitude, approval, sadness, and nervousness—falls within a relatively narrow range of approximately 43–48% human share, showing substantially smaller differences between the two groups. Notably, several negatively valenced emotions, including grief, disappointment, fear, and sadness, occur within this middle range rather than showing the stronger human skew seen for anger, disgust, annoyance, and disapproval. This suggests that the human–AI difference cannot simply be characterised as AI-generated text avoiding negative emotion altogether. Instead, the strongest divergence involves more reactive, confrontational, or judgment-oriented negative emotions, while more vulnerable or internally oriented negative emotions are distributed more evenly between the two groups.

Overall, the categorical emotion results offer a more nuanced interpretation of the authorship effect than aggregate sentiment or VAD scores alone. Human-authored text is relatively more associated with reactive, critical, and evaluative emotions—anger, disgust, annoyance, disapproval, surprise, and amusement—while AI-generated text is relatively more associated with caring, admiration, optimism, joy, curiosity, and love. This pattern indicates that the emotional distinction between the two sources is not simply a general difference in positivity, but reflects differences in the types of emotional responses expressed: AI-generated text appears more strongly associated with supportive and encouraging registers, whereas human-authored text shows greater representation of evaluative emotional responses.

V. DISCUSSION AND CONCLUSIONS

This study extends theoretical accounts of generative AI beyond a purely cognitive framing by positioning emotional expression as a measurable and systematic property of language models in its own right (Opara, 2024; Rodrigues et al., 2026). Using four complementary approaches—sentiment scoring, rule-based sentiment categorisation, VAD analysis, and categorical emotion classification—this study examined whether AI-generated text displays a systematically different emotional profile from human-generated text, and whether these differences remain consistent across domains and languages. Drawing on a large-scale dataset spanning multiple models, domains, and languages, the findings support both hypotheses: across ten of the eleven metrics examined, AI-generated text showed significantly higher valence, dominance, and emotional neutrality, alongside significantly lower negative sentiment, sadness, fear, and surprise, than human-generated text, with significant multivariate separation overall. This convergent evidence indicates that the AI positivity/neutrality effect is a robust feature of contemporary LLMs, not attributable to any single dataset, domain, language, or measurement method, addressing a methodological limitation identified in earlier reviews. These findings suggest that generative AI is not merely an information-transmission tool: its outputs carry a measurable and systematic emotional signature distinct from human writing, rather than reflecting incidental variation.

The study also moves beyond description by linking these emotional patterns to reinforcement learning from human feedback (RLHF) and related alignment mechanisms discussed in the literature on sycophancy and affective alignment (AlKhamissi et al., 2024; Bai et al., 2022). This provides an empirically grounded explanation for why an AI–human emotional gap may emerge and why its magnitude and direction may differ across languages. Rooted in asymmetries in training data and alignment processes, this account distinguishes the present work from purely observational studies and offers a testable pathway for future research, such as comparing base models with instruction-tuned versions of the same model.

For organisations using AI in customer service, education, marketing, or mental-health-adjacent settings, these findings indicate that a system's emotional tone cannot be assumed to remain consistent across languages: validating a chatbot's tone in English, for example, may not ensure comparable performance in Urdu, or Vietnamese, where the AI–human emotional gap may range from a strong positive bias to

reversed emotional cues, creating a potential mismatch between intended and actual system behaviour. This issue is particularly important in mental health and education, which have been identified as emotionally sensitive contexts where inappropriate AI tone may carry heightened risks; in such settings, even small discrepancies between intended and expressed emotional tone may have disproportionate consequences. As generative AI becomes further embedded in human-facing settings, understanding how emotional tone is expressed—and how consistently it is maintained across domains, languages, and vulnerable-user contexts.

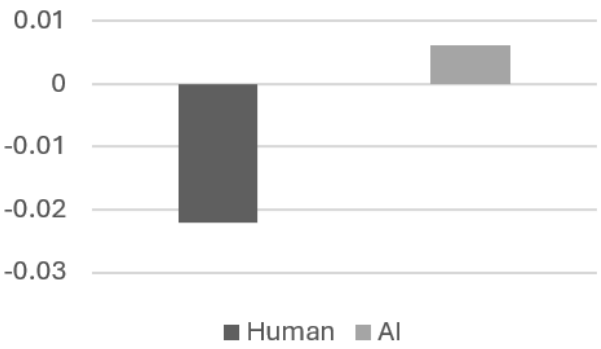
References

1. AlKhamissi B, ElNokrashy M, Alkhamissi M, Diab M (2024) Investigating cultural alignment of large language models. Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1)
2. Bai Y, Jones A, Ndousse K, Askell A, Chen A, DasSarma N, Drain D, Fort S, Ganguli D, Henighan T (2022) Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*
3. Bardol F (2025) ChatGPT Reads Your Tone and Responds Accordingly–Until It Does Not–Emotional Framing Induces Bias in LLM Outputs. *arXiv preprint arXiv:2507.21083*
4. Buscemi A, Lothritz C, Morales S, Gomez-Vazquez M, Clarisó R, Cabot J, Castignani G (2025) Mind the language gap: Automated and augmented evaluation of bias in llms for high-and low-resource languages. *arXiv preprint arXiv:2504.18560*
5. Demszky D, Movshovitz-Attias D, Ko J, Cowen A, Nemade G, Ravi S (2020) GoEmotions: A dataset of fine-grained emotions. *arXiv preprint arXiv:2005.00547*
6. Ekman P (1984) Expression and the nature of emotion. Approaches Emot 3(19):344
7. Foo LG, Rahmani H, Liu J (2025) Ai-generated content (aigc) for various data modalities: A survey. ACM-CSUR 57(9):1–66
8. Kishnani D (2025) The uncanny valley: An empirical study on human perceptions of AI-generated text and images. Massachusetts Institute of Technology]
9. Li S, Lin X, Li G, Liu Z, Wulianghai A, Ding L, Wu J, Li J (2026) Model-agnostic sentiment distribution stability analysis for robust LLM-generated texts detection. Proceedings of the AAAI Conference on AI
10. Liao W, Liu Z, Dai H, Xu S, Wu Z, Zhang Y, Huang X, Zhu D, Cai H, Li Q (2023) Differentiating ChatGPT-generated and human-written medical texts: quantitative study. JMIR Med Educ 9(1):e48904
11. Macko D, Kopal J, Moro R, Srba I (2025) MultiSocial: Multilingual benchmark of machine-generated text detection of social-media texts. Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)
12. Macko D, Moro R, Uchendu A, Lucas J, Yamashita M, Pikuliak M, Srba I, Le T, Lee D, Simko J (2023) MULTITuDE: Large-scale multilingual machine-generated text detection benchmark. Proceedings of

the 2023 Conference on Empirical Methods in Natural Language Processing

13. Mindner L, Schlippe T, Schaaff K (2023) Classification of human-and ai-generated texts: Investigating features for chatgpt. International conference on artificial intelligence in education technology
14. Moreno-Ortiz A, García-Gámez M (2026) Emotion in context: human and AI perspectives on the affective landscape of severance. Corpus Pragmat 10(1):15
15. Opara C (2024) StyloAI: Distinguishing AI-generated content with stylometric analysis. International conference on artificial intelligence in education
16. Opara C (2025) Distinguishing AI-generated and human-written text through psycholinguistic analysis. International Conference on Artificial Intelligence in Education
17. Pyreddy SR, Zaman TS Emoxpt: Analyzing emotional variances in human comments and llm-generated responses. 2025 IEEE 15th Annual Computing and Communication, Workshop (2025) and Conference (CCWC)
18. Rodrigues FA, Sturm NF, Pinheiro FL (2026) A linguistic comparison between human-and AI-generated content. Iscience, *29*(3)
19. Sui M (2025) Writing with emotion? Assessing emotional valence and appeals in AI-generated vs. human-written articles. AI Soc, 1–15
20. Tran Q-D, Nguyen V-Q, Pham Q-H, Nguyen K, Do T-H (2024) Vietnamese AI Generated Text Detection. *arXiv preprint arXiv:2405.03206*
21. VarastehNezhad A, Tavasoli R, Elyasi S, LotfiNia M, Farbeh H (2025) AI in Mental Health: Emotional and Sentiment Analysis of Large Language Models' Responses to Depression, Anxiety, and Stress Queries. *arXiv preprint arXiv:2508.11285*
22. Wang Y, Mansurov J, Ivanov P, Su J, Shelmanov A, Tsvigun A, Whitehouse C, Afzal OM, Mahmoud T, Sasaki T (2024) M4: Multi-generator, multi-domain, and multi-lingual black-box machine-generated text detection. Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1:)

Figures



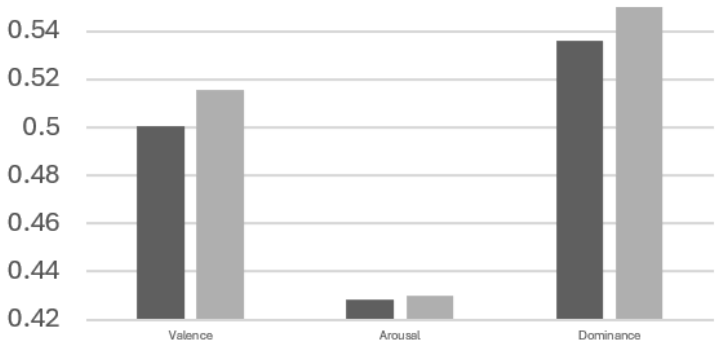


Figure 1

Differences in Sentiment and VAD Scores Between Human- and AI-Generated Texts

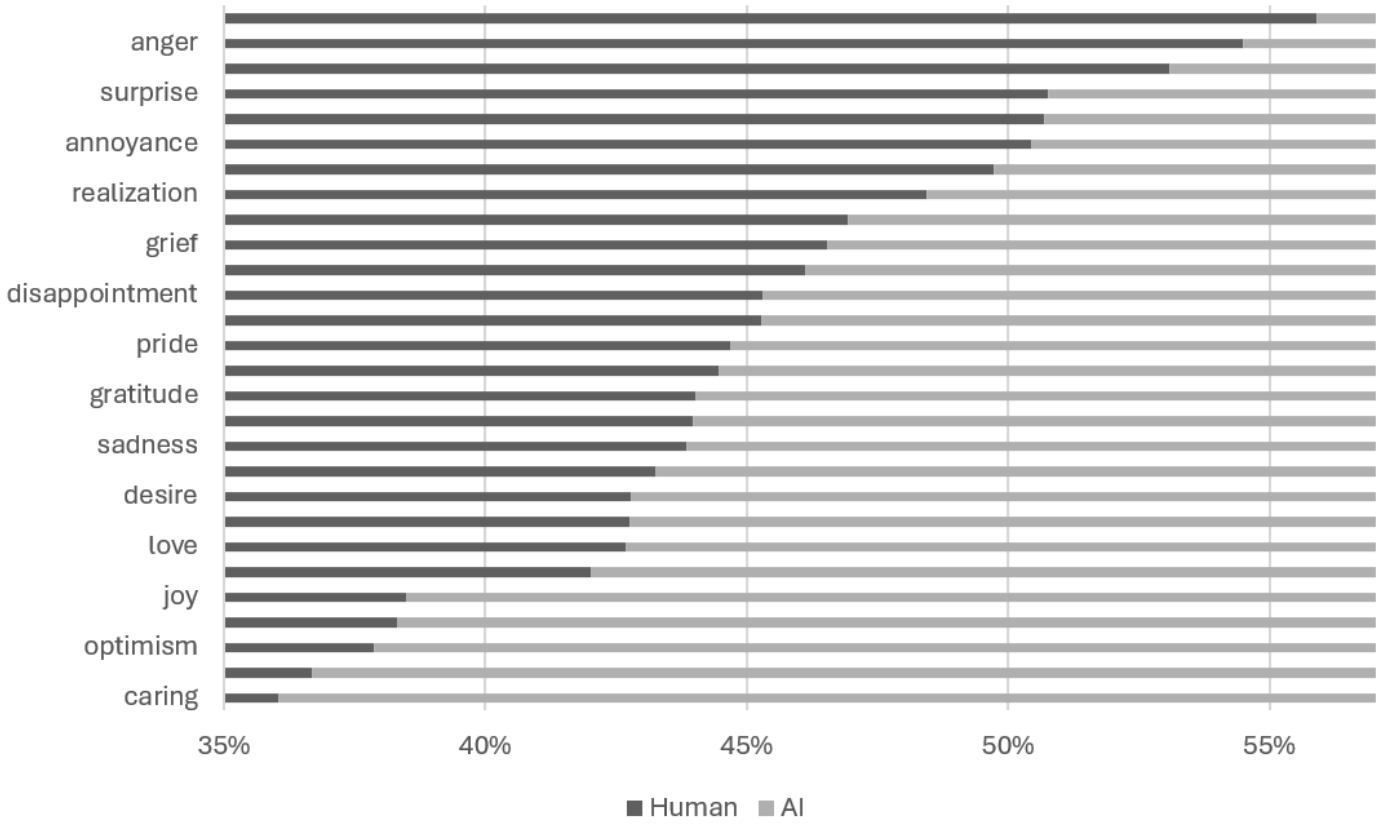


Figure 2

Differences in Emotional Intensities Between Human- and AI-Generated Text