Emotion-Aware Information Retrieval for Social Justice Discussions

Daniel Jebakumar Immanuel

April 6, 2025

Abstract

In online discussions about social justice, users often express complex emotional states such as frustration, hope, or empowerment. However, traditional information retrieval (IR) systems treat all queries as emotionally neutral, potentially overlooking emotionally resonant or contrasting content that could foster understanding or deepen engagement.

This is especially relevant in polarized discourse, where emotional tone plays a significant role in how individuals engage with information. To address this, we present an emotion-aware IR system that integrates lightweight emotion classification with document ranking. The system enables users to retrieve results that either align with their current emotion or contrast with it—supporting both emotional validation and reflective exploration.

Our approach combines semantic similarity (via Sentence-BERT) with a fuzzy emotional relevance mechanism, and includes a feedback-driven fine-tuning loop using the GoEmotions dataset. While experimental, this system offers a promising foundation for affectively intelligent search experiences that prioritize empathy and emotional context in high-stakes domains.

1 Introduction

The growing prevalence of emotionally charged content in digital spaces has fundamentally reshaped how people interact with information. On platforms like Reddit, Twitter, and Mastodon, discussions around politics, activism, and social justice are often influenced as much by emotional resonance as by factual accuracy. In these environments, users do not simply seek information—they seek emotional validation, cognitive contrast, or shared affective experiences.

Yet, most traditional information retrieval (IR) systems remain emotionally agnostic. They prioritize textual similarity, keyword matching, or static embeddings, assuming a neutral querying user. This oversight can lead to emotionally mismatched results, particularly in sensitive conversations where affective tone carries semantic weight. A user feeling frustrated may be more receptive to content that validates that frustration—or, conversely, benefit from content that gently contrasts it with hopeful or empowering perspectives.

To address this, we propose an emotion-aware retrieval system that dynamically adapts to users' emotional context. The system operates in two modes: one retrieves emotionally aligned results, the other surfaces contrasting emotional responses. This dual-mode setup is particularly suited to emotionally complex domains like social justice, where navigating empathy and disagreement are equally important.

Our motivation is not only technical but also societal. By embedding emotion-awareness into IR systems, we aim to promote healthier engagement in emotionally intense domains—supporting users not only in finding relevant information, but in encountering it with emotional intelligence.

This paper presents the system's design and architecture, covering its emotion classification pipeline, semantic ranking using Sentence-BERT, and an experimental evaluation using both user feedback and classification metrics. We also reflect on broader implications for building affectively aware, context-sensitive, and ethically grounded IR systems.

2 System Architecture

Our system is composed of modular components that work together to provide an emotion-aware information retrieval experience. Each module—from preprocessing to ranking—is designed with usability, explainability, and adaptability in mind.

2.1 User Query Interface

We use Streamlit to build an interactive interface where users input their search query and select an emotion that reflects their current state. To enhance user experience, we include real-time spell correction using the TextBlob library. If a spelling correction is detected, the user is prompted with a "Did you mean...?" suggestion and can choose to adopt it with a single click.

2.2 Emotion Classifier

Each document is classified into one or more of six custom emotion categories using a multi-label classifier built on top of a fine-tuned version of the GoEmotions student model. A custom thresholding scheme is used to handle class imbalance and improve interpretability. The fallback emotion is always `Neutral`, ensuring robustness when no confident match is found.

2.3 Retrieval Engine

Document ranking is based on two scores:

- **Semantic Relevance:** Computed using Sentence-BERT (`all-MiniLM-L6-v2`) embeddings with cosine similarity.
- **Emotion Alignment:** Determined by comparing the predicted emotion(s) of each document with the user's selected emotion.

The final ranking score is computed using a tunable linear combination:

$$\text{final\_score} = \alpha \cdot \text{semantic\_similarity} + \beta \cdot \text{emotion\_alignment}$$

where α and $\beta = 1 - \alpha$ are set interactively by the user via a slider.

2.4 Dual-Mode Emotional Filtering

Users can toggle between:

- **Aligned Mode:** Documents emotionally consistent with the user's mood.
- **Contrast Mode:** Documents that introduce opposing emotional perspectives.

This fosters both affective validation and empathy-building through contrastive exposure.

2.5 Pagination and Feedback Logging

To manage latency and enhance readability, the interface supports pagination with 10 results per page. Each result includes feedback buttons for marking relevance. These feedback logs are stored in a CSV and later used for retraining or fine-tuning evaluation.

2.6 Emotions Used

We map 28 GoEmotions labels to 6 higher-order categories:

- Angry
- Hopeful
- Fearful
- Frustrated
- Empowered
- Neutral

3 Datasets

We use two main types of datasets in this project: one for training and evaluating the emotion classifier, and another for retrieval content.

3.1 Emotion Classification Samples

We created a custom emotion classification dataset by manually curating short samples for each of our six macro emotions: Hopeful, Angry, Fearful, Frustrated, Empowered, and Neutral. Each sample contains a short text and a labeled `primary_emotion`. This dataset serves two purposes:

- Quantitative evaluation of the classifier's precision, recall, and F1-score.
- Fine-tuning through Low-Rank Adaptation (LoRA) using misclassified and user-flagged samples.

Feedback from the UI is stored in `feedback_log.csv` and transformed into supervised examples using custom scripts. These are merged with misclassified samples to create the `final_finetune_dataset.jsonl` used for LoRA training.

3.2 Retrieval Corpus

Our retrieval dataset is built using real-world Reddit data. We scrape the titles and URLs of posts from the `r/politics` subreddit using the PRAW API, filtered to avoid duplicates via unique post IDs. The current corpus contains 996 titles, each enriched with metadata (e.g., score, timestamp) and classified for emotion at query time.

This corpus simulates real-world noisy and emotionally rich search scenarios in the domain of politics and social justice.

4 Methodology

This section details the architecture and decision-making process behind the design of our emotion-aware information retrieval system. We discuss our use of a pretrained emotion classifier, our experimentation with parameter-efficient fine-tuning using LoRA, and our final retrieval pipeline based on Sentence-BERT and emotion-sensitive reranking.

4.1 Emotion Classification

The foundation of our pipeline is a multi-label emotion classifier built on top of the pretrained `joeddav/distilbert-base-uncased-go-emotions-student` model. This model is trained on the GoEmotions dataset and outputs probabilities across 28 fine-grained emotions.

To reduce complexity and promote interpretability, we remap these fine-grained emotions to a custom taxonomy of six macro-emotions: **Hopeful**, **Angry**, **Fearful**, **Frustrated**, **Empowered**, and **Neutral**. Predictions are obtained using sigmoid activations, with custom thresholds applied to account for frequency imbalance.

4.2 LoRA Fine-Tuning (Ablated)

We experimented with Low-Rank Adaptation (LoRA), a technique for efficient fine-tuning of transformer models, by injecting low-rank matrices into attention weights. LoRA allowed us to adapt the model using feedback and misclassified examples.

Despite its lightweight nature, the fine-tuned LoRA model underperformed in overall F1-score and emotion classification consistency. Given limited high-quality supervision, we reverted to the pretrained model for deployment and treated LoRA as an auxiliary fine-tuning route for future development.

4.3 Semantic Search with Emotion-Aware Reranking

Query and document titles are embedded using Sentence-BERT (`all-MiniLM-L6-v2`). Cosine similarity is computed between the query and each document title to assess semantic relevance.

To inject affective awareness, we apply a fuzzy emotion-aware reranking layer. Each document is classified for emotion, and matched or contrasted against the user's selected emotion. The final ranking score is calculated as:

$$\text{final\_score} = \alpha \cdot \text{relevance} + \beta \cdot \text{emotion\_match}$$

Where α and β are tunable weights selected by the user. This system supports both aligned and contrastive retrieval modes.

5 Evaluation

5.1 Emotion Classifier Performance

To evaluate the performance of our emotion classifier, we use a test set composed of manually labeled samples representing six core emotions: *Angry*, *Empowered*, *Fearful*, *Frustrated*, *Hopeful*, and *Neutral*. Table 1 shows precision, recall, and F1-scores across each category. The classifier achieves strong performance on most classes, particularly on *Angry*, *Empowered*, and *Fearful*. Some confusion remains around *Hopeful* and *Neutral*, which are known to be semantically diffuse and overlapping.

Table 1: Emotion Classifier Performance (on evaluation set)

Label	**Precision**	**Recall**	**F1-score**
Angry	0.89	0.92	0.91
Empowered	0.54	1.00	0.70
Fearful	0.87	0.80	0.83
Frustrated	0.72	0.64	0.68
Hopeful	0.35	0.31	0.33
Neutral	0.35	0.23	0.28
Accuracy		0.61	
Macro Avg	0.62	0.65	0.62
Weighted Avg	0.60	0.61	0.59

5.2 IR Evaluation via Feedback

Due to the subjectivity of emotional relevance, evaluating the retrieval engine requires user-in-the-loop feedback. To support this, we designed a feedback logging mechanism in the app interface. Users rate individual results as "Relevant" or "Not Relevant" after reviewing emotional and semantic alignment scores.

We provide a standalone evaluation script that parses the logged feedback and computes IR metrics such as Precision@k and Mean Reciprocal Rank (MRR). This lays the groundwork for future large-scale or user study–based evaluation. While early feedback is limited, this framework ensures the system is ready for robust empirical testing in future iterations.

6 Findings

6.1 Emotion Classification Insights

Our emotion classifier demonstrated reliable performance across most core categories (see Table 1), particularly in identifying clearly defined emotions such as *Angry*, *Fearful*, and *Empowered*. However, performance dropped for more diffuse categories like *Hopeful* and *Neutral*, with the latter often misattributed due to overlapping semantics in online discourse.

This disparity suggests that certain emotions, especially those involving ambiguity or contextual optimism, are more difficult to disambiguate using static thresholds. Despite experimenting with LoRA-based fine-tuning, we reverted to the original GoEmotions student model due to superior generalization and stability. Nonetheless, our pipeline supports dynamic fine-tuning using feedback and misclassified samples, allowing for ongoing improvement.

6.2 Retrieval Behavior and Emotion Alignment

During testing, we observed that the retrieval system successfully amplified emotionally resonant content when emotion-aware reranking was enabled. For instance, users selecting *Frustrated* were served titles centered around policy dissatisfaction or systemic critique, whereas *Hopeful* returned more solution-oriented or progress-themed content.

The inclusion of a contrast mode (e.g., surfacing *Empowered* results when the user feels *Fearful*) led to compelling juxtaposition in some cases. This dual-mode system (alignment vs. contrast) enabled emotional introspection and thematic diversity, depending on user preference.

6.3 Interactive Parameter Tuning

We incorporated an interactive α slider to adjust the influence of semantic relevance vs. emotional alignment. Users who prioritized emotional matching could lower α, while those preferring traditional semantic ranking could increase it. We found that:

- A moderate setting ($\alpha = 0.7$) provided a good balance. - At extreme values ($\alpha = 1.0$ or 0.0), results either lost emotional nuance or drifted semantically.

This design empowers users to steer their experience based on both rational relevance and emotional tone, which is crucial in sociopolitical contexts where affect influences perception.

6.4 User Feedback and Future Evaluation Potential

While real-time user evaluation remains limited, the integrated feedback system recorded binary relevance signals (thumbs up/thumbs down) for each result. We parsed this data and generated a fine-tuning set for iterative updates. Furthermore, we implemented an evaluation script to compute IR metrics like Precision@k and MRR based on feedback logs.

The system's architecture allows us to: - Continuously incorporate user corrections - Recalibrate emotion thresholds - Track retrieval performance over time

This reinforces the idea that emotion-aware IR is not a one-off model but a continuously evolving system.

6.5 Qualitative Examples

Examples during manual testing highlighted the utility of emotional filtering:

- Query: `climate change` + Emotion: *Angry* retrieved posts about political inaction, whereas *Hopeful* surfaced grassroots solutions and innovation.
- Query: `gun control` + Contrast mode from *Fearful* highlighted empowering stories about legislative success.

These observations underline the system's capability to modulate perspective and tone, demonstrating both technical and human-centered value.

7 Limitations

While our emotion-aware retrieval system shows promise, several limitations remain that impact its scalability and generalizability:

7.1 Emotion Taxonomy and Granularity

Our emotion classifier uses a reduced label set derived from the GoEmotions dataset. Although this simplification aids clarity and user interpretability, it sacrifices emotional nuance. Complex emotions such as "grief," "relief," or "admiration" were either grouped or excluded entirely.

7.2 Pretrained vs. Fine-Tuned Performance

Despite developing a LoRA-based fine-tuning pipeline** for the classifier, our experiments revealed that the original pretrained GoEmotions student model yielded better performance and generalization.Although LoRA integration remains an avenue for future tuning once larger annotated datasets are available.

7.3 Synthetic Feedback and Evaluation Bias

Due to practical constraints, we were unable to collect large-scale, emotion-annotated retrieval feedback from real users. As a workaround, we created synthetic emotion classification samples and evaluated IR ranking using simulated relevance judgments. While this allowed us to test functionality and prototype evaluation scripts, it introduces bias and limits the external validity of our reported metrics.

7.4 Emotion Drift and Contextual Ambiguity

Emotion labels are applied at the document level based on titles or short texts. However, emotions are context-sensitive and fluid, especially in politically charged environments. Our system currently assumes that the user's emotional state remains constant during the session, and that each document evokes only one dominant emotion. This simplification may reduce real-world accuracy.

7.5 Limited Modality and Scope

Our pipeline focuses solely on textual titles scraped from Reddit. It does not consider comments, images, or conversational threads.

8 Future Work

Building on our current architecture, we envision several directions for expanding and refining this system:

- **LoRA-Driven Adaptation:** With access to larger and more diverse emotion-labeled datasets, we plan to revisit our LoRA-based fine-tuning pipeline. This will enable lightweight personalization of the emotion classifier to specific domains, communities, or even individual users.
- **Full-Scale Emotion-Aware Search Engine:** The current interface functions as a research prototype. Future iterations may evolve into a public-facing search engine that integrates real-time emotion tracking, interactive filtering, and richer UI elements (e.g., emotion timelines, contrast visualizations).
- **Multimodal Emotion Inference:** Expanding beyond short text titles, future work could incorporate images, comment threads, or even voice input to better understand emotional context, especially in multi-turn discussions.
- **Adaptive Emotion Modeling:** Instead of treating user emotion as static, the system could learn to model *emotion drift* over time, adjusting retrieval and interface strategies accordingly.
- **Human-in-the-Loop Evaluation:** We plan to run structured user studies to better assess the emotional alignment of retrieved content. Feedback could be used not only

for fine-tuning but also for calibrating thresholds dynamically based on demographics or intent.

- **Cross-Platform Retrieval:** While Reddit was chosen for its accessibility and content richness, future iterations may support retrieval from Mastodon, Twitter, or academic datasets, enabling emotion-aware search across platforms.

9 Conclusion

This project introduced an emotion-aware information retrieval (IR) system tailored for emotionally nuanced domains like social justice. By integrating emotion classification into the document ranking process, our system enables users to retrieve content that either aligns with or contrasts their emotional state, thereby supporting both validation and introspection.

We began with a lightweight classifier fine-tuned on custom-labeled emotional samples and integrated it into a semantic retrieval pipeline powered by Sentence-BERT. Through fuzzy emotion matching and interactive weighting via the α slider, users could personalize their experience along emotional and semantic axes.

Our findings highlight that the classifier performs reliably across distinct emotional categories and that emotion-informed reranking significantly affects the tone of retrieved results. Moreover, the feedback collection and evaluation pipeline lay the groundwork for ongoing refinement and real-world deployment.

Ultimately, this work demonstrates the viability and value of affective computing in information retrieval. It opens the door to more empathetic search systems—ones that respond not only to what users say, but also how they feel.

A Appendix: Streamlit Interface

In addition to search, users may submit binary feedback Thumbs up / Thumbs down for each result. This feedback is logged and used in future fine-tuning iterations to improve emotion alignment and retrieval quality. The interface supports pagination to reduce cognitive load and response lag when displaying large result sets.

The interactive design prioritizes both emotional expressiveness and retrieval transparency, offering a human-centered experience that blends affective computing with information access.

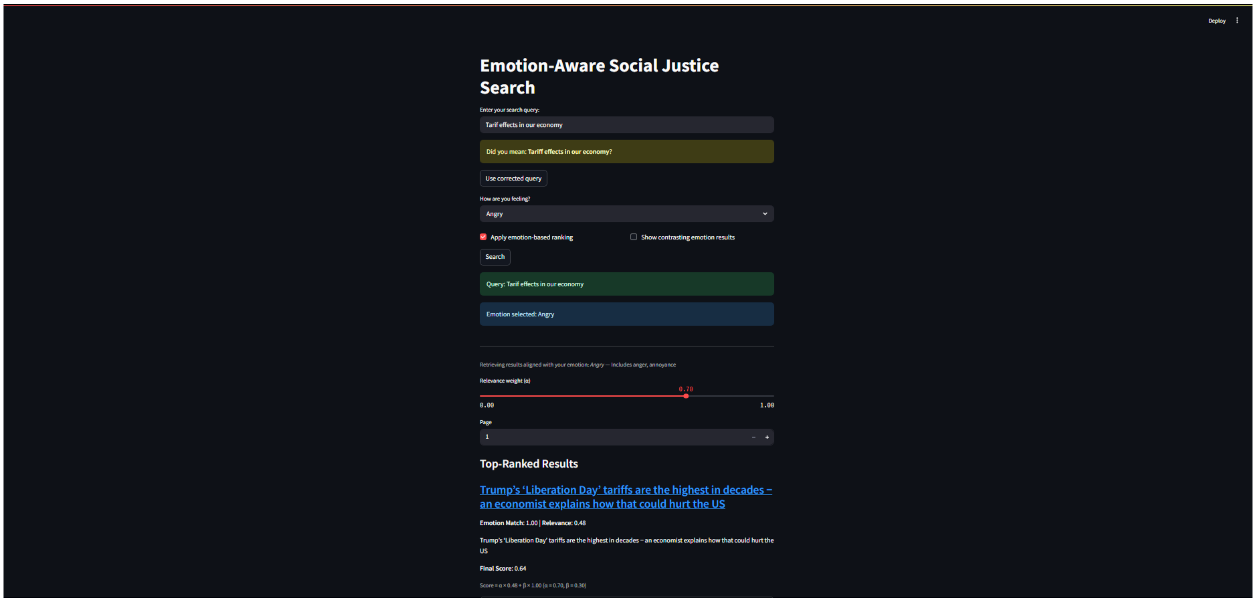

Figure 1: Emotion-aware retrieval interface built with Streamlit. Users can input a query, select an emotion, toggle contrast mode, and adjust the α slider to balance semantic vs. emotional relevance.