Special Feature

KPoEM: A Human-Annotated Dataset for Emotion Classification and RAG-Based Poetry Generation in Korean Modern Poetry

LIM Iro, JI Haein, and KIM Byungjun

The Review of Korean Studies Volume 29 Number 1 (June 2026): 161–206
doi: 10.25024/review.2026.29.1.006

Introduction

Poetry is widely regarded as one of the most expressive forms of literature, capable of capturing subtle nuances of human emotion. Unlike straightforward prose, however, poetic language often conveys feelings indirectly—through metaphor, imagery, and symbolic reference—requiring readers to infer meaning beyond the literal words. This richness of expression makes poetry evocative but also difficult to analyze, even for human readers, and more so for computational models.

Although recent advances in large language models (LLMs) have greatly improved emotion classification in general text, these models often falter when encountering metaphorically dense poetic language. This limitation reveals an interpretive gap, in which the statistical pattern recognition of current AI models fails to adequately capture the subtle emotional expressions embedded in literary texts. In particular, culturally embedded emotional concepts in Korean literature—such as *seoreoum* (sorrow) and *bijangham* (resolute)—further highlight the need for domain-specific resources.

To address this challenge, we developed KPoEM (Korean Poetry Emotion Mapping), the first expert-annotated dataset designed specifically for emotion analysis in modern Korean poetry. For the annotation process, five trained experts provided both line-level and work-level emotional labels, enabling a multi-layered analysis of poetic expression. This dataset, to our knowledge, is the first of its kind for Korean literature, and it serves as a crucial foundation for applying and evaluating AI models in this context. KPoEM is constructed from the works of five major modern Korean poets—Han Yong-un 한용운, Im Hwa 임화, Kim So-wol 김소월, Yi Sang 이상, and Yun Dong-ju 윤동주—whose

* This research was supported by the Academy of Korean Studies under Grant No. AKSR2025-RE04 (Development of Advanced Natural Language Processing and Large Language Model-Based Digital Korean Studies and Education Methodology, 2025). The authors would like to express their sincere gratitude to the AKS for its technical and financial support and to Sul Koo, Jonghoon Yun, and Song-yi Jung for their valuable contributions to data annotation.

A preliminary version of this work was made available as a preprint on arXiv (arXiv:2509.03932). This article has been substantially revised following peer review. Early versions and related findings of this study were presented at the 2nd Korea Conference for Modern Literature Researchers (Sungkyunkwan University, Seoul), the Digital Humanities and Social Sciences Korea Conference (James Joo-Jin Kim Center for Korean Studies, University of Pennsylvania, Philadelphia), and the 180th Regular Conference of the Society of Bangyo Language and Literature (Sungkyunkwan University, Seoul).

writings capture the diverse emotions and nuanced cultural sensibilities of early twentieth-century Korea (Kim and Cheon 2020; Seoul Shinmun 2007).

The purpose of this study is threefold: (1) to construct KPoEM, a specialized dataset for emotion-aware literary computing; (2) to develop and validate a poetry-specific emotion classification model through a sequential fine-tuning strategy that addresses the complexities of poetic language; and (3) to apply this framework to a Retrieval-Augmented Generation (RAG)-based poetry generation system that reflects culturally grounded Korean emotional sensibilities.

In summary, the key contributions of this paper are as follows:

1. **Introduction of KPoEM:** This study presents an expert-annotated dataset for modern Korean poetry, featuring 7,622 entries with dual-layered (line-level and work-level) emotional labels.

2. **Development of an Emotion Classification Model for Poetry:** This study develops a specialized emotion classification model optimized for the nuances of poetic language. Implementing a sequential fine-tuning strategy—beginning with KcELECTRA,[1] then fine-tuning on the KOTE (Korean Online That-gul Emotions) dataset (Jeon, Lee, and Kim 2024), and finally adapting to the specialized KPoEM dataset—is highly effective for capturing poetic nuances. This methodological approach resulted in a micro-F1 score of 0.60, representing a significant improvement over the 0.43 baseline and establishing a new performance benchmark for emotion analysis in modern Korean poetry.

3. **Bridging Analysis and Generation:** This study demonstrates that the KPoEM dataset can be effectively integrated as structured metadata into a vector database within a RAG-based framework. This approach allows for emotion-aware retrieval, enabling the system to produce creative outputs that reflect culturally grounded Korean emotional sensibilities.

Ultimately, this study bridges the gap between natural language processing

1 KcELECTRA is a pretrained model trained on Korean online news comments, built upon the ELECTRA architecture (Clark et al. 2020). For the development of the KPoEM emotion classification model, the 2022 version (KcELECTRA-base-v2022) was utilized. See the following repository, KcELECTRA, for details (https://github.com/Beomi/KcELECTRA).

and the humanities by integrating humanistic insight with computational methodology across dataset construction, model evaluation, and generative modeling. By extending the scope of NLP into figurative and culturally specific domains, our approach enables the systematic investigation of creative questions that have traditionally eluded quantification. Central to this endeavor is the preservation and revitalization of the essence of modern Korean poetry; by transforming subtle emotional nuances into structured data, we ensure that the distinctive sensibilities of Korean literary expression are not only protected but actively utilized in digital environments.

In this respect, this work goes beyond establishing a robust foundation for large-scale poetic analysis; it aims to provide a generative substrate for digital literary scholars, computational literary practitioners, and theorists of generative poetics. Furthermore, it seeks to offer crucial methodological scaffolding for researchers in cultural analytics and algorithmic criticism who navigate the intersection of human affect and machine intelligence. By empowering these diverse academic communities with structured emotional data, this study strives to actualize a synergy between humanistic inquiry and technological innovation—enabling the examination of literature from new methodological perspectives without compromising literature's essential contextual nuance.

Background

Emotion Classification in Text-Based Language Models and Literary Texts

Recent advancements in transformer-based LLMs such as BERT, RoBERTa, and GPT-3 have significantly improved emotion classification, consistently outperforming earlier methods (Acheampong, Chen, and Nunoo-Mensah 2020; Brown et al. 2020; Devlin et al. 2019; Liu et al. 2019). When fine-tuned on large, high-quality annotated datasets like GoEmotions, these models effectively capture subtle emotional nuances (Demszky et al. 2020). This confirms that integrating high-quality human annotations with pre-trained LLMs is the most effective approach for accurate text emotion recognition.

However, literary texts pose distinct challenges because metaphorical and stylistic complexities often obscure direct emotional cues. As a result, models trained primarily on literal, contemporary texts frequently misclassify figurative or affectively implicit expressions in literary works (Ji 2024; S and Mahalakshmi 2019; Sprugnoli and Redaelli 2024). To address this interpretive gap, recent scholarship has emphasized the necessity of domain adaptation strategies, including fine-tuning LLMs on corpora composed of literary texts (Li et al. 2021). Zhao, B. Wang, and Z. Wang (2024), for instance, demonstrated notable advancements in the interpretive capabilities of LLMs by fine-tuning them on a curated corpus of ancient Chinese poetry and incorporating literary features such as stylistic patterns and thematic diversity, underscoring the importance of contextual familiarity with literary forms and conventions.

Empirical studies show that general-purpose sentiment lexicons and binary classification are inadequate for literary analysis due to the affective complexity of works like Horace's *Odes* (Sprugnoli et al. 2023). These limitations necessitate a shift toward multi-class and multi-label frameworks (Ahmad et al. 2020) and techniques like continual pre-training on literary corpora to capture genre-specific expressions and subtle tones. Though LLMs provide a robust foundation for emotion classification in general-purpose texts, their application to literary materials requires methodological refinement. Integrating domain-specific annotated corpora, adopting multi-dimensional emotion taxonomies, and incorporating insights from literary theory significantly enhance model performance in this complex domain.

Emotion Datasets and Annotation Practices for Korean Language and Poetry

The development of Korean-language emotion datasets has seen notable progress in recent years, laying a crucial foundation for computational analysis of emotional expression in Korean texts. Early lexical resources (Park et al. 2018; Sohn et al. 2012), while providing useful polarity and intensity data, are insufficient for contemporary deep learning. For this reason, GoEmotions-Korean was developed through the translation and manual correction of the English-language GoEmotions dataset.[2] While this effort expanded the availability of Korean emotion-labeled data, scholars caution that directly

2 See the following repository, GoEmotions-Korean, for details (https://github.com/monologg/GoEmotions-Korean).

importing emotion taxonomies from English corpora may overlook culturally embedded emotions unique to Korean language and literature (Jeon, Lee, and Kim 2024).

To address such limitations, large-scale annotated corpora have emerged. The KOTE dataset represents one of the most comprehensive Korean emotion datasets, comprising 50,000 online comments with over 250,000 human-annotated labels across 44 emotion categories, including culturally specific emotions (Jeon, Lee, and Kim 2024). Derived through clustering of emotion terms in embedding spaces, KOTE captures nuanced emotional expressions reflective of Korean sociocultural contexts. Notably, KOTE has been extensively utilized to fine-tune transformer-based models such as KoBERT[3] and KcELECTRA, enabling them to recognize complex, multi-dimensional emotional states beyond simple polarity. However, as KOTE consists primarily of colloquial online comments, it remains inherently limited in capturing the highly refined, metaphorical, and aesthetically elevated emotional expressions characteristic of literary works. Similarly, Kang's (2024) taxonomy for classical novels addresses narrative-specific needs, yet a gap remains for the metaphorical complexity of poetry.

Despite these advances, poetry remains an underexplored domain because its figurative density, symbolic richness, and interpretive multiplicity introduce significant ambiguity for annotators. Prior research on poetic emotion annotation provides valuable methodological insights. For instance, international precedents like PO-EMO and PERC highlight the necessity of expert-led, multi-label annotation to manage the multifaceted nature of poetic effect (Haider et al. 2020; S and Mahalakshmi 2019). However, capturing these subtle nuances remains a significant challenge for computational models—a limitation that recent deep learning approaches have begun to address by effectively modeling long-term dependencies and contextual focus (Ahmad et al. 2020).

These limitations are particularly acute for early twentieth-century Korean poetry, whose historical depth and culturally distinct affective lexicons demand a domain-specific resource. Consequently, constructing an emotion dataset tailored to Korean poetry is vital to bridging the gap between computational linguistics and culturally nuanced AI development.

3 See the following repository, KoBERT, for details (https://github.com/SKTBrain/KoBERT).

Advancements in Computational Poetry Generation

Computational poetry generation integrates Natural Language Generation, Computational Creativity, and AI to produce creative, aesthetically pleasing text. Unlike early knowledge-intensive, symbolic systems like PoeTryMe that enforced rigid formal constraints (Oliveira 2012), contemporary approaches focus on balancing meaningful output with literary nuances (Oliveira 2017).

Recent advancements utilize pre-trained LLMs and sophisticated decoding. Fine-tuning models like GPT-2 or GPT-Neo on domain-specific corpora have proven superior to general-purpose LLMs in generating emotionally evocative and semantically coherent poems (Bena and Kalita 2019; Bhat et al. 2025; Lo, Ariss, and Kurz 2022). Notably, the TPPoet model achieves a balance between diversity and quality through dynamic temperature and Anti-LM decoding[4] (Panahandeh, Asemi, and Nourani 2023). However, evaluating generated poetry remains difficult: while automatic metrics offer baseline validation, the inherently subjective and figurative nature of poetic language makes human assessment of poetic quality indispensable (Panahandeh, Asemi, and Nourani 2023), and existing approaches continue to raise unresolved questions about aesthetic judgment and interpretive validity (Van Heerden and Bas 2024).

This trajectory suggests that domain-specific emotion resources are essential for building culturally grounded poetry generation models. Heflin (2020) argues that AI-generated literature is not a monolithic practice but rather a complex assemblage of human and machine labor, emphasizing the role of vector representations in making language legible to generative models. Because emotional expression in modern Korean poetry is highly metaphorical and symbolic, the vectorization of nuanced affective states becomes a prerequisite for effective computational modeling. Recent studies therefore highlight the need for emotion-aware representations that can support faithful modeling of complex literary affect beyond surface-level sentiment.

---

4 The Anti-LM is a contrastive decoding strategy designed to suppress prior bias in Large Language Models (LLMs), particularly the tendency to repeat source text or follow repetitive patterns in zero-shot settings. By penalizing the probabilities (logits) of a simple language model conditioned only on the source, this method encourages more diverse and instruction-aligned generation. For the technical formulation involving exponential decay, see Sia, DeLucia, and Duh 2024.

Methodology[5]

Overview of Dataset Construction[6]

This study constructed KPoEM (Korean Poetry Emotion Mapping), an emotion-annotated dataset for quantitative analysis and emotion classification modeling of modern Korean poetry. It was developed through a structured preprocessing and refinement workflow, detailed in Appendix A, including poet and text selection, text normalization, and emotion annotation. Five expert annotators assigned multiple emotion labels to each instance based on the 44 fine-grained emotion categories defined in KOTE. Emotion categories are ordered alphabetically by their Korean labels for neutrality and ease of reference (Table 1). KPoEM is structured for fine-tuning transformer-based models and supports both emotion-focused literary analysis and downstream NLP tasks. Examples of the finalized dataset are presented in Tables 2 and 3.[7]

Table 1. Emotion Categories (n = 44) Used for KPoEM (Based on KOTE[8])

Valence	Korean	Romanized	Interpretation
Negative	경악	*gyeongak*	shock
	공포/무서움	*gongpo/museoum*	fear
	귀찮음	*gwichaneum*	laziness
	당황/난처	*danghwang/nancheo*	embarrassment
	부끄러움	*bukkeureoum*	shame
	부담/안 내킴	*budam/an naekim*	reluctant
	불쌍함/연민	*bulssangham/yeonmin*	compassion
	불안/걱정	*buran/geokjeong*	anxiety
	불평/불만	*bulpyeong/bulman*	dissatisfaction
	슬픔	*seulpeum*	sadness

5 To ensure the reproducibility and extensibility of the research, all source code used in this study is available at the following link: https://github.com/AKS-DHLAB/KPoEM.

6 The KPoEM dataset constructed in this research is available at the following link: https://doi.org/10.57967/hf/6303.

7 The English translations provided in parentheses within the text, subtitle, and title columns in Table 2 and Table 3 were generated, using Gemini 3 Pro developed by Google, and subsequently reviewed and validated by the authors.

8 The valence classification and English interpretations of the emotion categories are adopted from the original KOTE (Korean Online That-gul Emotions) dataset, following Jeon, Lee, and Kim (2024). The dataset is publicly available at https://github.com/searle-j/KOTE.

Valence	Korean	Romanized	Interpretation
	서러움	*seoreoum*	sorrow
	안타까움/실망	*antakkaum/silmang*	disappointment
	어이없음	*eoieopseum*	preposterous
	역겨움/징그러움	*yeokgyeoum/jinggeureoum*	disgust
	의심/불신	*uisim/bulsin*	distrust
	짜증	*jjajeung*	irritation
	재미없음	*jaemieopseum*	boredom
	절망	*jeolmang*	despair
	죄책감	*joechaekgam*	guilt
	증오/혐오	*jeungo/hyeomo*	contempt
	지긋지긋	*jigeutjigeut*	fed up
	패배/자기혐오	*paebae/jagihyeomo*	gessepany
	한심함	*hansimham*	pathetic
	화남/분노	*hwanam/bunno*	anger
	힘듦/지침	*himdeum/jichim*	exhaustion
Positive	감동/감탄	*gamdong/gamtan*	admiration
	고마움	*gomaum*	gratitude
	기대감	*gidaegam*	expectancy
	기쁨	*gippeum*	joy
	뿌듯함	*ppudeutham*	pride
	신기함/관심	*singiham/gwansim*	interest
	아껴주는	*akkyeojuneun*	care
	안심/신뢰	*ansim/silloe*	relief
	존경	*jongyeong*	respect
	즐거움/신남	*jeulgeoum/sinnam*	excitement
	편안/쾌적	*pyeonan/kwaejeok*	comfort
	행복	*haengbok*	happiness
	환영/호의	*hwanyeong/houi*	welcome
	흐뭇함(귀여움/예쁨)	*heumutham(gwiyeoum/yeppeum)*	attracted
Neutral	깨달음	*kkaedareum*	realization
	놀람	*nollam*	surprise
	비장함	*bijangham*	resolute
	우쭐댐/무시함	*ujjuldaem/musiham*	arrogance
ETC.	없음	*eopseum*	NO EMOTION

Table 2. Examples from the KPoEM Line-Level Dataset

line_id	poem_id	Text	sub_title	title	poet	annotator_01	annotator_02	annotator_03	annotator_04	annotator_05
36	4	바람이 부는데 While the wind blows		바람이 불어 In the Blowing Wind	Yun Dong-ju	interest, NO EMOTION	interest, NO EMOTION, embarrassment, anxiety, resolute, sorrow	NO EMOTION	anxiety, reluctant, boredom	admiration, joy, surprise, comfort, welcome
6885	472	꽃이 보이지 않는다. No flowers are in sight.	절벽 (Precipice)	위독 Critical Condition	Yi Sang	despair, disappointment, fear, anxiety, embarrassment, reluctant, distrust	fear, embarrassment, anxiety, shock, sorrow, disappointment, despair, distrust, realization	embarrassment, disappointment	shock, fear, embarrassment, reluctant, sadness, despair	embarrassment, anxiety, disappointment

Table 3. Examples from the KPoEM Work-Level Dataset

seg_id	poem_id	text	sub_title	title	poetry_book	poet	annotator_01	annotator_02	annotator_03	annotator_04	annotator_05
6	6	계절이 지나가는 하늘에는 가을로 가득 차 있습니다. 나는 아무 걱정도 없이 가을 속의 별들을 다 헤일 듯합니다... 가슴 속에 하나 둘 새겨지는 별을 이제 다 못 헤는 것은 쉬이 아침이 오는 까닭이요, 내일 밤이 남은 까닭이요, 아직 나의 청춘이 다하지 않은 까닭입니다. 별 하나에 추억과 별 하나에 사랑과 별 하나에 쓸쓸함과 별 하나에 동경과 별 하나에 시와 별 하나에 어머니, 어머니 The sky where seasons pass by is filled with autumn. I feel as though I could count all the stars in the autumn air, without a single care...		별 헤는 밤 The Night Counting Stars	하늘과 바람과 별과 시 The Sky, the Wind, the Stars, and the Poem	Yun Dong-ju	attracted, admiration, care, sadness, joy, expectancy, realization, welcome, respect	admiration, expectancy, sorrow, sadness, resolute, care, attracted, disappointment, realization, respect	attracted, expectancy, disappointment, sorrow, sadness	admiration, joy, attracted, happiness, care, comfort	respect, admiration, interest, realization, sorrow, sadness, disappointment

seg_ id	poem_ id	text	sub_title	title	poetry_ book	poet	annotator_01	annotator_02	annotator_03	annotator_04	annotator_05
		The reason I cannot count all the stars being engraved one by one in my heart now is because morning comes too soon, because tomorrow night still remains, and because my youth is not yet spent. Memory for one star, Love for another, Loneliness for another, Longing for another, Poetry for another, And mother, mother for one star.									
7	6	어머님, 나는 별 하나에 아름다운 말 한 마디씩 불러 봅니다. 소학교 때 책상을 같이했던 아이들의 이름과, 패, 경, 옥 이런 이국 소녀들의 이름과, 벌써 아기 어머니 된 계집애들의 이름과, 가난한 이웃 사람들의 이름과, 비둘기, 강아지, 토끼, 노새, 노루, '프랑시스 잠', '라이너 마리아 릴케', 이런 시인의 이름을 불러 봅니다. 이네들은 너무나 멀리 있습니다. 별이 아스라이 멀 듯이, 어머님, 그리고 당신은 멀리 북간도에 계십니다 나는 무엇인지 그리워 이 많은 별빛이 내린 언덕 위에 내 이름자를 써 보고, 흙으로 덮어 버리었습니다. 딴은, 밤을 새워 우는 벌레는 부끄러운 이름을 슬퍼하는 까닭입니다. 그러나 겨울이 지나고 나의 별에도 봄이 오면 무덤 위에 파란 잔디가 피어나듯이 내 이름자 묻힌 언덕 위에도 자랑처럼 풀이 무성할 게외다.		별 헤는 밤 The Night Counting Stars	하늘과 바람과 별과 시 The Sky, the Wind, the Stars, and the Poem	Yun Dong-ju	care, sadness, shame, compassion, respect, admiration, gratitude, attracted, welcome	care, attracted, welcome, sorrow, sadness, realization, guilt, shame, compassion, expectancy	disappointment, sorrow, expectancy, sadness, realization	anxiety, sorrow, sadness, respect, welcome	sorrow, sadness, respect

seg_ id	poem_ id	text	sub_title	title	poetry_ book	poet	annotator_01	annotator_02	annotator_03	annotator_04	annotator_05
		Mother, I call out a beautiful word for every star. The names of the children I shared a desk with in elementary school; the names of foreign girls like Pae, Kyeong, and Ok; the names of girls who have already become mothers; the names of poor neighbors; and the names of poets like "Francis Jammes" and "Rainer Maria Rilke," along with pigeons, puppies, rabbits, mules, and roe deer. They are all so far away. As distantly far as the stars, Mother, and you are far away in Northern Kando. Longing for I know not what, I wrote my name upon this hill where so much starlight falls, and then I covered it over with dirt. Surely, the reason the insects chirp all through the night is because they grieve for their shameful names. But when winter passes and spring comes to my star as well, just as green grass sprouts upon a grave, Grass will grow thick like pride upon the hill where my name is buried.									

In this study, the KPoEM dataset was constructed in two forms: line-level and work-level. For the line-level dataset, poetry texts were cleaned and segmented into individual lines, all of which were randomly shuffled to allow annotators to focus on line-specific emotional expression without broader contextual influence. This design enables an experimental examination of whether individual lines can convey emotions independently of their surrounding context. The dataset was constructed from 483 poems following the metadata schema shown in Table 4.

To account for the contextual nature of poetic emotion, a work-level dataset was also created in which annotators read the full text of each poem in its entirety and assigned emotion labels with full contextual awareness. Each poem was treated as a single data instance; however, texts exceeding 512 characters were segmented into paragraphs and ordered sequentially according to the metadata schema in Table 5. The original line and stanza structures from Wikisource were preserved, and emotion annotation was performed on texts that retained these poetic formal structures, ensuring that the formal and rhythmic characteristics of the source poems were reflected in the annotation process.

Table 4. Line-Level Metadata Schema of KPoEM

Field Name	Description
line_id	Unique identifier for each line-level entry in the dataset
poem_id	Individual identifier assigned to each poem included in the dataset
text	Textual content of the individual poetic line
sub_title	Subtitle of an individual piece in a series (if applicable) (e.g., Pursuit)
title	Title of the poem (e.g., Azaleas)
poet	Name of the poem's author (e.g., Kim So-wol)
annotator_XX	Identifier for the annotator who annotated the given line (e.g., annotator_01)

Table 5. Work-Level Metadata Schema of KPoEM

Field Name	Description
seg_id	Unique identifier for each work-level entry in the dataset
poem_id	Individual identifier assigned to each poem included in the dataset
text	Full text of the poem
sub_title	Subtitle of an individual piece in a series (if applicable) (e.g., Sinking)
title	Title of the poem (e.g., In the Blowing Wind)
poetry_book	Title of the poetry collection in which the poem appears (e.g., Sky, Wind, Stars, and Poetry)
poet	Name of the poem's author (e.g., Yun Dong-ju)
annotator_XX	Identifier for the person who annotated the given work (e.g., annotator_05)

In both dataset types, a multi-label structure was adopted, allowing five annotators to assign up to ten emotion labels to each line (or work). The order of metadata fields was designed so that annotators would first encounter the poem text immediately after the ID field, followed by the title and author. For series-based poems (e.g., Yi Sang's *Widok*), a subtitle field was added to distinguish between individual pieces, such as *Chugu* (Pursuit) and *Chimmol* (Sinking).

Table 6 presents the number of poetic lines and poems per poet in KPoEM. In total, the dataset comprises 7,622 entries, consisting of 7,007 line-level segments and 615 work-level texts drawn from 483 representative works of the 1920s–1940s. An analysis of the emotion label distribution revealed the ten most frequently assigned categories (Table 7). Beyond its role as an analytical corpus, this dataset serves as the foundation for emotion-conditioned poetry generation experiments, enabling the modeling of nuanced affective transitions in modern Korean literature.

Table 6. Composition of the KPoEM Dataset by Poet

Category	Han Yong-un	Im Hwa	Kim So-wol	Yi Sang	Yun Dong-ju	Total
Line-level	1,198	2,163	2,071	464	1,111	7,007
Work-level	138	110	176	77	114	615
Subtotal						**7,622**
Number of Poems	**117**	**43**	**165**	**46**	**112**	**483**

Table 7. Top 10 Most Frequently Used Emotion Labels

Rank	Emotion Label	Frequency
1	anxiety	10,126
2	sadness	8,715
3	expectancy	8,399
4	disappointment	8,016
5	sorrow	7,423
6	interest	7,094
7	resolute	6,808
8	care	6,786
9	admiration	5,316
10	embarrassment	5,249

Table 8 presents statistics on inter-annotator agreement for emotion labels. Across the dataset, 99% of the texts show agreement by at least two annotators on one or more emotion labels, indicating a generally consistent labeling pattern across annotators, with no severe disagreements observed. This high level of agreement reflects not only the effectiveness of the annotation guidelines but also the shared cultural and linguistic interpretive context among annotators. In addition, texts labeled as NO EMOTION exhibit a balanced distribution (Table 9), acknowledging that not all poetic lines explicitly convey emotion. This distribution enhances the dataset's representativeness of real reading experiences and provides a foundation for training models to distinguish emotionally weak or absent expressions.

Table 8. Statistics of Inter-Annotator Agreement on Emotion Labels

Agreement					
at least x annotators agreed on at least one label	x=1	x=2	x=3	x=4	x=5
# of texts	7,622	7,613	7,052	4,659	1,725
(% to total)	100	99.88	92.52	61.13	22.63

Table 9. Statistics of Lines Labeled as NO EMOTION

Texts Labeled for NO EMOTION						
NO EMOTION	0	1	2	3	4	5
# of texts (% to total)	6,047 (79.34%)	897 (11.77%)	413 (5.42%)	208 (2.73%)	50 (0.66%)	7 (0.09%)

The KPoEM dataset was made publicly available via Zenodo and Hugging Face. During distribution, the shuffled line-level dataset was reordered according to the original “line_id” sequence. For the work-level dataset—which preserves the structural features of poems from Wikisource and was annotated in that form—the release version was standardized by removing newline characters and retaining only the continuous text. This adjustment was made to enhance dataset consistency and usability, as inconsistent newline symbols can affect tokenization and downstream comparisons.

Model Construction[9]

Training, validation, test (TVT) distribution

For model training and evaluation, we established a rigorous data partitioning protocol. The procedure began by handling the two distinct data formats—the poem work-level dataset and the poem line-level dataset—separately. First, each of the two datasets was independently split into training, validation, and test sets with an 8:1:1 distribution (Table 10). Following this initial split, the corresponding sets were merged: the training set from the work-level data was combined with the training set from the line-level data, and this process was repeated for the validation and test sets to yield the final, unified datasets.

9 The KPoEM emotion classification model constructed in this research is available at the following link: https://doi.org/10.57967/hf/6301.

Table 10. Distribution of Data across TVT Sets

	Train (80%)	Validation (10%)	Test (10%)	Total
The Number of Rows	6,096	763	763	7,622

Processing of multi-annotator data for fine-tuning

The process for handling multi-annotator data involves three main steps: aggregation, vectorization, and normalization.

1. **Label Aggregation:** For each data instance, all discrete labels assigned by the 5 annotators are collected and compiled into a comprehensive bag of labels, ensuring all annotations (annotator_01 to annotator_05) are preserved.
2. **Score Vectorization:** The aggregated labels are transformed into a numerical "score vector" (s). This vector's dimension is equal to the total number of possible emotions (L). Each entry in s represents the frequency (0 to 5) of a specific emotion from L across all annotators, thereby quantifying the level of consensus for each label.
3. **Normalization:** To account for varying levels of agreement across instances, the raw score vector is normalized on an instance-by-instance basis using Min-Max scaling. This rescales the scores to a continuous range between 0 and 1, emphasizing the relative importance of each emotion within that specific data instance.

The normalized score vectors were used as soft multi-label targets during fine-tuning, with the model trained via binary cross-entropy loss over the 44 emotion categories. The foundation model used for classification was KcELECTRA-base-v2022, initially fine-tuned on the KOTE dataset. An additional fine-tuning process was then performed using the KPoEM training data, with validation data employed for monitoring progress and mitigating overfitting.

Results

Model performance was evaluated using the test set (n = 763), applying a classification threshold of 0.30. The threshold of 0.30 indicates that, among the 44 emotion categories, an emotion label is regarded as correctly predicted when its score exceeds 0.30. This standard directly adopts the threshold proposed in previous research by Jeon, Lee, and Kim (2024).

Table 11 presents a comparative analysis of the performance of models trained on the KPoEM and KOTE datasets, respectively. For this research phase, we utilized Optuna (v4.6.0), a hyperparameter optimization framework, to identify the optimal values for key hyperparameters within a Python (v3.12.3) environment. The search was conducted for 3 epochs within the following ranges: a learning rate between 1e-6 and 5e-5, a batch size between 8 and 16, and a dropout rate between 0.1 and 0.5. The resulting emotion probability distributions were normalized using min-max scaling (min=0, max=1), and the final labels were derived by applying a classification threshold of 0.3, subsequent to the data preprocessing, which also involved a min-max scaling factor of 0.2.

Table 11. Performance Comparison of KPoEM Emotion Classification Models *Threshold = 0.3

Model	Accuracy	Precision_ micro	Precision_ macro	Recall_ micro	Recall_ macro	F1_ micro	F1_ macro	MCC
KcELECTRA (KOTE only)	0.77	0.49	0.46	0.38	0.33	0.43	0.34	0.29
KcELECTRA (KPoEM only)	0.79	0.53	0.43	0.66	0.50	0.59	0.45	0.45
KcELECTRA (KOTE → KPoEM)	**0.79**	**0.53**	**0.47**	**0.69**	**0.54**	**0.60**	**0.49**	**0.47**

The KcELECTRA model, when subjected to fine-tuning solely on the KOTE dataset,[10] exhibited relatively low performance, achieving an Accuracy of 0.77, a micro-averaged Recall of 0.38, and a macro-averaged F1-score of 0.34. This outcome suggests that the KOTE dataset, which primarily comprises emotion data from colloquial online language, has limitations in capturing the nuanced contextual emotions inherent in literary texts. In contrast, the model directly fine-tuned on our KPoEM dataset[11] recorded a superior performance with an Accuracy of 0.79 and a macro F1-score of 0.45. This result demonstrates that the KPoEM dataset provides an effective foundation for the task of emotion classification in Korean modern poetry.

10 The KcELECTRA model fine-tuned only on the KOTE dataset is available at the following link: https://huggingface.co/AKS-DHLAB/KcELECTRA_KOTEOnly.

11 The KcELECTRA model fine-tuned only on the KPoEM dataset is available at the following link: https://huggingface.co/AKS-DHLAB/KcELECTRA_KPoEMOnly.

Notably, the model employing a sequential fine-tuning approach—pre-training on the KOTE dataset before transfer learning on the KPoEM dataset—yielded the best performance across all metrics: an Accuracy of 0.79, a micro-averaged Recall of 0.69, a macro F1-score of 0.49, and an MCC of 0.47. This finding indicates that a sequential training strategy is highly effective for literary emotion classification, significantly improving overall classification performance—particularly recall (0.38 → 0.69)—as reflected by the F1-score. This shift toward higher recall reflects the model's enhanced sensitivity to the multi-layered affective expressions characteristic of poetic language. More broadly, this implies that for domains with limited data, such as literary emotion analysis, integrating a broader, general-domain emotion dataset offers a productive strategy for improving model performance. These results were obtained using hyperparameters optimized via Optuna.

To further examine model behavior qualitatively, emotion classification was applied to a selection of representative modern and contemporary Korean poems. The sample comprises works by eminent poets, notably Han Kang 한강, laureate of the 2024 Nobel Prize in Literature, and Jeong Ji-yong 정지용, a canonical figure of modern Korean poetry. Comparative analysis revealed distinct tendencies between the three models.

Table 12 presents the results of a qualitative evaluation conducted by applying the models trained on the KPoEM and KOTE datasets to actual poetic texts. For this purpose, Han Kang's (2013) "Hyoege. 2002. Gyeoul" (To Hyo: Winter 2002) and Jeong Ji-yong's "Hyangsu" (Nostalgia)[12] were analyzed as case studies.[13]

In a case study of Han Kang's poem, the KOTE-only model struggled with poetic nuances, predicting less contextually aligned emotions such as "despair." In contrast, the KPoEM-only model showed improved alignment with "realization" and "sorrow," though with residual noise. The transfer learning model (KOTE → KPoEM) achieved the most stable distribution, identifying deeper affirmative tones (e.g., relief or welcome) beyond the surface-level sorrow.

In a subsequent case study involving Jeong Ji-yong's "Hyangsu," the

12 For the original Korean text of Jeong Ji-yong's "Hyangsu," we referred to Wikisource (https://ko.wikisource.org/wiki/향수/향수).

13 The English translations provided in parentheses within the Text and Title columns in Table 12 were generated using Gemini 3 Pro and subsequently reviewed and validated by the authors.

Table 12. Comparative Emotion Classification Results for Poems by Han Kang and Jeong Ji-yong

Text	Title	Poet	KcELECTRA (KOTE only)	KcELECTRA (KPoEM only)	KcELECTRA (KOTE → KPoEM)
하지만 곧 너도 알게 되겠지 내가 할 수 있는 일은 기억하는 일뿐이란 걸 저 번쩍이는 거대한 흐름과 시간과 성장(成長), 집요하게 사라지고 새로 태어나는 것들 앞에 우리가 함께 있었다는 걸 But soon You too will come to realize That the only thing I can do Is to remember That flashing, immense flow, and Time, and Growth, Before things that relentlessly vanish And are born anew That we were together	효에게. 2002. 겨울 To Hyo: Winter 2002	Han Kang	realization: 0.80 expectancy: 0.58 resolute: 0.55 disappointment: 0.51 sadness: 0.48 anxiety: 0.45 NO EMOTION: 0.42 exhaustion: 0.39 despair: 0.32	resolute: 0.89 realization: 0.86 sorrow: 0.77 anxiety: 0.72 disappointment: 0.69 sadness: 0.67 expectancy: 0.63 exhaustion: 0.42 admiration: 0.42 dissatisfaction: 0.33 distrust: 0.33 reluctant: 0.32	realization: 0.93 resolute: 0.91 expectancy: 0.75 anxiety: 0.50 admiration: 0.48 sadness: 0.45 **relief: 0.44** disappointment: 0.43 sorrow: 0.39 joy: 0.39 **welcome: 0.36** **care: 0.35** **pride: 0.32**
흙에서 자란 내 마음 파아란 하늘 빛이 그립어 함부로 쏜 화살을 찾으려 풀섶 이슬에 함추름 휘적시든 곳, — 그 곳이 참하 꿈엔들 잊힐 리야. 전설바다에 춤추는 밤물결 같은 검은 귀밑머리 날리는 어린 누이와 아무렇지도 않고 여쁠 것도 없는 사철 발벗은 안해가 따가운 해ㅅ살을 등에 지고 이삭 줏던 곳, — 그 곳이 참하 꿈엔들 잊힐 리야. My mind, raised from the soil Longing for the blue sky light To find the arrow I shot at random Where I was drenched in the dew of the grass thickets, — Could that place ever be forgotten, even in dreams? Like the night waves dancing in the sea of legend With my young sister, her black side-locks flying And my wife, neither extraordinary nor pretty, Barefoot throughout the four seasons Gleaning ears of grain, the stinging sunlight on her back — Could that place ever be forgotten, even in dreams?	향수 Nostalgia	Jeong Ji-yong	sadness: 0.91 disappointment: 0.72 compassion: 0.66 despair: 0.56 exhaustion: 0.54 sorrow: 0.49 anxiety: 0.42 realization: 0.35 NO EMOTION: 0.31	disappointment: 0.94 sorrow: 0.94 sadness: 0.94 anxiety: 0.85 care: 0.79 compassion: 0.78 exhaustion: 0.68 realization: 0.63 expectancy: 0.59 interest: 0.51 admiration: 0.40 embarrassment: 0.38 reluctant: 0.34 dissatisfaction: 0.33	sadness: 0.97 **sorrow: 0.94** disappointment: 0.90 **compassion: 0.75** anxiety: 0.68 care: 0.64 exhaustion: 0.63 expectancy: 0.42 **despair: 0.34** **realization: 0.32** reluctant: 0.31

KOTE-only model exhibited a diffuse emotional distribution, overemphasizing less contextually appropriate emotions such as despair and exhaustion. In contrast, the KPoEM-only model showed improved alignment with core emotions of sorrow and compassion, though with residual noise due to uneven emotional concentration. The transfer learning model (KOTE → KPoEM) achieved the most stable emotional hierarchy; it effectively suppressed excessive negative affect while simultaneously capturing both the poem's wistful tone and a profound affection for the subject. These findings indicate that general-domain pretraining followed by domain-specific refinement achieves the best balance between emotional intensity and semantic alignment in literary emotion classification. We refer to this optimized transfer model as the KPoEM emotion classification model (the KPoEM model).

Emotion-Aware Generative Poetics with KPoEM

Overview: *Emotion-Aware Poetry Generation Framework*

This section presents an emotion-aware poetry generation framework based on Retrieval-Augmented Generation (RAG), which integrates semantic similarity

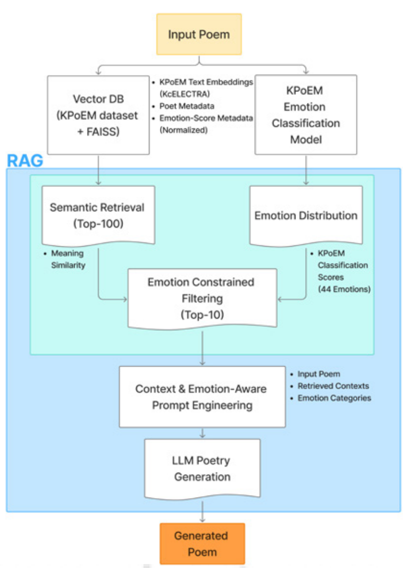

Figure 1. Emotion-Aware RAG-Based Poetry Generation Architecture

with emotion classification using the KPoEM dataset. The system generates poems by retrieving semantically relevant and emotionally aligned poetic lines and using them as contextual input for a large language model (LLM). As illustrated in Figure 1, the framework consists of emotion classification, emotion-constrained filtering, and LLM-based generation. The following sections provide step-by-step examples demonstrating how this pipeline operates in practice.[14]

Poetry Generation Experiments and Results

Construction of the vector database

This study constructs a vector database from the KPoEM dataset to support RAG-based poetry generation. In this framework, the vector database acts as a specialized external memory that provides the LLM with domain-specific knowledge of modern Korean poetic sensibilities (Jing et al. 2025). The dataset is processed at both the line and work levels; each poetic unit is embedded using the KcELECTRA model, which was also employed as the backbone of the KPoEM emotion classification model, ensuring consistency between semantic representation and emotion interpretation. Each instance is associated with normalized emotion-score metadata derived from annotations by five expert annotators. Emotion scores are min–max normalized to the 0–1 range, and only categories with values of 0.2 or higher are retained, following the same thresholding strategy used in the KOTE schema and the KPoEM model. This procedure ensures that the stored metadata reflects the relative emotional salience of both individual lines and entire poetic works. Through this design, the retrieval module selects relevant poetic units based on two criteria: (1) semantic similarity to the input text and (2) emotional alignment within the 44-category KOTE taxonomy.

The resulting vector database is indexed using FAISS (Facebook AI Similarity Search, v1.13.2) and integrated into a LangChain (v1.2.0) pipeline, enabling retrieved poetic units to be incorporated directly as contextual input during poem generation. The implementation relies on langchain-core (v1.2.5), langchain-community (v0.4.1), and langchain-huggingface (v1.2.0).

14 The English translations of the input poems and the generated poetic outputs presented in this section were produced using Gemini 3 Pro and subsequently reviewed and validated by the authors.

Pipeline design and execution

(1) Emotion classification of the input poem

The provided input poem is first classified by the KPoEM emotion classification model. This model generates probability-based scores across 44 fine-grained emotion categories (Table 1). These scores serve as reference values for computing the emotional alignment between the input poem and the poetic instances stored in the vector database during the subsequent retrieval stage.

Table 13 presents an example of emotion scores produced by the KPoEM classifier when an excerpt from Kim Chunsu's (2004) poem, "Kkot" (Flower), is used as the input text. Kim Chunsu's poetry is a notable case for this study, given that his concept of i*ngong-sihak* (artificial poetics) has been identified as forming a conceptual precursor of Korean AI generative poetics (Jeong 2025). As shown in the table, the input text exhibits high scores for emotion categories such as "care" (0.90), "realization" (0.87), and "admiration" (0.86).

The proposed poetry generation model leverages these classification results by using the input poem as a source of poetic imagery and thematic cues, while relying on the identified emotion categories to determine the affective tone of the generated poem, thereby aligning generation with both the input text and the emotional structures encoded in the KPoEM dataset.

Table 13. Example of KPoEM Emotion Classification Results for the Poem by Kim Chunsu

Original Input Poem	KPoEM Emotion Classification Results
내가 그의 이름을 불러주기 전에는 그는 다만 하나의 몸짓에 지나지 않았다. 내가 그의 이름을 불러 주었을 때 그는 나에게로 와서 꽃이 되었다. Before I called his name He was nothing But a mere gesture. When I called his name He came to me And became a flower.	care: 0.90 realization: 0.87 admiration: 0.86 joy: 0.79 expectancy: 0.78 resolute: 0.78 welcome: 0.64 gratitude: 0.64 sadness: 0.63 happiness: 0.62 respect: 0.60 relief: 0.60 sorrow: 0.59 pride: 0.54 disappointment: 0.52 attracted: 0.47 interest: 0.32 anxiety: 0.30

(2) Semantic retrieval and emotion-based filtering

To mitigate the performance degradation caused by irrelevant context in RAG systems (Leto et al. 2024), we introduce a two-stage filtering strategy. The embedding vector of the input poem is compared against the poetic unit vectors stored in the KPoEM vector database. Based on cosine similarity, the system first retrieves the top 100 semantically similar poetic instances—a candidate pool size determined through empirical experimentation to ensure sufficient thematic diversity. It then computes affective similarity by comparing the emotion scores of the input text with the emotion metadata attached to each retrieved instance. Finally, the system selects the top 10 poetic instances that satisfy both semantic similarity and emotional alignment. By limiting the context to the top 10 emotionally aligned instances, we aim to maximize generation quality while maintaining emotional coherence. This design aligns with empirical findings that optimal RAG performance is frequently observed with a context size of either 10 or 20 documents (Leto et al. 2024).

Table 14. Examples of Poetic Instances Retrieved from the KPoEM Vector Database Based on Semantic and Affective Similarity to “Kkot” by Kim Chunsu

Retrieved Text	Emotion Scores (Top 3)	Poet
혀끝에서 물결이 솟고 붓 아래에 꽃이 피어요. Waves rise from the tip of the tongue, and flowers bloom beneath the brush.	admiration: 0.8 attracted: 0.8 joy: 0.8	Han Yong-un
인간(人間)에 이 세상에 다시 있으랴. Could such a person ever be found in this world again?	resolute: 0.6 realization: 0.6 admiration: 0.4	Kim So-wol
…		
홈싹홈싹 숨치우는 보드라운 모래 바닥과 같은 긴 길이, 항상 외롭고 힘없는 저의 발길을 그리운 당신한테로 인도하여 주겠지요. A long path, like a soft bed of sand breathing with gentle gasps, will surely lead my always lonely and feeble footsteps toward you, for whom I long.	sadness: 0.8 expectancy: 0.8 gratitude: 0.6	Kim So-wol

Table 14 presents three examples from the top ten poetic instances selected from the vector database when Kim Chunsu’s poem is used as the input text. As shown in the table, the retrieved instances are selected based not only on semantic similarity but also on emotional alignment.

(3) Prompt engineering and poetry generation

The final set of ten retrieved poetic units is provided as contextual input to the generative language model (LLM). In this study, Midm-2.0-Base-Instruct,[15] an open-source Korean LLM, was employed, and the retrieved instances and emotion information were dynamically injected into the prompt using the LangChain framework. The generation pipeline utilized the following hyperparameters, determined through empirical experimentation to optimize poetic output: temperature=0.7, top_p=0.9,

max_new_tokens=128, and repetition_penalty=1.2.

Under this configuration, the model generates emotionally coherent poetic text by jointly referencing the retrieved units and the emotion scores classified by the KPoEM model from the input text. The prompt instruction used in this generation process is provided in Appendix B. The following tables present the poetry generated from the input poetic texts, with semantically similar expressions shared between the original sources and the generated outputs highlighted in bold to illustrate content relevance.

Table 15. Input Poem (Kim Chunsu's "Kkot") and the Resulting Generated Poem

Input Poem	Generated Poem
내가 그의 이름을 불러주기 전에는 그는 다만 하나의 몸짓에 지나지 않았다 내가 그의 이름을 불러 주었을 때 그는 나에게로 와서 **꽃이 되었다** **Before I called his name** He was nothing more But a mere gesture When I called his name He came to me **And became a flower**	<이름 부르는 순간> **나 그대를 부르기 이전엔** **그저 스쳐가는 바람결이었네** **수많은 얼굴 속 익명인 채** **존재했지만 존재하지 않던** 그러나 한 글자 한 글자 마음 실어 당신의 이름을 부르니 어둠 속 별처럼 빛나기 시작하네 마침내 내게 다가와 **영혼의 꽃송이 되어 피어오르네** 아껴주던 모든 것들이 이제 비로소 진짜 모습 드러내고 그리움 깊었던 시간들이 환영하는 눈빛 되어 돌아오네 깊은 밤 홀로 서 있을 때조차 누군가 알아봐 준다는 안도감 그것은 곧 삶의 의미이며 세상을 살아갈 이유였구나 <The Moment I Call Your Name> **Before I called your name** **You were merely a passing breeze**

15 See the following repository, K-intelligence, for details (https://huggingface.co/K-intelligence/Midm-2.0-Base-Instruct/discussions).

	Anonymous amidst countless faces **Existing, yet not truly existing** But as I called your name With each syllable laden with my heart You began to shine like a star in the darkness Finally, you come to me **And bloom as a flower of the soul** All the things I cherished Now finally reveal their true forms And the times of deep longing Return as eyes filled with welcome Even when standing alone in the deep night The relief that someone recognizes me— That was the very meaning of life And the reason to live in this world

Table 15 presents the poem generated by the proposed RAG-based model, titled "Ireum bureuneun sungan" (The Moment I Call Your Name). The generated poem demonstrates high semantic coherence by effectively inheriting the input's core metaphor, in which the act of naming defines the existence of the flower. This close thematic and metaphorical correspondence illustrates how the generation model preserves the original emotional tone of the input.

Furthermore, this study extended the generation experiments to include other representative works of Korean poetry: "Sarang" (Love) by Kim Su-young (1974) and "Nalgae hwansangtong" (Phantom Pain Wings) by Kim Hyesoon (2019), the latter of whom recently won the 2024 National Book Critics Circle Award for Poetry. The resulting generated poems, which demonstrate the model's adaptability to diverse poetic styles and historical contexts, are presented in Tables 16 and 17.

Table 16. Input Poem (Kim Su-young's "Sarang") and the Resulting Generated Poems

Input Poem	Generated Poem
어둠 속에서도 불빛 속에서도 변치 않는 사랑을 배웠다 너로 해서 그러나 너의 얼굴은 **어둠에서 불빛으로 넘어가는** **그 찰나에 꺼졌다 살아났다** **너의 얼굴은 그만큼 불안하다** Unchanging in the dark and in the light, I learned love through you But your face, **passing from darkness to light,**	<밤길의 약속> 밤하늘 별들이 하나둘 떨어져 그대와 나누었던 그 순간처럼 **어둠과 빛 사이 경계에서** 얼굴마저 흐릿해지네 **불안한 눈동자 안에 담긴** 깊은 한숨 같은 그리움이여 서로 다른 길 위에 서서 보이지 않아도 느낄 수 있는

flickered out and came alive in that fleeting moment **Your face is that precarious**	변하지 않을 마음 하나 불빛 깜빡이며 사라지는 그 짧은 찰나가 주는 아픔을 그래도 견뎌내는 것이 우리였기에 별똥별 지나며 흩어진 조각마다 다시 만나자던 그 약속 잊혀진 듯하지만 잊혀지지 않은 그리움이란 이름으로 남아있어 <Promise on a Night Road> Stars in the night sky fall one by one, Just like that moment I shared with you. **At the boundary between darkness and light**, Even your face grows faint Oh, this longing like a deep sigh, **Held within anxious eyes** Standing on different paths, Though unseen, we can feel This one unchanging heart The pain of that brief, fleeting moment **As the light flickers and fades**— Because enduring it was who we were In every scattered fragment of a passing shooting star, The promise we made to meet again Seems forgotten, yet is never forgotten, Remaining under the name of longing

Table 17. Input Poem (Kim Hyesoon's "Nalgae hwansangtong") and the Resulting Generated Poems

Input Poem	Generated Poem
하이힐을 신은 새 한마리 **아스팔트 위를 울면서 간다** **마스카라는 녹아 흐르고** 밤의 깃털은 무한대 무한대 그들은 말했다 애도는 우리 것 **너는 더러워서 안 돼** **A bird in high heels** **goes crying across the asphalt** **Mascara melting and running down,** the night's feathers are infinite, infinite They said Mourning is ours **You cannot, for you are dirty**	<밤의 새들> **하이힐 신겨진 발소리 아스팔트 위로** **눈물방울들이 흘러내려** 무수한 별들이 떨어진 자리마다 새들은 울음 섞인 노래를 부른다 **마스카라 번지는 눈가엔 꿈도 함께 번져** 무한히 이어지는 어둠 속에서 길을 잃고 누군가 던진 차가운 목소리가 메아리친다 **'더러워서 너는 안된다'** 그 한마디에 날개 꺾인 채 주저앉으며 새벽까지 걸어가는 그 발걸음 소리 도시의 불빛 사이로 스며드는 달빛처럼 희망이란 작은 씨앗 하나 품고 있지만 가슴 깊은 곳에선 여전히 울고 있는 밤의 깃털들을 나는 바라본다 <Birds of the Night> **The sound of high-heeled footsteps upon the asphalt,** **Teardrops flowing down** Wherever countless stars have fallen, The birds sing a song laced with weeping

	Around eyes where mascara smudges, dreams smudge too Lost in the infinitely stretching dark, A cold voice cast by someone echoes **"You cannot, for you are dirty."** Sinking down with broken wings at that single word, The sound of those footsteps pacing until dawn Like moonlight seeping through the city lights, Though harboring a tiny seed called hope, I gaze at the feathers of the night That are still weeping deep within

Conclusions

This study proposed a new methodology for the quantitative analysis of Korean modern poetry and its application to AI-driven poetry generation through the construction of KPoEM, an emotion-labeled dataset annotated at both the line and work levels. KPoEM comprises 7,622 entries, each annotated with 44 fine-grained emotion categories in a multi-label scheme by five expert annotators. The resulting dataset was then used for sequential fine-tuning of KcELECTRA, which was first fine-tuned on the KOTE dataset.

Quantitative evaluation on a held-out test set of 763 entries demonstrated that the proposed KPoEM model outperformed alternative configurations fine-tuned exclusively on the KOTE dataset and on KPoEM alone across all metrics, achieving an accuracy of 0.79, a micro F1-score of 0.60, and an MCC of 0.47. In qualitative evaluation, the KPoEM model accurately captured not only the dominant emotions but also the contextual emotions embedded in the poetic text, showing clear recognition of the affective registers characteristic of early twentieth-century Korean poetry, the period from which the training corpus was drawn.

Nevertheless, this study has several limitations. Due to copyright constraints, the dataset is restricted to works by five representative poets, resulting in limited temporal and authorial diversity, as well as an underrepresentation of female poets. In addition, the inherent ambiguity of poetic language and the subjectivity of emotional interpretation remain fundamental challenges that cannot be fully resolved through computational approaches alone. Despite these limitations, this study provides experimental evidence that the emotional structures embedded in poetic texts can be systematically explored through data-driven methods, offering foundational resources for expanding the intersection of literature and artificial intelligence.

Aligned with recent advancements in LLM-based poetry generation and evaluation, KPoEM provides a practical foundation for both AI-assisted creative education and the preservation of Korean literary affect as structured data. In particular, future systems could integrate KPoEM to help learners explore the affective dimensions of Korean poetry. Furthermore, by integrating these datasets with increasingly advanced large language models, future research could enhance the generative pipeline to achieve higher levels of emotional and stylistic sophistication. Ultimately, KPoEM contributes to the broader project of preserving Korean literary sensibility through structured, machine-readable data.

Beyond these contributions, KPoEM is conceived as the foundation for a Co-Reading environment—a human–AI collaborative system for poetic text reconstruction—in which humans and AI work together through sensory and emotional colors of the poetry (Lim 2025). Future research could extend this framework by constructing a dataset of sensory elements in Korean modern poetry (Ji 2025), thereby enabling more holistic literary analysis that integrates both sensory and emotional dimensions of poetic experience. More broadly, this study contributes to ongoing conversations in digital humanities about how culturally specific literary corpora can inform—and challenge—the development of language models trained predominantly on contemporary, non-literary data. By centering Korean modern poetry as both a computational object and an interpretive horizon, KPoEM offers a model for sustaining literary specificity within an era of large-scale, general-purpose AI.

References

Acheampong, Francisca Adoma, Wenyu Chen, and Henry Nunoo-Mensah. 2020. "Text-Based Emotion Detection: Advances, Challenges, and Opportunities." *Engineering Reports* 2 (7): e12189. https://doi.org/10.1002/eng2.12189.

Ahmad, Shakeel, Muhammad Zubair Asghar, Fahad Mazaed Alotaibi, and Sherafzal Khan.2020. "Classification of Poetry Text into the Emotional States Using Deep Learning Technique." *IEEE Access* 8: 73865–78.

https://doi.org/10.1109/access.2020.2987842.

Bena, Brendan, and Jugal Kalita. 2019. "Introducing Aspects of Creativity in Automatic Poetry Generation." Paper presented at the 16th International Conference on Natural Language Processing, International Institute of Information Technology, Hyderabad, India, December 18–21. https://aclanthology.org/2019.icon-1.4/.

Bhat, Pranav, Karthik K P, Shrishail Golappanavar, Rohit Mendigeri, Uday Kulkarni and Shashank Hegde. 2025. "Poetry Generation Using Transformer Based Model GPT-Neo." Paper presented at the 3rd International Conference on Futuristic Technology, Hotel Crowne Plaza Pune Citi Centre, Pune, Maharashtra, February 21–22.

Brown, Tom B. et al. 2020. "Language Models are Few-Shot Learners." *arXiv*. https://doi.org/10.48550/arXiv.2005.14165.

Clark, Kevin, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. "ELECTRA: Pre-Training Text Encoders as Discriminators Rather Than Generators." *arXiv*. https://doi.org/10.48550/arXiv.2003.10555.

Demszky, Dorottya, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade and Sujith Ravi. 2020. "GoEmotions: A Dataset of Fine-Grained Emotions." *arXiv.* https://doi.org/10.48550/arXiv.2005.00547.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." Paper presented at the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Hyatt Regency, Minneapolis, June 2–7.

Haider, Thomas, Steffen Eger, Evgeny Kim, Roman Klinger, and Winfried Menninghaus. 2020. "PO-EMO: Conceptualization, Annotation, and Modeling of Aesthetic Emotions in German and English Poetry." *arXiv*. https://doi.org/10.48550/arXiv.2003.07723.

Han, Kang 한강. 2013. *Seorab e jeonyeok eul neoeo dueotta* 서랍에 저녁을 넣어 두었다 [I Put the Evening in the Drawer]. Seoul: Moonji Publishing.

Han, Yong-un 한용운. 2016. *Nim ui chimmuk* 님의 침묵 [Silence of My Beloved 1]. Gwacheon: The Story.

Heflin, Judy. 2020. "AI-Generated Literature and the Vectorized Word." MA diss., Massachusetts Institute of Technology.

Jeon, Duyoung, Junho Lee, and Cheongtag Kim. 2024. "User Guide for KOTE: Korean Online That-gul Emotions Dataset." Paper presented at the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, Lingotto Conference Centre, Torino, May 20–25.

Jeong, Keutbyul 정끝별. 2025. *Aei i wa Hanguk hyeondaesi* AI와 한국 현대시 [AI and Modern Korean Poetry]. Seoul: CommunicationBooks.

Ji, Haein 지해인. 2024. “Dijiteol gamgak gamjeong bunseok eul tonghan Isang munhak ui epipani yeongu” 디지털 감각·감정 분석을 통한 이상 문학의 에피파니 연구 [A Study on the Epiphanies in Yi Sang’s Literature Using Digital Sense and Emotion Analysis]. *Sangheo hakbo* 상허학보 [The Learned Society of Sanghur`s Literature] 72: 753–827.

———. 2025. “Dib leoning giban ui munhak gamgak yoso bunseok eul wihan yebijeonk gochal” 딥러닝 기반의 문학 감각 요소 분석을 위한 예비적 고찰 [Preliminary Study on Deep Learning-Based Analysis of Sensory Elements in Literature]. *Inmunhak yeongu* 인문학 연구 [Journal of Humanities] 39: 83–120.

Jing, Zhi, Yongye Su, Yikun Han, Bo Yuan, Haiyun Xu, Chunjiang Liu, Kehai Chen, and Min Zhang. 2025. “When Large Language Models Meet Vector Databases: A Survey.” *arXiv*. https://doi.org/10.48550/arXiv.2402.01763.

Kang, Wookyu 강우규. 2024. “Gungmun jangpyeon soseol gamjeong deiteo modelling ui hyeonsiljeok bangan gochal: gamjeong bullyu chegye jeongnip eul jungsim euro” 국문장편소설 감정 데이터 모델링의 현실적 방안 고찰: 감정분류체계 정립을 중심으로 [Consideration of Realistic Ways to Model Sentiment Data for Korean Full-length Novels: Centered on Establishing an Emotion Classification System]. *Minjok munhaksa yeongu* 민족문학사연구 [Journal of Korean Literary History] 84: 397–432.

Kim, Byung-jun 김병준, and Jung-hwan Cheon 천정환. 2020. “Baksa hagwi nonmun 2000–2019 deiteo bunseok eul tonghae bon Hanguk hyeondae munhak yeongu ui byeonhwa wa jeonmang” 박사학위 논문(2000–2019) 데이터 분석을 통해 본 한국 현대문학 연구의 변화와 전망 [The Changes and Prospects of Studies on Modern Korean Literature: Data Analysis of Doctoral Dissertations from 2000 throughout 2019]. *Sangheo hakbo* 상허학보 [The Learned Society of Sanghur’s Literature] 60: 443–517.

Kim, Chunsu 김춘수. 2004. *Gimchunsu sijeonjip* 김춘수 시전집 [The Collected Poems of Kim Chunsu]. Seoul: Hyundae munhak.

Kim, Hyesoon 김혜순. 2019. *Nalgae hwansangtong* 날개 환상통 [Phantom Pain

Wings] Seoul: Moonji Publishing.
Kim, Su-young 김수영. 1974. *Geodaehan ppuri* 거대한 뿌리 [The Great Root]. Seoul: Minumsa.
Leto, Alexandria, Cecilia Aguerrebere, Ishwar Bhati, Ted Willke, Mariano Tepper, and Vy Ai Vo. 2024. "Toward Optimal Search and Retrieval for RAG." *arXiv*. https://doi.org/10.48550/arXiv.2411.07396.
Li, Wenhao, Fanchao Qi, Maosong Sun, Xiaoyuan Yi, and Jiarui Zhang. 2021. "CCPM: A Chinese Classical Poetry Matching Dataset." *arXiv*. https://doi.org/10.48550/arXiv.2106.01979.
Lim, Iro 임이로. 2025. "Deiteoset ui hwangyeongjeok jeonhwan: ingong jineung gwa hamkke ilggi reul tonghan Hanguk geunhyeondae si gamjeong saekchae meoltimodal midieo" 데이터셋의 환경적 전환: 인공지능과 함께 읽기(Co-Reading)를 통한 한국 근현대 시–감정–색채 멀티모달 미디어 [Transforming a Dataset into an Environment: Human–AI Co-Reading of Korean Modern Poetry through Emotion–Color Multimodal Media]. Paper presented at 2025 Chung-Ang University Graduate Student Conference in Film and Media Studies, Chung-Ang University, Seoul, December 8. https://doi.org/10.5281/zenodo.17850832.
Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." *arXiv*. https://doi.org/10.48550/arXiv.1907.11692.
Lo, Kai-Ling, Rami Ariss, and Philipp Kurz. 2022. "GPoeT-2: A GPT-2 Based Poem Generator." *arXiv*. https://doi.org/10.48550/arXiv.2205.08847.
Oliveira, Hugo Gonçalo. 2012. "PoeTryMe: A Versatile Platform for Poetry Generation." Paper presented at the ECAI 2012 Workshop on Computational Creativity, Concept Invention, and General Intelligence, University Montpellier 2, Montpellier, on August 27–28.
———. 2017. "A Survey on Intelligent Poetry Generation: Languages, Features, Techniques, Reutilisation and Evaluation." Paper presented at the 10th International Natural Language Generation Conference, University of Santiago de Compostela, Santiago de Compostela, Spain, September 4–7.
Panahandeh, Amir, Hanie Asemi, and Esmaeil Nourani. 2023. "TPPoet: Transformer-Based Persian Poem Generation Using Minimal Data and Advanced Decoding Techniques." *arXiv*. https://doi.org/10.48550/arXiv.2312.02125.

Park, Sang-Min 박상민, Chul-Won Na 나철원, Min-Seong Choi 최민성, Da-Hee Lee 이다희, and Byung-Won On 온병원. 2018. "Bi-LSTM giban ui hangugeo gamseong sajeon guchuk bangan" Bi-LSTM 기반의 한국어 감성사전 구축 방안 [KNU Korean Sentiment Lexicon: Bi-LSTM-Based Method for Building a Korean Sentiment Lexicon]. *Jineung jeongbo yeongu* 지능정보연구 [Journal of Intelligence and Information Systems] 24 (4): 219–40.

S, Sreeja P., and G. S. Mahalakshmi. 2019. "PERC-An Emotion Recognition Corpus for Cognitive Poems." Paper presented at 2019 International Conference on Communication and Signal Processing (ICCSP), Adhiparasakthi Engineering College, Melmaruvathur, Tamil Nadu, April 4–6.

Seoul Shinmun 서울신문. 2007. "Siin hyeopoe 'Hanguk 10-dae siin' seonjeong" 시인협회 '한국 10대 시인' 선정 [Korean Poets Association Selects Top 10 Korean Poets]. Accessed January 15, 2026. https://www.seoul.co.kr/news/society/2007/10/15/20071015027028.

Sia, Suzanna, Alexandra DeLucia, and Kevin Duh. 2024. "Anti-LM Decoding for Zero-Shot In-Context Machine Translation." *arXiv*. https://doi.org/10.48550/arXiv.2311.08324.

Sohn, Sun-Ju 손선주, Mi-Sook Park 박미숙, Ji-Eun Park 박지은, and Jin-Hun Sohn 손진훈. 2012. "Hangugeo gamjeong pyohyeon daneo ui chuchul gwa beomjuhwa" 한국어 감정표현단어의 추출과 범주화 [Korean Emotion Vocabulary: Extraction and Categorization of Feeling Words]. *Gamseong gwahak* 감성과학[Korean Journal of the Science of Emotion & Sensibility] 15 (1): 105–20.

Sprugnoli, Rachele, Francesco Mambrini, Marco Passarotti, and Giovanni Moretti. 2023. "The Sentiment of Latin Poetry. Annotation and Automatic Analysis of the Odes of Horace." *Italian Journal of Computational Linguistics* 9 (1): 53–71.

Sprugnoli, Rachele, and Arianna Redaelli. 2024. "How to Annotate Emotions in Historical Italian Novels: A Case Study on I Promessi Sposi." Paper presented at the Third Workshop on Language Technologies for Historical and Ancient Languages, Lingotto Conference Centre, Torino, Italy, May 20–25. https://aclanthology.org/2024.lt4hala-1.13.

Van Heerden, Imke, and Anil Bas. 2024. "A Perspective on Literary Metaphor in the Context of Generative AI." *arXiv*. https://doi.org/10.48550/

arXiv.2409.01053.
Zhao, Cheng, Bin Wang, and Zhen Wang. 2024. "Understanding Literary Texts by LLMs: A Case Study of Ancient Chinese Poetry." *arXiv*. https://doi.org/10.48550/arXiv.2409.00060.

LIM Iro (poetdeveloper.iro@gmail.com/bkksg.studio@gmail.com) is a master's student in Cultural Informatics at the Graduate School of Korean Studies, the Academy of Korean Studies. Her research focuses on artificial intelligence and digital humanities, particularly AI-assisted interpretation of literature, media aesthetics, and artistic approaches to computational literary analysis
(ORCID) https://orcid.org/0009-0006-1600-8818

JI Haein (cihayin@gmail.com) is a PhD student in Cultural Informatics at the Graduate School of Korean Studies, the Academy of Korean Studies. His research focuses on Computational Literary Studies, with a broader interest in the field of Digital Humanities.
(ORCID) https://orcid.org/0000-0001-6864-6789

KIM Byungjun (bjkim@aks.ac.kr/bjkim@byungjunkim.com) received a PhD in Data Science from Sungkyunkwan University. He is an Assistant Professor of Cultural Informatics at the Graduate School of Korean Studies, the Academy of Korean Studies. His research focuses on digital humanities and computational social science, with an emphasis on natural language processing and computational literary studies applied to Korean studies. He is the corresponding author of this paper.
(ORCID) https://orcid.org/0000-0001-9925-9343

Abstract

This study introduces KPoEM (Korean Poetry Emotion Mapping), a novel dataset that serves as a foundation for both emotion-centered analysis and generative applications in modern Korean poetry. Despite advancements in NLP, poetry remains underexplored due to its complex figurative language and cultural specificity. We constructed a multi-label dataset of 7,622 entries (7,007 line-level and 615 work-level), annotated with 44 fine-grained emotion categories, drawn from the works of five influential Korean poets. The KPoEM emotion classification model, fine-tuned through a sequential strategy—moving from general-domain emotion corpus (KOTE) to the specialized KPoEM dataset—achieved a micro F1-score of 0.60, significantly outperforming the baseline model (0.43). The model demonstrates an enhanced ability to identify temporally and culturally specific emotional expressions while preserving core poetic sentiments. Furthermore, applying the structured emotion dataset to a Retrieval-Augmented Generation (RAG)-based poetry generation model demonstrates the feasibility of generating poetic texts that reflect the emotional and cultural sensibilities of Korean literature. This integrated approach strengthens the connection between computational techniques and literary analysis, opening new pathways for quantitative emotion research and generative poetics.

Keywords: emotion classification, human-annotated dataset, Korean modern poetry, poetry generation, retrieval augmented generation (RAG)

Submission: 2026. 3. 12. Referee/Revision: 2026. 4. 28. Confirm: 2026. 5.26.

Appendix A: Dataset Construction Workflow

The KPoEM (Korean Poetry Emotion Mapping) dataset was constructed through a structured preprocessing and refinement workflow.[1] Specifically, (1) criteria for poet and poem selection were first established; (2) poetic texts were collected from Wikisource; (3) the texts then underwent cleaning and normalization procedures; (4) an emotion annotation framework was applied; and finally, (5) each entry was independently labeled by five expert annotators to ensure high data reliability.

Poet and Poem Selection Criteria

For the construction of the KPoEM, poet selection is conducted using the following three criteria. First, poets are selected based on bibliographic data of doctoral dissertations in the field of modern Korean literature published between 2000 and 2019. Specifically, poets whose names appear ten or more times in dissertation titles or subtitles are included (Kim and Cheon 2020). This criterion ensures the representativeness of poets who have been recurrent subjects of literary scholarship. Second, additional poets are selected based on public recognition, referencing "the Top 10 Korean Poets" list published by the Korean Poets Association (Seoul Shinmun 2007). Third, five representative modern Korean poets—Han Yong-un, Im Hwa, Kim So-wol, Yi Sang, and Yun Dong-ju—whose copyright restrictions have expired, are chosen to ensure legal clarity for dataset publication.

The primary sources are public domain poetry collections available on Wikisource, including *Nim ui chimmuk* (*Silence of My Beloved*) by Han Yong-un, *Hyeonhaetan* (*The Korean Strait*) by Im Hwa, *Jindallaekkot* (*Azaleas*) by Kim So-wol, and *Haneul gwa baram gwa byeol gwa si* (*Sky, Wind, Stars and Poetry*) by Yun Dong-ju. In the case of Yi Sang, the dataset focuses on his Korean-language series poems such as "Widok" (Critical Condition), "Ogamdo" (Crow's Eye

---

1 In the data sample examples presented in Appendix A, titles and subtitles were translated using Gemini 3 Pro and subsequently reviewed and validated by the authors. For proper nouns, the Romanized transliteration was provided alongside the translation to preserve the original phonetics. Furthermore, as the data refinement and preprocessing procedures hold greater significance than the specific illustrative content, the main bodies of the poems were not translated.

View), and “Yeokdan” (Reverse).

A total of 483 poems is collected, comprising 165 from Kim Sowol, 112 from Yun Dong-ju, 46 from Yi Sang, 43 from Im Hwa, and 117 from Han Yong-un. While each poem is preserved as a complete work-level entry, it is also further segmented into individual lines to support fine-grained emotional annotation. These segmented lines form the primary basis of the constructed dataset and its subsequent analysis.

Data Collection from Public Sources

Figure 1. Public Data Source Example: Wikisource. “Dangsin eul boatseumnida” (I Saw You)[2] by Han Yong-un.

Wikisource, a free online digital library operated by the Wikimedia Foundation, publishes XML-formatted dump files twice a month, offering complete access to its structured content (Figure 1). Each dump file includes metadata such as document titles, text bodies, and revision histories, making it a valuable resource for large-scale corpus construction.

In this study, we construct a poetry dataset using the Korean Wikisource (kowikisource) dump file dated November 1, 2024. The dump file[3] is converted

2 Available at https://ko.wikisource.org/wiki/님의_침묵/당신을_보았습니다.

3 Available at https://dumps.wikimedia.org/kowikisource/20241101/kowikisource-20241101-pages-articles.xml.bz2. Since Wikimedia dump files are periodically removed, this file may no longer be

into JSON format using the Wikiextractor tool, and each document is subsequently parsed at both the line level and the work level. The resulting JSON objects include keys such as id, revid, url, title, and text. Table 1 presents an example of raw text extracted from the Korean Wikisource before any preprocessing steps. This excerpt is from the poem "Dangsin eul boatseumnida" (I Saw You) by Han Yong-un.

Table 1. Example of Raw Text Extracted from Wikisource (Before Preprocessing)

```
{"id": "3647", "revid": "5442", "url": "https://ko.wikisource.org/wiki?curid=3647", "title": "님의 침묵/당신을 보았습니다", "text": "&lt;poem&gt;
당신이 가신 뒤로 나는 당신을 잊을 수가 없습니다
까닭은 당신을 위하느니보다 나를 위함이 많습니다
나는 갈고심을 땅이 없으므로 秋收가 없습니다
저녁거리가 없어서 조나 감자를 꾸러 이웃집에 갔더니 主人은「거지는 人格이 없다 人格이 없는 사람은 生命이 없다 너를 도아주는 것은 罪惡이다」고 말하였습니다
그 말을 듣고 돌아나올 때에 쏟아지는 눈물 속에서 당신을 보았습니다
나는 집도 없고 다른 까닭을 겸하여 民籍이 없습니다
「民籍 없는 者는 人權이 없다 人權이 없는 너에게 무슨 貞操냐」하고 凌辱하려는 將軍이 있었습니다
그를 抗拒한 뒤에 남에게 대한 激憤이 스스로의 슬픔으로 化하는 刹那에 당신을 보았습니다
아아 온갖 倫理, 道德, 法律은 칼과 黃金을 祭祀 지내는 烟氣인 줄을 알았습니다
永遠의 사랑을 받을까 人間歷史의 첫 페이지에 잉크칠을 할까 술을 마실까 망서릴 때에 당신을 보았습니다&lt;/poem&gt;"}
```

The converted JSON data is loaded in a Python (v3.10.18) environment and organized into a unified DataFrame structure using the pandas library. Poems by specific authors are retrieved via keyword search in the title field, and the corresponding text field is extracted as the primary content. HTML escape characters (e.g., <poem>) within the text are restored using the html.unescape function, and tags such as <poem> and </poem> are removed directly through string processing using regular expressions. The cleaned texts were subsequently split by newline characters (\n), segmenting each poem into individual lines, and were also divided into work-level versions split into 512-character segments, resulting in two structured versions of the dataset.

Each line is organized into a DataFrame with the column names line_id (unique identifier), poem_id (poem identifier), text (line content), sub_title (subtitle, if applicable), title (poem title), and poet (poet name). This initial line-level structuring of the poetic content forms the foundational layer for subsequent emotion annotation and computational analysis (Table 2).

available. Use the latest dump file for parsing.

Table 2. Example of Line-Level Segmentation from Processed Wikisource Data

line_id	poem_id	text	sub_title	title	poet
2849	197	당신이 가신 뒤로 나는 당신을 잊을 수가 없습니다		당신을 보았습니다 I Saw You	Han Yong-un
2850	197	까닭은 당신을 위하느니보다 나를 위함이 많습니다		당신을 보았습니다 I Saw You	Han Yong-un
2851	197	나는 갈고심을 땅이 없으므로 秋收가 없습니다		당신을 보았습니다 I Saw You	Han Yong-un
2852	197	저녁거리가 없어서 조나 감자를 꾸러 이웃집에 갔더니 主人은 「거지는 人格이 없다 人格이 없는 사람은 生命이 없다 너를 도아주는 것은 罪惡이다」고 말하였습니다		당신을 보았습니다 I Saw You	Han Yong-un
2853	197	그 말을 듣고 돌아나올 때에 쏟아지는 눈물 속에서 당신을 보았습니다		당신을 보았습니다 I Saw You	Han Yong-un
2854	197	나는 집도 없고 다른 까닭을 겸하여 民籍이 없습니다		당신을 보았습니다 I Saw You	Han Yong-un
2855	197	「民籍 없는 者는 人權이 없다 人權이 없는 너에게 무슨 貞操냐」하고 凌辱하려는 將軍이 있었습니다		당신을 보았습니다 I Saw You	Han Yong-un
2856	197	그를 抗拒한 뒤에 남에게 대한 激憤이 스스로의 슬픔으로 化하는 刹那에 당신을 보았습니다		당신을 보았습니다 I Saw You	Han Yong-un
2857	197	아아 온갖 倫理, 道德, 法律은 칼과 黃金을 祭祀 지내는 烟氣인 줄을 알았습니다		당신을 보았습니다 I Saw You	Han Yong-un
2858	197	永遠의 사랑을 받을까 人間歷史의 첫 페이지에 잉크칠을 할까 술을 마실까 망서릴 때에 당신을 보았습니다		당신을 보았습니다 I Saw You	Han Yong-un

Preprocessing Pipeline

Prior to labeling, the poetry texts collected from Wikisource—both line-level and work-level—undergo additional cleaning and normalization to ensure their suitability for training emotion classification models.

Text selection criteria and Hangeul–Sino-Korean notation rules

When poems contain both archaic Korean and modern Korean translations, only the modern version is retained in the dataset to ensure consistency and improve classification accuracy. In cases where multiple versions of a poem exist, preference is given to the version that preserves the original orthography and includes accurate notation of Sino-Korean characters 漢字. In such cases, the selected version is further normalized by researchers, who apply consistent principles to handle Hangeul–Sino-Korean pairings (e.g., 한글 [漢字]), ensuring both readability and fidelity to the poetic source (Table 3). Additionally, if discrepancies in line order are found, the parsed data is manually realigned to match the canonical order in the source. This process ensures that each annotated line preserves its semantic position within the work.

Table 3. "Tagor ui si reul ilkko" (Reading Tagore's Poems) by Han Yong-un

Preserved Original Version	Duplicate Version	Key Differences	Version Selected for Research Use
벗이여 나의 벗이여 愛人의 무덤 위의 피어 있는 꽃처럼 나를 울리는 벗이여 작은 새의 자취도 없는 沙漠의 밤에 문득 만난 님처럼 나를 기쁘게 하는 벗이여 …	벗이여, 나의 벗이여. 애인의 무덤 위에 피어 있는 꽃처럼 나를 울리는 벗이여. 작은 새의 자취도 없는 사막의 밤에 문득 만난 님처럼 나를 기쁘게 하는 벗이여. …	Preserved original orthography; accurate Sino-Korean character notation	벗이여 나의 벗이여 애인(愛人)의 무덤 위의 피어 있는 꽃처럼 나를 울리는 벗이여 작은 새의 자취도 없는 사막(沙漠)의 밤에 문득 만난 님처럼 나를 기쁘게 하는 벗이여 …

Data exclusion

Poems written in archaic Korean script (e.g., ᄒᆞ) are excluded from the initially cleaned dataset due to incompatibilities with modern encoding environments. A representative example is Kim Sowol's "5-ilbam sanbo" (Five-Night Walk), which is removed during this filtering step. In the work-level dataset, duplicated expressions are preserved as they appear in the original text; however, in the line-level dataset, duplicate lines are identified and eliminated to prevent unnecessary redundancy. Poems with strong visual formatting or image elements—such as Yi Sang's "Ogamdo sije 4-ho and 5-ho"—are also excluded, as their visual structure cannot be accurately reproduced for quantitative analysis (Figure 2).

Character normalization and cleaning

Character normalization is performed to ensure consistency in textual representation. Sino-Korean words are retained alongside Hangul in the form of Hangul (Sino-Korean), e.g., 공기 (空氣), while alternative notation forms already present in Wikisource (e.g., 단[熱한]) are preserved as-is. Non-textual symbols, excessive punctuation (e.g., "……"), and formatting artifacts are retained in the work-level dataset but removed in the line-level dataset as unnecessary noise. In cases where typographical errors are identified in Wikisource, corrections are made by referencing official printed editions of the poetry collections. For example, the misspelling *sadiri* 사디리 is corrected to the standard form *sadari* (ladder 사다리) (Y. Han 2016).

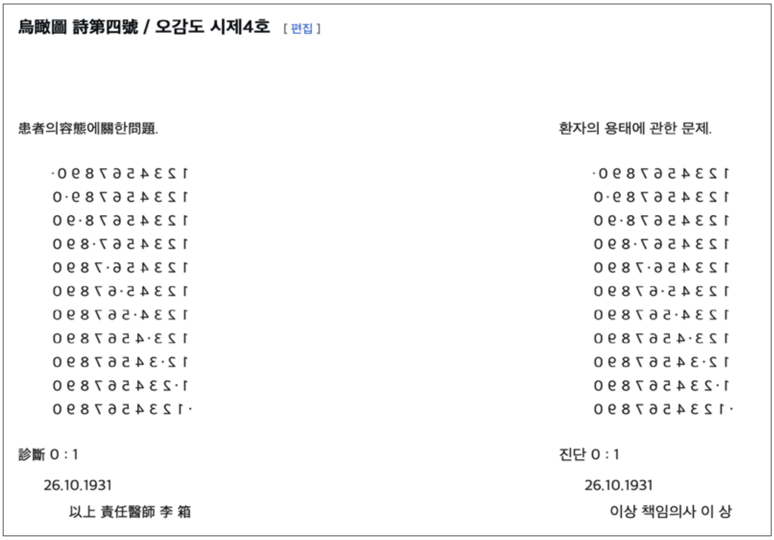

Figure 2. Public Data Source Example: Wikisource
"Ogamdo sije 4-ho" (Crow's Eye View No. 4)[4] by Yi Sang

Preservation of poetic features

To maintain the literary integrity of the source material, poetic license, older expressions, and archaisms are preserved without modernizing the language. For instance, expressions such as "ani nunmul heullioridа" (Will I not shed tears? 아니 눈물 흘리오리다) are retained in their original form. Repetitive notation such as 々, used to indicate duplication in Sino-Korean, is also preserved to avoid semantic distortion.

Exception handling in line and sentence segmentation

In the line-level dataset, most poetic lines are segmented based on newline characters (\n). Most poems are segmented using newline characters (\n). However, some poems are written in prose-like format without explicit line breaks. In such cases, if the text exceeds four sentences, segmentation is performed at sentence boundaries—typically identified by sentence-final endings such as *-da* 다, *-yo* 요, or *-orida* -오리다—as shown in Tables 4 and 5, in order to preserve semantic integrity.

4 Available at https://ko.wikisource.org/wiki/오감도.

Table 4. Original Poem Text

Original Poem Text	Sub-Title	Title	Poet
벌판한복판에 꽃나무하나가있소. 근처(近處)에는꽃나무가하나도없소. 꽃나무는제가생각하는꽃나무를 열심(熱心)으로생각하는것처럼 열심으로꽃을피워가지고섰소. 꽃나무는제가생각하는꽃나무에게갈수없소. 나는막달아났소. 한꽃나무를위(爲)하여 그러는것처럼 나는참그런이상스러운흉내를내었소.		꽃나무 Blossom Tree	Yi Sang

Table 5. Line-Level Segmentation (Processed)

line_id	poem_id	text	sub-title	title	poet
6699	450	벌판한복판에 꽃나무하나가있소.		꽃나무 Blossom Tree	Yi Sang
6700	450	근처(近處)에는꽃나무가하나도없소.		꽃나무 Blossom Tree	Yi Sang
6701	450	꽃나무는제가생각하는꽃나무를 열심(熱心)으로생각하는것처럼 열심으로꽃을피워가지고섰소.		꽃나무 Blossom Tree	Yi Sang
6702	450	꽃나무는제가생각하는꽃나무에게갈수없소.		꽃나무 Blossom Tree	Yi Sang
6703	450	나는막달아났소.		꽃나무 Blossom Tree	Yi Sang
6704	450	한꽃나무를위(爲)하여 그러는것처럼 나는참그런이상스러운흉내를내었소.		꽃나무 Blossom Tree	Yi Sang

Emotion Annotation Framework

Emotion label taxonomy

To capture the nuanced and compound emotional expressions in modern Korean poetry, we adopt a multi-label annotation scheme based on the KOTE dataset (Jeon, Lee, and Kim 2024). This taxonomy comprises 44 fine-grained emotion categories, refined and clustered from Korean emotion vocabulary using word embeddings and human validation (Table 1 in the main text). The emotion labels span a wide range of affective states—including primary emotions (e.g., anger, joy, or sadness), social or relational states (e.g., care, welcome, or respect), and culturally specific expressions (e.g., *bijangham* [resolute] or *seoreoum* [sorrow]). Each poetic line can be annotated with up to 10 concurrent emotion tags, acknowledging the inherent emotional multiplicity in poetic expression.

Controlling contextual dependency via shuffling

In the line-level dataset, to empirically test whether individual poetic lines contain sufficient lexical cues for emotion classification independent of broader

Table 6. A Randomly Shuffled Representative Sample from the Final Line-Level Annotated KPoEM

line_id	poem_id	text	sub_title	title	poet	annotator_01	annotator_02	annotator_03	annotator_04	annotator_05
6999	483	장난감신부에게 내가 바늘을주면 장난감신부는 아무것이나 막 찌른다.일력. 시집. 시계. 또 내몸 내 경험이들어앉아있음직한곳.	2밤 Night 2	I WED A TOY BRIDE	Yi Sang	anxiety, fear, guilt	fear, surprise, embarrassment, sadness, realization	fear, embarrassment, reluctant, dissatisfaction, disappointment, preposterous, contempt	shock, fear, anxiety, surprise, embarrassment, contempt	shock, fear, surprise, embarrassment, anxiety, despair, gessepany, exhaustion
6930	480	천고로창천이허방빠져있는함정에유언이석비처럼은근히침몰되어있다.	자상 Self-Injury	위독 Critical Condition	Yi Sang	despair, shock, anxiety, surprise, fear	despair, gessepany, exhaustion, realization, sadness	embarrassment, reluctant, disappointment	shock, fear, despair, gessepany, exhaustion	disappointment, despair, compassion, anxiety, sadness
4760	364	삼수갑산이 어디뇨 내가 오고 내 못 가네		삼수갑산 - 차안서삼수갑산운 Samsugapsan - Rhyming with Anseo's "Samsugapsan"	Kim So-wol	sadness, expectancy, disappointment, sorrow, anxiety, dissatisfaction, exhaustion	embarrassment, dissatisfaction, sorrow, sadness, disappointment, preposterous, despair, fed up, pathetic	embarrassment, disappointment, sorrow, dissatisfaction	embarrassment, anxiety, sorrow, sadness	embarrassment, pathetic, exhaustion, gessepany, guilt

poetic context, we constructed a shuffled dataset and carried out annotation on it (Table 6). In this dataset, the original order of lines within each poem was randomly permuted, effectively removing the narrative context and structural continuity that could otherwise influence emotional interpretation.

This experimental design draws on the structural premise of the KOTE dataset. KOTE is built to classify emotions embedded in single, self-contained comments under 512 characters, without relying on surrounding textual context. Similarly, we treat each poetic line as an atomic unit, equivalent to a comment, and create an annotation environment in which raters assess the emotional content of a line based solely on its intrinsic lexical and stylistic features, isolated from neighboring lines.This randomized arrangement serves two primary purposes: (1) to enable a focused analysis of intra-line emotional signals by eliminating inter-line dependencies and (2) to encourage emotion classification models to rely on semantic properties of the line itself, rather than on positional or narrative flow cues.

Annotation process

To ensure reliability and consistency, five annotators participated in labeling emotions for the same dataset. All annotators were trained researchers specializing in Korean literature and/or digital humanities, and each independently labeled the dataset in separate annotation sheet environments without interference from one another. In cases where conflicts arose in the annotation results, a third-party reviewer cross-validated the data and adjudicated disagreements.

To minimize subjectivity, all annotators received prior training and were provided with a standardized guideline document. This guideline included: (1) operational definitions of each emotion label, (2) illustrative labeling examples prepared by the author, (3) explicit instructions to label the work-level dataset only after completing the line-level dataset in order to remain faithful to the dataset's purpose, and (4) clear guidance on assigning multiple emotion labels. Annotators were encouraged to assign multiple emotions—up to a maximum of ten per line—whenever semantically justified. This approach follows the latest standard procedures for multi-annotator human annotation studies in NLP (Jeon, Lee, and Kim 2024).

Appendix B: Prompt Design for Emotion-Aware Poetry Generation

This appendix provides the specific prompt template used for the generative phase of the framework. The generation process utilizes the KPoEM dataset and the KPoEM Emotion Classification Model to ensure that the output maintains both poetic quality and emotional consistency.

Prompt Structure and Input Variables

The prompt is structured using a PromptTemplate that integrates three core components (Figure 1 in the main text):

1. **sample_text (Input Poem):** The original source poem provided by the user.
2. **context_snippets (Augmented Context):** 10 poetic units retrieved from the KPoEM vector database. These snippets are selected through emotion-constrained filtering to provide stylistic and lexical inspiration aligned with the target emotion.
3. **top_emotion (Emotion Scores):** An array of emotion classification scores generated by the KPoEM model. These scores act as the foundational benchmark to guide the tone of the generated poem.

Generative Prompt Template

The following template is used to instruct the Large Language Model (LLM) to act as a creative modern poet (Table 7):

Table 7. Generative Prompt Template: Original Korean Text and English Translation

Original Prompt (Korean)
PromptTemplate(input_variables=[“sample_text”, “context_snippets”, “top_emotion”], template=””” ### 시스템: 당신은 창의적이고 감성적인 근현대 시인입니다. 아래 조건을 모두 만족하는 시를 작성하세요. - 한국어로만 작성할 것 - 이모지나 그림 없이 순수 시어만 사용할 것 - 원문 텍스트는 시의 소재·상징·분위기 아이디어로 활용 - 참고 문장들에서는 표현 방식과 어휘의 결을 배우되 그대로 베끼지 말 것 - 감정 목록은 시 전체의 정서적 축으로 삼을 것 - 제목을 지을 것 - 해설 금지, 시만 작성 --- **원문 텍스트:** {sample_text} --- **참고 문장들:** {context_snippets} --- **핵심 감정 목록:** {top_emotion} --- ### 시: “””.strip())

English Translation (Gemini 3 Pro)
PromptTemplate(input_variables=["sample_text", "context_snippets", "top_emotion"], template=""" ### System: You are a creative and emotional modern Korean poet. Write a poem that satisfies all of the following conditions: - Must be written in Korean only. - Use only pure poetic language (no emojis or images). - Use the Source Text as a conceptual idea for the subject, symbolism, and atmosphere. - Learn the expression style and nuance from the Reference Sentences, but do not copy them directly. - Use the Core Emotion List as the affective axis for the entire poem. - Create a title for the poem. - Do not provide an explanation; write the poem only. --- **Source Text:** {sample_text} --- **Reference Sentences:** {context_snippets} --- **Core Emotion List:** {top_emotion} --- ### Poem: """.strip())