ELSA: A STYLE-ALIGNED DATASET FOR EMOTIONALLY INTELLIGENT LANGUAGE GENERATION

Vishal Gandhi and Sagar Gandhi

Joyspace AI, WA, USA

ABSTRACT

Advancements in emotion-aware language processing increasingly shape vital NLP applications ranging from conversational AI and affective computing to computational psychology and creative content genera- tion. Existing emotion datasets either lack emotional granularity or fail to capture necessary stylistic diver- sity, limiting the advancement of effective emotion-conditioned text generation systems. Seeking to bridge this crucial gap between granularity and style diversity, this paper introduces a novel systematically con- structed dataset named ELSA (Emotion and Language Style Alignment Dataset)[1] leveraging fine-grained emotion taxonomies adapted from existing sources (dair-ai/emotion dataset and GoEmotions taxonomy). This dataset comprises multiple emotionally nuanced variations of original sentences regenerated across distinct contextual styles (conversational, formal, poetic, and narrative) using advanced Large Language Models (LLMs). Rigorous computational evaluation using metrics such as perplexity, embedding variance, readability, lexical diversity, and semantic coherence measures validates the dataset's emotional authentic- ity, linguistic fluency, and textual diversity. Comprehensive metric analyses affirm its potential to support deeper explorations into emotion-conditioned style-adaptive text generation. By enabling precision-tuned emotionally nuanced language modeling, our dataset creates fertile ground for research on fine-grained emotional control, prompt-driven explanation, interpretability, and style-adaptive expressive language gen- eration with LLMs.

KEYWORDS

Emotion-aware language modeling, fine-grained emotion recognition, stylistic variation, emotion-conditioned text generation, large language models (LLMs), text augmentation, emotion and style transfer, affective text generation, emotion-centric NLP, multistyle text synthesis, Natural Language Generation (NLG)

1. INTRODUCTION

Emotion-aware language processing is a cornerstone of human-computer interactions, shaping applications from affective computing and conversational AI to computational psychology and literary analysis [1, 2]. With the rapid advancement of Large Language Models (LLMs), recent research has significantly expanded the capabilities of automated emotional expression generation, providing unprecedented levels of expressive and affective control in generated text [3, 4]. However, effectively leveraging these advances still fundamentally depends on having access to high-quality datasets with sufficiently nuanced emotion annotations and economically diverse language expressions [5, 6].

[1] https://huggingface.co/datasets/joyspace-ai/ELSA-Emotion-and-Language-Style-Alignment-Dataset

David C. Wyld et al. (Eds): NATL, MLTEC, SIGEM, SEAPP, CSEA, FUZZY, DMDBS – 2025
pp. 01-10, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.152201

Currently, popular datasets such as the dair-ai/emotion collection cover broadly defined emotional categories including sadness, anger, love, surprise, fear, and joy [6]. While useful and widely embraced by the emotion recognition community, these emotion labels lack granularity, failing to represent wide variations in emotional intensity and complexity frequently encountered in realistic human language scenarios [3]. Conversely, although fine-grained datasets, such as Google's GoEmotions [5], encompass a richer and more nuanced taxonomy of 27 emotions, they focus primarily on short and informal conversational texts. Consequently, utilizing these datasets to train and evaluate LLM-based controlled emotion generation systems proves challenging [7], as direct transfer of nuanced conversational emotional annotations to diverse linguistic contexts (e.g., professional writing, poetry, storytelling) often results in unsuitable or unnatural outputs [8].

Moreover, emotions are inherently expressed differently across contexts and styles: the same emotional intent may translate with radically different lexical, syntactic, and semantic choices across poetic, conversational, formal, or narrative settings [9]. Prior work underscores that ignoring these contextual variations significantly limits the expressiveness and authenticity of computationally generated emotion-laden sentences [10]. Thus, a central research challenge emerges: existing emotion datasets either lack granularity or overlook stylistic contextual diversity, thereby restricting LLMs' potential to produce genuinely nuanced, contextually accurate emotional content.

Addressing this notable research gap, this paper introduces a systematically constructed ELSA dataset explicitly designed to enrich emotion-controlled text generation frameworks. Our proposed ELSA dataset effectively bridges the granularity of GoEmotion's fine-grained emotional taxonomy with the wide stylistic contextual diversity needed for realistic text generation. By employing mapping strategies linking dair-ai's coarse emotional categories [6] with GoEmotion's fine-grained categories [5], and through targeted prompt-based augmentation using advanced LLMs, we systematically generate multiple emotionally nuanced rewrites of textual samples across distinct stylistic expressions such as conversational, poetic, formal, and narrative.

Recognizing the importance of rigorous validation, we evaluate our ELSA dataset across established computational metrics, including perplexity for fluency, emotional distinctiveness through embedding variance, lexical diversity (distinct-n and self-BLEU), and style coherence measures [11]. This careful methodological process guarantees that generated texts exhibit emotional authenticity while preserving stylistic diversity. Thus, our ELSA dataset provides a robust foundation for future NLP research aimed at deeper exploration into emotion-conditioned text generation and style-adaptive affective language modeling.

In summary, the contributions of this paper include: (1) proposing an integrative ELSA dataset bridging existing coarse-grained emotion taxonomies with richer fine-grained emotional labels, (2) introducing nuanced stylistic contextual variation essential for realistic human textual emotion expressions, and (3) providing systematic empirical quality assessments designed to standardize future research benchmarks. Through enhancing emotional and stylistic complexity, the presented dataset promises to significantly advance the capacity of models to accurately replicate human emotional expression across varied communicative situations, thereby setting a valuable precedent for future research in affective computing and computational linguistics.

2. RELATED WORK

Emotion recognition and generation in natural language processing (NLP) have undergone significant advancements, primarily driven by the development of annotated emotional datasets and the evolution of generative models capable of producing emotionally nuanced text.

Early approaches to emotion recognition relied heavily on lexicon-based tools such as the Linguistic Inquiry and Word Count (LIWC) and WordNet-Affect, which extended the WordNet database to include affective concepts, facilitating emotion analysis in textual data [12]. While foundational, these lexicon-based methods were limited by manual creation constraints and often struggled to capture the nuanced, context-dependent expressions of emotion.

To address scalability challenges, researchers employed distant supervision techniques, leveraging social media platforms like Twitter to collect large-scale emotion-labeled data. For instance, Wang et al. [13] automatically created a dataset of approximately 2.5 million tweets by harnessing emotion-related hashtags, enabling broader emotion identification studies. However, these methods introduced substantial labeling noise and often focused on coarse emotional categories, such as those defined by Ekman's six basic emotions such as anger, disgust, fear, happiness, sadness, and surprise [14] or Plutchik's eight primary emotions.

Recent efforts have aimed to enhance the granularity of emotion taxonomies. Notably, Demszky et al. introduced GoEmotions, a dataset comprising 58,000 Reddit comments annotated with 27 distinct emotion categories [5]. This dataset provides a more nuanced understanding of emotional expression in text. However, its domain specificity, primarily encompassing informal, conversational-style text from Reddit, may limit its applicability across diverse writing genres.

In parallel, the field of text generation has seen significant progress in controlling stylistic and emotional attributes. Techniques such as controlled text generation have been employed to modulate the emotional tone of generated content. For example, Singh et al. proposed adapting language models to generate affect-driven and topic-focused sentences by incorporating emotion as a prior, allowing control over both the category and intensity of emotion in the generated text [15]. Similarly, Liu et al. introduced a method for modulating language models with emotions using a technique inspired by computer vision, enabling the generation of context-aware language that embodies diverse emotions [16]. Recent advances in LLM-based augmentation and emotion-controlled generation have opened new avenues for affect-aware text generation [15, 17]. Additionally, models have been developed to condition output on stylistic or emotional goals, bridging the gap between emotion recognition and controlled generation [18].

Contemporary models like EmoLLMs and others have begun integrating emotion representations more directly into language modeling [16]. Other studies focus on augmenting low-resource or complex affective phenomena, such as irony, through LLM-powered augmentation [19]. More broadly, text augmentation remains a core method to increase emotional data diversity and robustness [20]. Studies have also attempted to benchmark the emotional expressivity and comprehension of various LLMs [21].

Despite these advancements, a notable gap remains in datasets that simultaneously offer fine-grained emotion labels and encompass a broad range of stylistic contexts. Existing resources often lack the stylistic diversity necessary to train models capable of generating emotionally expressive text across various genres, such as formal writing, poetry, storytelling, and conversational language. Addressing this gap is crucial for developing models that can understand and generate text with appropriate emotional and stylistic nuances. Beyond text, emotion conditioning is also extending into other modalities like image generation, enabling broader multimodal affect-aware applications [22].

The present study seeks to bridge this gap by introducing ELSA (Emotion and Language Style Alignment) dataset that combines fine-grained emotion annotations with diverse stylistic contexts. By leveraging recent advances in large language model-based text augmentation techniques, we aim to enrich stylized emotional expression for diverse NLP applications, facilitating the devel-

opment of models capable of generating emotionally and stylistically nuanced text across various domains.

3. ELSA DATASET CREATION METHODOLOGY

To generate a nuanced dataset, an iterative pipeline leveraging advanced LLM-based augmentation was designed:

3.1.Initial Dataset Preparation

The initial data is derived from dair-ai/emotion dataset, containing 10,434 annotated text samples across six emotional classes (sadness, anger, love, surprise, fear, joy). First, each dair-ai emotion was mapped systematically to fine-grained GoEmotions subcategories, yielding higher granularity: as shown in the Table 1

Table 1: Primary emotions and their mapped emotions.

Primary Emotion	Mapped Emotions
Sadness	Sadness, Grief, Remorse
Anger	Anger, Annoyance, Disapproval, Embarrassment
Love	Love, Admiration, Caring
Surprise	Surprise, Realization
Fear	Fear, Nervousness
Joy	Joy, Excitement, Pride, Gratitude, Amusement

3.2.Emotion-conditioned Stylistic Augmentation

Each original textual instance was expanded through automated LLM-generated augmentation. The original text was fed into an OpenAI GPT o-1 conditioned explicitly with the mapped GoEmotions subcategories to create multiple emotionally varied, yet semantically genuine rewrites. For each emotionally labeled instance, the dataset includes systematically generated stylistic variations: formal, conversational, poetic, and narrative.

3.3. Quality Control and Validation

Generated texts were automatically evaluated with off-the-shelf classification models (e.g., GoEmotions classification model) to verify emotional and categorical accuracy. Text embeddings (Sentence-BERT embeddings) were utilized to ensure adequate semantic distance, ensuring novelty and diversity from original sources.

4. DATASET METRICS AND ANALYSIS

In this section, we present a comprehensive quantitative analysis of our ELSA dataset. To systematically evaluate the quality, emotional distinctiveness, semantic coherence, linguistic diversity, and readability of the generated texts, we employ multiple established metrics. Below, we formally define each metric, summarize their empirical statistics (Table 2), and discuss our results in detail thereafter.

4.1.Metric Definitions

Embedding Variance: Given sentence embedding vectors $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$ corresponding to n gen-

erated samples derived from a common base sentence, the embedding variance is defined as:

$$\sigma^2_{\text{emb}} = \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{s}_i - \bar{\mathbf{s}}\|^2 \quad (1)$$

where $\bar{\mathbf{s}}$ is the mean embedding vector:

$$\bar{\mathbf{s}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{s}_i \quad (2)$$

Average Emotion Distance: For a given emotion embedding representation $\mathbf{e}_{\text{orig}}$ of the original text and generated emotion embeddings $\mathbf{e}_i$, the average emotion distance is computed as the mean cosine distance:

$$D_{\text{emotion}} = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\mathbf{e}_{\text{orig}} \cdot \mathbf{e}_i}{\|\mathbf{e}_{\text{orig}}\|\|\mathbf{e}_i\|}\right) \quad (3)$$

Average Readability: The readability score is calculated using the Flesch–Kincaid readability test:

$$R = 206.835 - 1.015\left(\frac{\text{Total Words}}{\text{Total Sentences}}\right) - 84.6\left(\frac{\text{Total Syllables}}{\text{Total Words}}\right) \quad (4)$$

Distinct-n: Distinct-n measures lexical diversity as the proportion of unique n-grams over all n-grams in a corpus. For bigrams (n = 2), it is computed as:

$$\text{Distinct-2} = \frac{\text{Unique-bigrams-in-corpus}}{\text{Total-bigrams-in-corpus}} \quad (5)$$

Self-BLEU: Self-BLEU measures textual similarity within a set by calculating the average BLEU score of each sentence against all remaining generated sentences. A lower Self-BLEU implies greater textual diversity:

$$\text{Self-BLEU} = \frac{1}{N}\sum_{i=1}^{N} \text{BLEU}(s_i, \{s_j \mid j \neq i\}) \quad (6)$$

Average Perplexity: Given a probabilistic language model with probabilities $p(w_t \mid w_{<t})$ for tokens w_t in a sentence of length T, the perplexity is defined as:

$$PP = \exp\left(-\frac{1}{T}\sum_{t=1}^{T}\log p(w_t \mid w_{<t})\right) \quad (7)$$

Cosine Similarity: To assess semantic consistency between the original embeddings $\mathbf{s}_{\text{orig}}$ and the generated embeddings $\mathbf{s}_i$, we calculate their cosine similarity, averaged over all generated sentences:

$$\text{Cosine Similarity}_{\text{avg}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{s}_{\text{orig}} \cdot \mathbf{s}_i}{\|\mathbf{s}_{\text{orig}}\|\|\mathbf{s}_i\|} \quad (8)$$

4.2. Empirical Metrics Summary

Table 2: Statistical Summary of Various Metrics

	Mean	Median	Std	Min	Max
embedding_variance	0.000867	0.000869	0.000247	0.000179	0.001715
avg_emotion_distance	0.525050	0.527698	0.146301	0.029889	0.890717
avg_readability	57.671280	58.027500	10.606247	13.865000	92.630000
distinct_2	0.930821	0.940594	0.052273	0.601563	1.000000
self_bleu	0.036381	1.369605e-78	0.066719	8.548274e-232	0.525877
avg_perplexity	67.522142	60.261040	31.579396	17.430520	400.349832
cosine_sim_avg	0.525050	0.527698	0.146301	0.029889	0.890717

4.3. Discussion and Analysis

The computed metrics reveal several important insights regarding the ELSA dataset's emotional and stylistic complexity, diversity, readability, and semantic coherence.

Embedding Variance (mean=0.000867, std=0.000247): The low variance observed in sentence embeddings indicates highly stable semantic embedding vectors across stylistically diverse text versions derived from the same input base sentences. This stability aligns with expectations, as generated texts differ stylistically and emotionally while maintaining semantic coherence.

Average Emotion Distance (mean=0.525, std=0.146): The moderate mean emotional distance suggests that generated variations indeed shift significantly in emotion from the original samples. Thus, our generated ELSA dataset successfully exhibits moderate emotional diversity. However, the presence of high extreme values (max=0.89) implies occasional pronounced emotional shifts, warranting caution when very strict emotional stylistic retention is necessary.

Average Readability (mean=57.67, std=10.6): Overall readability scores provide evidence of good linguistic accessibility (Flesch-Kincaid readability scale). Nonetheless, the relatively high standard deviation and wide range of readability scores (minimum 13.86, maximum 92.63) reflect stylistic differences—for instance, markedly complex sentences common in poetic or formal narrative generation may reduce readability, suggesting a potential trade-off between stylistic richness and accessibility.

Distinct-2 (mean=0.9308, std=0.052): A mean Distinct-2 score approaching one underscores a high degree of lexical richness across the generated samples. Strong lexical diversity assures minimal repetitiveness, supporting the dataset's effectiveness for diverse generation applications that demand linguistic creativity and richness.

Self-BLEU (mean=0.036, std=0.067): The very low Self-BLEU values indicate notable textual novelty and minimal overlap among generated outputs. These low values are particularly advantageous for tasks that value output variety and originality, highlighting dataset quality in ensuring robust diversity.

Average Perplexity (mean=67.52, std=31.57): Moderate perplexity levels suggest the generated sentences generally maintain fluency and grammatical structures. However, the significant variations and the presence of high maximum values reflect complexities introduced by stylistic types. Particularly, poetic and highly expressive genres may lead to occasional lower predictability, an issue future research might address to optimize coherence and fluency further.

Cosine Similarity (mean=0.525, std=0.146): This value closely mirrors the average emotion distance metric, emphasizing moderate semantic retention relative to original texts while allowing sample room for stylistic and emotional variations.

In summary, these metrics collectively underscore our ELSA dataset's potential, showcasing distinctive emotional and stylistic variations, acceptable readability, high lexical richness and textual diversity, and stable semantic coherence. Such balanced metrics indicate its suitability for various nuanced NLP applications, ranging from emotion modeling and controlled generation to stylistic text generation tasks.

5. POTENTIAL RESEARCH USE-CASES

The proposed ELSA dataset opens several valuable avenues for downstream research in affective NLP, LLM fine-tuning, and emotion-conditioned generation. Below, we outline key areas where the ELSA dataset can significantly contribute:

1. **Fine-tuning for Emotionally Nuanced Stylistic Transfer.** Large language models currently excel in generic text generation but struggle with reliably transferring nuanced emotional semantics across diverse stylistic contexts. By fine-tuning these models with the proposed emotionally labeled dataset containing carefully mapped stylistic variations between conversational, poetic, formal, and narrative texts researchers can systematically train LLMs to precisely control emotional intensity, subtlety, and semantic appropriateness simultane- ously with effective style transfer [23]. This would significantly advance capacities currently limited in state-of-the-art LLM-driven style-adaptive generation systems.

2. **Precision-guided Prompt Engineering and Emotional Steering.** Controlling generated emotional outputs of large-scale generative models currently depends primarily on ad-hoc prompt engineering, leading to imprecise emotional or stylistic outcomes. With a comprehensive emotion-stylistic mapping, this dataset facilitates precise fine-tuning that supports systematic prompt-based control of emotional outputs. Consequently, it enables researchers and developers to create structured prompt methodologies such as moving beyond empirical trial-and-error prompt crafting,to rigorously guide emotional and style aspects of generated responses thus substantially advancing controllability and consistency levels currently unachievable through traditional prompting techniques alone [24].

3. **Context-aware Emotion Explanation and Interpretability.** Interpretability and explana- tion of how large language models internally encode subtle emotional distinctions remain open research challenges. The detailed emotional granularity and carefully mapped stylis- tic contextual variation embedded in the proposed dataset provide uniquely suitable train- ing data to fine-tune LLMs explicitly toward context-aware emotional explainability [25].

Through supervised and semi-supervised fine-tuning, researchers can equip LLMs to gener- ate explicit textual explanations concerning why particular linguistic structures evoke spe- cific emotions within a given context. Such improved interpretability capabilities would not only aid transparency and user-trust but also offer unprecedented insights into linguistic- emotional processing within NLP models, overcoming critical limitations of the current black-box nature of LLMs.

4. **Affective Computing and Sentiment Analysis.** Affective computing applications rely fundamentally on high-quality labeled datasets to detect subtle emotion signals within text. Leveraging both fine-grained and coarse-grained emotion labels organized across distinct stylistic contexts, the proposed dataset directly supports improved emotion recognition mod- els. Particularly, its granularity allows for enhanced detection and more precise emotion classification of implicit affective cues within diverse textual settings and contexts, applicable to social media sentiment detection, product review analysis, and textual affective computing in dialogues.

5. **Conversational AI and Chatbot Development.** The ability of conversational agents to generate plausible and emotionally expressive responses remains crucial for enhancing user engagement and communicative realism. By offering distinctly annotated emotional expressions across varying conversational styles such as formal communication, casual empathetic interactions, and narrative storytelling scenarios, the introduced dataset enables models to maintain stylistically coherent emotional responses. Thereby, it supports the development and improvement of emotionally adaptive dialogue systems capable of responding sensi- tively across diverse communicative scenarios such as therapeutic conversations, formal negotiations, or emotional social interactions [26].

6. **Creative and Literary Narrative Generation.** Emotional intensity and stylistically nuanced language underpin authentic literary and creative storytelling. Prior research stresses that computational narrative systems significantly benefit from data resources explicitly de- signed to capture emotional and stylistic diversity [27, 28]. This new dataset, featuring sys- tematic variations in emotion and stylistic contexts, provides essential resources enabling narrative generation models to achieve genuine emotional authenticity and literary quality. Positively impacting fields such as automated story writing, digital entertainment, and com- putational literary analysis, it helps model precise control of emotional affect in generated creative content.

7. **Psychological and Social Science Research.** Emotional variations expressed through different linguistic styles provide an important lens for studying human emotions in textual communications and other contextual interactions. Given its granularity in emotional annotations and its stylistically diverse text samples, the proposed dataset affords opportunities for computational psychologists and social scientists aiming to model and predict human emotional reactions across a spectrum of textual stimuli. It can further serve as an empirical basis for increased understanding of the relationships between linguistic style, emotion, and human psychological behaviors and perceptions.

8. **Comprehensive Evaluation Benchmarks.** Rigorous model benchmarking remains a fundamental objective in natural language processing. Reliable evaluation frameworks should incorporate metrics for theoretical linguistics assessment, style diversity, and emotional authenticity in generated text. The carefully controlled stylistic and emotional variations offered by this dataset establish ground for robust benchmarks, encompassing measures for fluency (e.g., perplexity), emotional distinctiveness, lexical diversity (e.g., distinct-n, self-BLEU), and contextual sensitivity. Thus, it helps in systematically evaluating model

performance, improving transparency, explainability, and practical effectiveness of future emotion-aware NLP systems.

Despite these contributions, careful attention needs to be paid regarding inherent biases within source datasets and possible semantic overlaps or unintended misclassifications during text generation processes. Therefore, we encourage continued research aimed at mitigating such limitations through explicit detection, evaluation, and controlled remediation strategies within downstream applications.

6. CONCLUSION

This study presents a rigorously constructed emotion-conditioned ELSA dataset, explicitly designed to address existing limitations in granular emotional modeling and contextual stylistic variations within NLP. By methodically integrating coarse-grained emotional categories with fine-grained emotional subcategories derived from established datasets (dair-ai and GoEmotions), we generate stylistically diversified emotional expressions via advanced LLM-driven augmentations. Comprehensive computational evaluation demonstrates promising levels of semantic stability, moderate emotional divergence, high linguistic diversity, readability, and fluency, highlighting the ELSA dataset's practical utility and scientific rigor.

The proposed ELSA dataset opens numerous research pathways conducive to fine-tuning LLMs for emotionally expressive, stylistic text generation tasks previously unattainable, such as nuanced stylistic transfer, controlled emotional prompting, context-driven explainability, and accurate emotional grounding within complex textual contexts. Given these clear benefits, we encourage the NLP community to leverage this resource for advancing emotional control and interpretability with large language models. Future research should focus on addressing inherent biases from source datasets and optimizing generations further for fluency and emotion-stylistic coherence, thereby continuously enhancing the robustness and applicability of emotion-aware NLP solutions.

ACKNOWLEDGEMENTS

The authors would like to thank the open-source community for providing the foundational datasets and tools that made this research possible.

REFERENCES

[1] Picard, R. W. (2003). "Affective Computing: Challenges", *International Journal of Human-Computer Studies*, Vol. 59, No. 1-2, pp. 55-64.

[2] Mohammad, S. M. (2022). "Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis", *Computational Linguistics*, Vol. 48, No. 2, pp. 239-278.

[3] Chen, H., Wang, B., Yang, K., and Song, Y. (2023). "ECCRG: A Emotion-and Content-Controllable Response Generation Model", In *Proceedings of the International Conference on Collaborative Computing*, Springer.

[4] Chen, Y., and Xiao, Y. (2024). "Recent Advancement of Emotion Cognition in Large Language Mod- els", *arXiv preprint arXiv:2409.13354*.

[5] Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., and Ravi, S. (2020). "GoEmo- tions: A Dataset of Fine-Grained Emotions", *arXiv preprint arXiv:2005.00547*.

[6] Saravia, E., Liu, H.-C. T., Huang, Y.-H., Wu, J., and Chen, Y.-S. (2018). "CARER: Contextualized Affect Representations for Emotion Recognition", In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 3687-3697.

[7] Chen, R., Wang, J., Yu, L.-C., and Zhang, X. (2023). "Decoupled Variational Autoencoder with Interac- tive Attention for Affective Text Generation", *Engineering Applications of Artificial*

Intelligence, Vol. 123, pp. 106447.

[8] Barnes, J., De Clercq, O., and Klinger, R. (2023). "Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis", *WASSA 2023, Association for Computational Linguistics*.

[9] Jurafsky, D., and Martin, J. H. (2000). "Lexicons for Sentiment, Affect, and Connotation", *Speech and Language Processing: An Introduction to NLP, Computational Linguistics, and Speech Recognition*.

[10] John, V., Mou, L., Bahuleyan, H., and Vechtomova, O. (2018). "Disentangled Representation Learning for Non-Parallel Text Style Transfer", *arXiv preprint arXiv:1808.04339*.

[11] Juola, P., Mikros, G. K., and Vinsick, S. (2019). "Correlations and Potential Cross-Linguistic Indicators of Writing Style", *Journal of Quantitative Linguistics*, Vol. 26, No. 2, pp. 146-171.

[12] Strapparava, C., and Valitutti, A. (2004). "WordNet Affect: An Affective Extension of WordNet", In *Proceedings of LREC*, vol. 4, no. 1083–1086, p. 40.

[13] Wang, W., Chen, L., Thirunarayan, K., and Sheth, A. P. (2012). "Harnessing Twitter "Big Data" for Automatic Emotion Identification", In *Proceedings of the 2012 International Conference on Privacy, Security, Risk and Trust, and the 2012 International Conference on Social Computing*, pp. 587-592, IEEE.

[14] Ekman, P. (1992). "An Argument for Basic Emotions", *Cognition and Emotion*, Vol. 6, No. 3-4, pp. 169-200.

[15] Singh, C., Askari, A., Caruana, R., and Gao, J. (2023). "Augmenting Interpretable Models with Large Language Models During Training", *Nature Communications*, Vol. 14, No. 1, pp. 7913.

[16] Liu, Z., Yang, K., Xie, Q., Zhang, T., and Ananiadou, S. (2024). "EMoLLMs: A Series of Emotional Large Language Models and Annotation Tools for Comprehensive Affective Analysis", *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5487-5496.

[17] Gandhi, V., and Gandhi, S. (2025). "Prompt Sentiment: The Catalyst for LLM Change", *arXiv preprint arXiv:2503.13510*.

[18] Zheng, C., Sabour, S., Wen, J., Zhang, Z., and Huang, M. (2022). "AugESC: Dialogue Augmentation with Large Language Models for Emotional Support Conversation", *arXiv preprint arXiv:2202.13047*.

[19] Lin, Y., Xia, Y., and Long, Y. (2024). "Augmenting Emotion Features in Irony Detection with Large Language Modeling", In *Workshop on Chinese Lexical Semantics*, pp. 196-206, Springer Nature Singapore.

[21] Mohammad, F., Khan, M., Marwat, S. N. K., Jan, N., Gohar, N., Bilal, M., and Al-Rasheed, A. (2023). "Text Augmentation-Based Model for Emotion Recognition Using Transformers", *Computers, Materials & Continua*, Vol. 76, No. 3, pp. 3523-3547.

[22] Klapach, N. (2024). "The Comparative Emotional Capabilities of Five Popular Large Language Models", *Critical Debates in Humanities, Science and Global Justice*, Vol. 2, No. 1.

[23] Zekun Li, Baolin Peng, Pengcheng He, Michel Galley, Jianfeng Gao, Xifeng Yan (2023). "Guiding Large Language Models via Directional Stimulus Prompting", *arXiv preprint arXiv:2302.11520*.

[24] Resendiz, Y. M., and Klinger, R. (2023). "Emotion-Conditioned Text Generation Through Automatic Prompt Optimization", *arXiv preprint arXiv:2308.04857*.

[25] Sun, T., Wang, C., Song, X., Feng, F., and Nie, L. (2022). "Response Generation by Jointly Modeling Personalized Linguistic Styles and Emotions", *ACM Transactions on Multimedia Computing, Communica- tions, and Applications (TOMM)*, Vol. 18, No. 2, pp. 1-20.

[26] Li, C., Wang, J., Zhang, Y., Zhu, K., Hou, W., Lian, J., Luo, F., Yang, Q., and Xie, X. (2023). "Large Language Models Understand and Can Be Enhanced by Emotional Stimuli", *arXiv preprint arXiv:2307.11760*.

[27] Lin, Q., Zhang, J., Ong, Y. S., and Zhang, M. (2024). "Make Me Happier: Evoking Emotions Through Image Diffusion Models", *arXiv preprint arXiv:2403.08255*.

[28] Wang, K., Jing, Z., Su, Y., and Han, Y. (2024). "Large Language Models on Fine-grained Emotion Detection Dataset with Data Augmentation and Transfer Learning", *arXiv preprint arXiv:2403.06108*.

©2025 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.