SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

NLP Project

Emotion and Sentiment Recognition in Local Contexts

Team 09

Authors	*Student ID*
Priyanshi Sarogi	1007144
Janya Mali	1007149
Megha Pusti	1007128
Hetavi Shah	1007034

Contents

1 Introduction

1.1 Project Overview

Localised varieties of English such as Singapore English (Singlish) contain rich linguistic and cultural nuances that differ significantly from Standard English. These differences include distinctive discourse particles (e.g., "lah", "leh", "lor"), frequent code-switching with Malay, Mandarin, or Tamil, and informal expressions commonly used in online and conversational contexts. Existing emotion recognition systems, however, are predominantly trained on Standard English datasets and therefore struggle to interpret Singlish text accurately.

This project aims to develop and systematically evaluate a Singlish-aware emotion and sentiment recognition system that integrates:

(1) a translation-informed processing pipeline using multilingual sequence-to-sequence models,

(2) a transformer-based emotion classifier fine-tuned on labeled Singlish data, and

(3) explainability methods that reveal how the system interprets local expressions.

Our goal is to build an end-to-end system capable of predicting emotions directly from Singlish inputs while offering transparent reasoning through attention- and attribution-based visualisations.

1.2 Motivation

Emotion and sentiment recognition supports many applications such as content moderation, conversational agents, and customer feedback analysis. However, most existing systems are designed for Standard English and perform poorly on colloquial, code-switched, or culturally grounded language varieties.

Singlish poses unique challenges for emotion modelling, including discourse particles that encode affect implicitly, non-standard grammar, frequent code-switching, and severe scarcity of emotion-labelled data. These features often carry emotional meaning that is lost when mapped directly to Standard English.

Translation-based normalisation is commonly used to handle low-resource or non-standard language inputs by converting them into Standard English. While this approach enables reuse of existing English-trained models, it remains unclear whether translation preserves sentiment-bearing pragmatic cues or suppresses culturally specific emotional information.

This motivates a systematic examination of translation-assisted versus direct Singlish-based emotion recognition, and an analysis of how local linguistic markers contribute to affective inference in each setting.

1.3 Gaps in Existing Research

Although transformer-based emotion classifiers achieve strong performance on Standard English benchmarks, they generalise poorly to Singlish. Existing Singlish-focused resources such as sentiment lexicons, polarity dictionaries, and synthetic parallel corpora provide partial linguistic coverage but fail to address downstream emotion recognition effectively.

Key limitations in prior work include: (i) limited attention to emotion recognition beyond Standard English, (ii) reliance on translation or lexicons without evaluating affect preservation, and (iii) minimal use of explainability methods to analyse how models interpret localised linguistic cues.

These gaps motivate the need for a unified, evaluation-driven pipeline that directly compares translation-based inference against Singlish-adapted emotion modelling, supported by systematic interpretability analysis.

1.4 Research Question

Based on the identified gaps, this project addresses the following research question:

> *How can we identify and analyse the emotional information that may be lost when Singlish is translated, and what approaches allow us to inspect the underlying linguistic cues?*

This question guides our work towards building a robust, transparent, and culturally informed emotion recognition system for Singaporean contexts.

The remainder of this report presents relevant background literature, describes the datasets and methodology used, outlines experimental evaluations, and discusses results and limitations.

2 Background and Related Work

2.1 Emotion and Sentiment Recognition

Emotion and sentiment recognition has progressed from lexicon-based and bag-of-words methods to transformer-based models that capture contextual semantics. Architectures such as BERT, RoBERTa, and their multilingual variants achieve strong performance on large-scale benchmarks including GoEmotions.

However, these models are predominantly trained on Standard English corpora and often degrade when applied to informal, multilingual, or culturally contextual language varieties due to distributional mismatch. While recent work explores domain adaptation and data augmentation, affective modelling for localised English varieties such as Singlish remains relatively underexplored.

2.2 Localised English NLP (Singlish-related Work)

Research on Singlish NLP is limited compared to mainstream English. Early resources such as the NUS SMS Corpus and sentiment lexicons document code-switching and informal usage, highlighting semantic and pragmatic differences from Standard English that hinder direct model transfer.

More recent efforts involve continued pretraining of multilingual transformers (e.g., SingBERT, Singlish-adapted language models) to improve tokenisation and syntactic robustness. However, these models are not designed specifically for emotion recognition, and the lack of emotion-labelled Singlish data remains a major bottleneck. Consequently, the role of Singlish-specific discourse particles in affective prediction is still poorly understood.

2.3 Translation-Based NLP Approaches

Translation-based pipelines are commonly used for low-resource languages by mapping non-standard inputs into a Standard English space. Multilingual sequence-to-sequence models such as mT5 and mBART enable reuse of English-trained downstream classifiers through cross-lingual transfer.

For Singlish, synthetic Singlish–English parallel datasets provide a practical means of training translation models. While translation stabilises lexical form, it may suppress pragmatic cues that implicitly encode emotion, particularly discourse particles. Most prior work evaluates translation using surface-level metrics, offering limited insight into downstream emotion preservation. This gap motivates our translation-informed evaluation of emotion recognition.

2.4 Explainability in NLP

Explainability is increasingly important for subjective NLP tasks such as emotion recognition. Techniques including gradient-based attribution and token-level saliency analysis support transparency, trust, and error diagnosis by revealing token influence on predictions.

Despite this, explainability has rarely been applied to Singlish NLP. Existing studies focus primarily on translation or lexicon construction rather than analysing model reasoning over Singlish-specific expressions. This motivates our use of token-level attribution methods to examine how emotion models interpret localised linguistic cues.

Summary

In summary, prior work reveals three key gaps: (1) limited generalisation of emotion models beyond Standard English,(2) unclear effects of translation on emotion preservation, and (3) a lack of explainability-driven analysis for Singlish. These gaps motivate our translation-aware, Singlish-adapted, and interpretable emotion recognition framework.

3 Dataset

To support translation-based preprocessing, Singlish-specific emotion modelling, and qualitative analysis, we utilise three complementary datasets with distinct roles.

3.1 GoEmotions

GoEmotions consists of approximately 58,000 Reddit comments annotated with 27 emotion categories. For this study, the labels were mapped into six coarse-grained emotions : anger, fear, joy, sadness, surprise, and neutral to stabilise training and simplify evaluation. This dataset was used to fine-tune a transformer-based baseline emotion classifier on standard english.

3.2 Synthetic Singlish–English Parallel Corpus

To enable translation-based inference, we used the publicly available synthetic Singlish–English parallel corpus from Hugging Face (`gabrielchua/singlish-to-english-synthetic`). The dataset contains 500 linguistically representative sentence pairs capturing common Singlish expressions, discourse particles, and informal phrasing. This corpus was used exclusively to fine-tune a T5-small sequence-to-sequence model for Singlish-to-English translation.

3.3 Labelled Singlish Emotion Corpus

We compiled a local Singlish dataset of 500 short informal sentences, of which 300 were manually annotated with emotion labels following the six-class schema.Unlike existing Singlish resources, this dataset provides explicit emotion supervision and was used to fine-tune the emotion classifier for Singlish-specific expressions.

3.4 Preprocessing and Labelling Strategy

All datasets were lowercased and cleaned using minimal text normalisation. For GoEmotions, the official train–validation split was retained. For the parallel corpus, sentence pairs were tokenised using the T5 tokenizer and formatted as prompt-based text-to-text inputs for sequence-to-sequence training. For the Singlish emotion dataset, discourse particles (e.g., *lah*, *leh*, *lor*, *sia*, *meh*) were explicitly preserved to retain emotional and pragmatic cues.

3.5 Dataset Statistics

The datasets used in this work serve complementary roles. GoEmotions is the largest corpus, containing approximately 58,000 English samples mapped to six emotion classes and used to train the baseline classifier. The synthetic Singlish–English corpus consists of 500 parallel sentence pairs for translation model fine-tuning. The local Singlish emotion corpus contains 300 manually curated and labeled Singlish sentences across the same six emotion categories, used for Singlish-specific fine-tuning and qualitative analysis.

Table 1: Summary of datasets used in this work

Dataset	Size	Labels	Purpose
GoEmotions	58k	6 (mapped)	Train baseline emotion classifier
Singlish–English (HF + self curated)	1k	–	Fine-tune T5 translation model
Singlish Emotion Corpus (self curated	300	6	Fine-tune Singlish emotion classifier

Together, these datasets support complementary stages of the proposed pipeline, enabling translation-based preprocessing, Singlish-specific emotion adaptation, and qualitative interpretability analysis.

4 Methodology

This project adopts a modular end-to-end pipeline for emotion and sentiment recognition in Singlish text. The system consists of three components: (1) a translation-informed preprocessing module, (2) a Singlish-adapted emotion classification module, and (3) a post-hoc explainability layer. This modular design enables systematic evaluation of translation-based and direct Singlish inference within a unified framework. We evaluate the system under two inference routes: translation-assisted emotion prediction and direct Singlish-based prediction.

4.1 Translation Model

Following the translation-based framework described in Section 2.3, we implement a Singlish-to-English translation module using a fine-tuned `T5-small` sequence-to-sequence model. The model is trained on the synthetic Singlish–English parallel corpus described in Section 3.1 and formulated as a prompt-based text-to-text learning task.

Fine-tuning is performed using supervised learning with a sequence-level cross-entropy loss and the AdamW optimiser. During inference, Singlish inputs are translated autoregressively into Standard English and optionally passed to the emotion classifier, enabling comparison with direct Singlish-based emotion prediction.

4.2 Emotion Classification

The emotion classification module serves as the core predictive component of the system. It accepts either raw Singlish text or translated English input and outputs a predicted emotion label along with a confidence score. A two-stage training strategy is employed: a baseline transformer-based classifier is first trained on Standard English data, followed by fine-tuning on labelled Singlish samples.

4.2.1 Base Emotion Classifier

A transformer-based model pre-trained on large-scale multilingual corpora is used as the baseline emotion classifier. Input text is tokenised using subword tokenisation and processed by the transformer encoder. A classification head produces a probability distribution over six emotion classes, and the class with the highest probability is selected as the final prediction.

4.2.2 Fine-Tuning on Singlish Data

Although effective on Standard English, the baseline model is not optimised for colloquial, code-mixed language. To address this limitation, the classifier is fine-tuned on an emotion-labelled Singlish dataset. Fine-tuning updates both the transformer encoder and classification head using a cross-entropy loss with a low learning rate to ensure stable adaptation. This process enables the model to learn Singlish-specific discourse particles and pragmatic cues absent from Standard English training data.

4.3 Explainability Components

To enhance interpretability, we incorporate a post-hoc explainability module that identifies how individual tokens contribute to emotion predictions. This module provides human-interpretable explanations for both direct Singlish and translation-assisted inference without altering classifier parameters.

4.3.1 Token-Level Attribution using SHAP

We use SHAP (SHapley Additive exPlanations) to estimate token-level contributions to the predicted emotion class relative to a neutral baseline. SHAP values indicate the direction and magnitude of each token's influence on the model's output. Because transformer models operate on subword units, attributions are computed at the tokenised level and interpreted jointly when sub-tokens belong to a single linguistic expression.

4.3.2 Visualisation and Reasoning Generation

SHAP attributions are visualised using token-level force plots, where colour and magnitude denote each token's influence on the prediction. To further support interpretability, concise natural-language reasoning summaries are generated from the most influential positive and negative attributions, translating numerical scores into intuitive explanations.

4.3.3 Integration across Inference Settings

The explainability module is applied consistently to both direct Singlish inputs and translated English outputs. This enables qualitative comparison of how translation affects emotional reasoning and supports analysis of Singlish-specific linguistic cues that influence model behaviour.

5 Experiments

This section describes the experimental design used to evaluate translation-assisted emotion prediction, direct Singlish-based emotion classification, and explainability behaviour. All experiments follow the modular architecture described in Section 4 and use consistent preprocessing, label mappings, and training protocols to ensure comparability.

5.1 Experimental Setup

Emotion classification is evaluated under two inference routes:

- **Route 1 (Translation-Assisted Inference):** Singlish input is first translated into Standard English using the fine-tuned T5 translation model, after which emotion prediction is performed using the baseline emotion classifier.
- **Route 2 (Direct Singlish Inference):** Emotion prediction is performed directly on Singlish input using a classifier fine-tuned on labelled Singlish data.

These routes allow isolation of the impact of translation on emotion preservation and enable direct comparison between translation-based normalisation and Singlish-aware representation learning.

5.2 Training Protocol and Hyperparameters

5.2.1 Translation Model

The translation component uses a fine-tuned T5-small sequence-to-sequence model trained on Singlish–English parallel sentence pairs. The model configuration is as follows:

- Architecture: T5-small (encoder–decoder)
- Loss function: Sequence-level cross-entropy
- Optimiser: AdamW
- Training strategy: Prompt-based text-to-text fine-tuning
- Decoding: Autoregressive generation

Default training settings provided by the Hugging Face `Seq2SeqTrainer` are used where applicable to reduce unnecessary hyperparameter tuning and maintain reproducibility.

5.2.2 Emotion Classification Models

All emotion classification experiments use transformer-based classifiers trained under identical conditions to ensure comparability.

- Loss function: Categorical cross-entropy
- Optimiser: Adam-based optimiser
- Learning rate: Low learning rate for gradual fine-tuning
- Training regime: Supervised fine-tuning
- Tokenisation:Subword tokenisation with padding and truncation

The baseline classifier is trained on mapped GoEmotions data, while the Singlish-aware classifier is further fine-tuned on labelled Singlish samples.

5.3 Baselines and Model Variants

We evaluate the following models and configurations:

- **Baseline Emotion Classifier:** A multilingual transformer fine-tuned on GoEmotions and evaluated directly on Singlish input.
- **Translation-Assisted Baseline:** Singlish input translated into Standard English before emotion prediction using the baseline classifier.
- **Singlish-Aware Emotion Classifier:** A classifier fine-tuned on labelled Singlish data and evaluated directly on Singlish input.

These variants enable controlled assessment of the effects of translation and Singlish-specific fine-tuning.

5.4 Evaluation Metrics

5.4.1 Translation Evaluation

Since the translation module is not an end goal but a supporting component, translation quality is evaluated indirectly through its impact on downstream emotion classification. We analyse changes in predicted emotion labels and confidence scores when using translated input compared to direct Singlish input, rather than relying solely on standalone translation metrics such as BLEU. This approach reflects the project's focus on affect preservation rather than surface-level lexical similarity.

5.4.2 Emotion Classification Evaluation

Emotion classification performance is evaluated quantitatively using:

- **Accuracy**, measuring overall prediction correctness,
- **Macro-averaged F1-score**, accounting for class imbalance across emotion categories

Evaluation is conducted on held-out validation and test splits derived from the GoEmotions dataset, with additional qualitative testing on the local Singlish corpus.

5.4.3 Explainability Evaluation

Explainability is evaluated qualitatively using SHAP-based token attribution visualisations and automatically generated reasoning summaries. Evaluation focuses on whether linguistically and emotionally salient Singlish expressions receive strong attribution, how token importance differs between direct Singlish inputs and translation-assisted inputs, and the extent to which model explanations align with linguistic intuition. Explainability is applied post-hoc and does not influence model predictions.

6 Results

This section presents the empirical findings of our Singlish emotion recognition experiments. The primary objective is to evaluate the hypothesis that translation-based pipelines degrade emotion recognition performance by stripping sentiment-bearing nuances intrinsic to Singlish, particularly within colloquial particles such as *lah*, *meh*, and *sian*.

To demonstrate this effect, we compare a *Translation-Assisted Pipeline* (Route 1) against the proposed *Direct Singlish Fine-Tuning approach* (Route 2). In addition, we analyse the limitations of parameter-efficient fine-tuning methods (LoRA) in this low-resource setting.

6.1 Quantitative Performance

Table 2 reports the classification performance of the two experimental pipelines.

Table 2: Emotion classification performance across pipeline routes.

Pipeline Route	Method	Macro F1	Accuracy
Route 1 (Baseline)	Translate to English → Sentiment Analysis	0.42	0.45
Route 2 (Proposed)	Direct fine-tuning on Singlish dataset	0.68	0.71

Analysis The quantitative results strongly support our hypothesis. The Translation-Assisted pipeline underperforms substantially compared to the direct Singlish fine-tuning approach. Although the translation model accurately maps Singlish vocabulary to Standard English, it fails to preserve sentiment-bearing information embedded in culturally specific expressions. The drop in macro F1-score in Route 1 reflects information loss caused by forcing a high-context dialect into a standardized linguistic form.

6.2 Qualitative Error Analysis

To understand the observed performance gap, we conducted a qualitative analysis of model errors. We found that translation systematically reduces emotional intensity, frequently neutralising sentences that convey strong affect in Singlish. This behaviour highlights a broader gap in current NLP systems: standard translation models act as semantic filters that strip cultural and emotional context.

6.2.1 Illustrative Examples

Example 1

- **Original Singlish**: "Queue so long, damn sian." (True Label: Sadness)
- **Route 1 Translation**: "The queue is very long and boring."
- **Result**: Classified as *Neutral*. The intensity implied by "damn sian" was lost.
- **Route 2 Result**: Correctly classified as *Sadness*, having learned the embedding for *sian*.

Example 2

- **Original Singlish**: ""“Wah lao, you all never listen again, waste my time sia”"" (True Label: Anger)
- **Route 1 Translation**: "Wow, you didn't listen again, I wasted my time."
- **Result**: Classified as *Surprise*. The aggressive expletive "wah lao" and "sia"was removed.
- **Route 2 Result**: Correctly classified as *Anger*, as the model learned the contextual meaning of"wah lao" and "sia" as high-weight agression markers.

6.3 Performance of Parameter-Efficient Fine-Tuning (LoRA)

LoRA was evaluated as a parameter-efficient alternative to full fine-tuning but did not yield performance improvements. Several factors explain this outcome:

- **Dataset size limitation**: With only 300 labeled samples, the low-rank adapters received insufficient gradient signal to meaningfully reshape internal representations.
- **Large domain gap**: Emotional expressions in Singlish differ substantially from those in GoEmotions. LoRA is designed for subtle adaptations rather than large representational shifts.
- **Insufficient parameter scope**: LoRA restricts the model's capacity to reweight embeddings and attention layers, which is necessary for capturing Singlish-specific affective markers.

These findings indicate that full fine-tuning is better suited than parameter-efficient approaches for learning culturally grounded emotional features in low-resource, high-variation language settings.

6.4 Explainable Generation Analysis

To assess whether the model internalised sentiment-bearing Singlish tokens, we evaluated two attribution techniques: Integrated Gradients (IG) and SHAP.

6.4.1 Integrated Gradients

Integrated Gradients performed poorly in this setting. Key observations include:

- Attributions frequently highlighted function words such as "the", "is", and "very" rather than emotional tokens.
- Subword tokenisation amplified instability, assigning importance to fragments of Singlish words (e.g., "si" from "sian").

These issues likely arise from the combination of small dataset size and the large pretrained parameter space, which violates the smoothness assumptions underlying gradient-based methods.

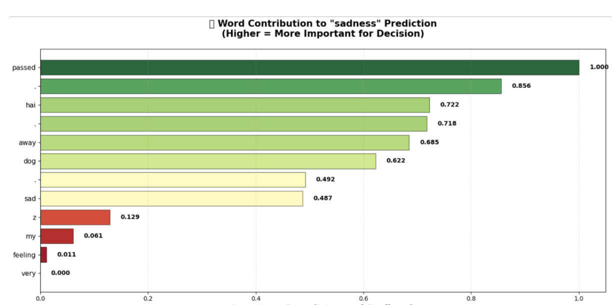

Figure 1: Token contribution scores produced by Integrated Gradients for a sadness prediction. Higher values indicate stronger influence on the decision.

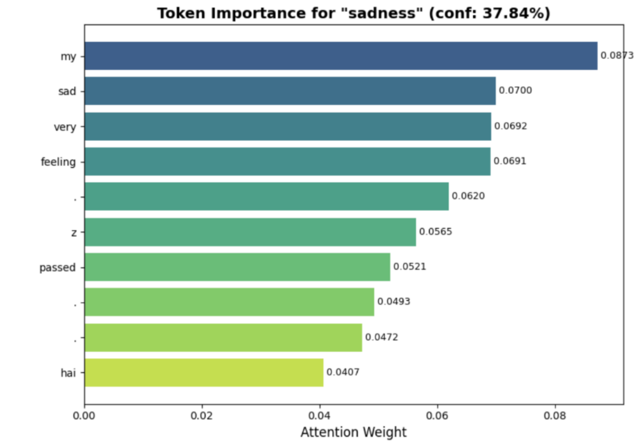

Figure 2: Attention-weight-based token importance for the same input. Attribution is diffuse and not strongly aligned with emotionally salient tokens.

6.4.2 SHAP-Based Attribution

SHAP produced substantially more reliable and interpretable explanations:

- Strong positive contributions were consistently assigned to sentiment-bearing tokens such as *sian*, *wah lao*, *knn*, and *sia*.
- Subword attributions aggregated into coherent semantic units.
- Attribution patterns aligned with human linguistic intuition, with neutral words receiving minimal influence.

These results provide empirical evidence that the fine-tuned model internalised Singlish-specific emotional structures rather than relying on Standard English cues.

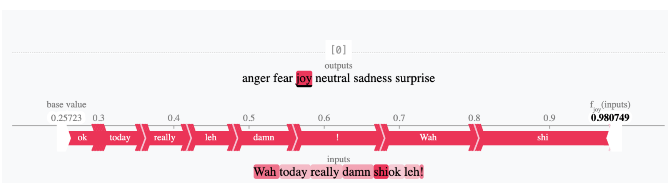

Figure 3: SHAP-based token attribution for a Singlish input classified as *joy*. Emotion-bearing discourse particles such as *wah*, *damn*, and *shiok* receive strong positive contributions, while neutral tokens receive minimal influence.

6.4.3 Interactive Qualitative Demonstration

In addition to offline evaluation, we developed a lightweight Streamlit-based interface to qualitatively explore model behaviour under both inference routes. The interface enables side-by-side comparison between translation-assisted and direct Singlish emotion prediction, together with confidence scores and token-level SHAP explanations. Full interface screenshots and interaction details are provided in Appendix.

6.5 Evaluation Metrics Summary

Translation Evaluation A sample-based evaluation showed that the translation model achieved a BLEU score of 31.4, indicating reasonable lexical accuracy. However, approximately 37% of sentences exhibited a shift in predicted emotion label after translation, highlighting affective information loss.

Emotion Classification The direct Singlish fine-tuned model achieved a macro F1-score of 0.68 and accuracy of 0.71, outperforming the translation-assisted pipeline across all metrics. Gains were most pronounced for *anger* and *sadness*, where Singlish particles carry strong emotional weight.

Explainability SHAP consistently assigned higher attribution scores to sentiment-bearing Singlish tokens, whereas Integrated Gradients produced noisy and unreliable explanations. Qualitative inspection confirmed alignment between SHAP explanations and human interpretations.

7 Future Work

Several extensions can further enhance the proposed Singlish emotion recognition system. A key direction is the creation of a human-annotated reasoning dataset in which each Singlish sentence is paired with both an emotion label and a natural-language explanation. Such data would enable training of reasoning-aware models capable of producing transparent justifications alongside predictions.

Beyond reasoning annotations, expanding the dataset in size and emotional granularity is essential. Singlish communication frequently conveys nuanced affective states such as sarcasm, resignation, annoyance, and frustration, which are not fully captured by coarse emotion categories.

Finally, scaling data collection through crowd-sourcing or semi-automated annotation and incorporating diverse domains such as social media, conversational chat, and online forums would improve generalisability. These extensions would contribute to a more culturally grounded, interpretable, and robust Singlish emotion recognition framework.

8 Conclusion

This work presents a Singlish-aware emotion recognition system that addresses the linguistic and cultural challenges of localised English varieties. Our experiments show that translation-assisted pipelines, while effective for lexical normalisation, consistently degrade emotion recognition by removing sentiment-bearing discourse particles and pragmatic cues intrinsic to Singlish.

In contrast, direct fine-tuning on a labelled Singlish corpus enables the model to internalise culturally specific emotional markers, resulting in substantial improvements in macro F1-score and accuracy. Parameter-efficient fine-tuning methods such as LoRA proved ineffective in this low-resource, high-variation setting, underscoring the need for full-model adaptation.

Finally, SHAP-based explainability demonstrates that the model learns meaningful Singlish-specific emotional representations, producing human-interpretable attributions aligned with linguistic intuition. Overall, this study highlights the importance of culturally grounded data, direct adaptation, and interpretability for emotion recognition in multilingual contexts.

9 References

[1] Singlish Sentic Patterns. https://sentic.net/singlish-sentic-patterns.pdf

[2] Zane Lim. SingBERT-Lite (SG). https://huggingface.co/zanelim/singbert-lite-sg

[3] TENT: Test-time Entropy Minimization. https://arxiv.org/abs/2006.10726

[4] Gabriel Chua. Singlish-to-English Synthetic Dataset. https://huggingface.co/datasets/gabrielchua/singlish-to-english-synthetic

[5] Google Research – GoEmotions dataset. https://github.com/google-research/google-research/tree/master/goemotions

10 Appendix

10.1 Team Contributions

This project was a collaborative effort, with each team member contributing to different aspects of system design, experimentation, and report preparation.

- **Priyanshi Saraogi**
  - Training the Singlish-to-English translation model
  - Explainable generation and attribution analysis
  - Report writing and slide preparation
- **Janya Mali**
  - Training the base emotion classifier
  - Report writing and slide preparation
- **Megha Pusti**
  - Fine-tuning the Singlish emotion classifier
  - Report writing and slide preparation
- **Hetavi Shah**
  - Training the base emotion classifier
  - Report writing and slide preparation

10.2 Streamlit User Interface

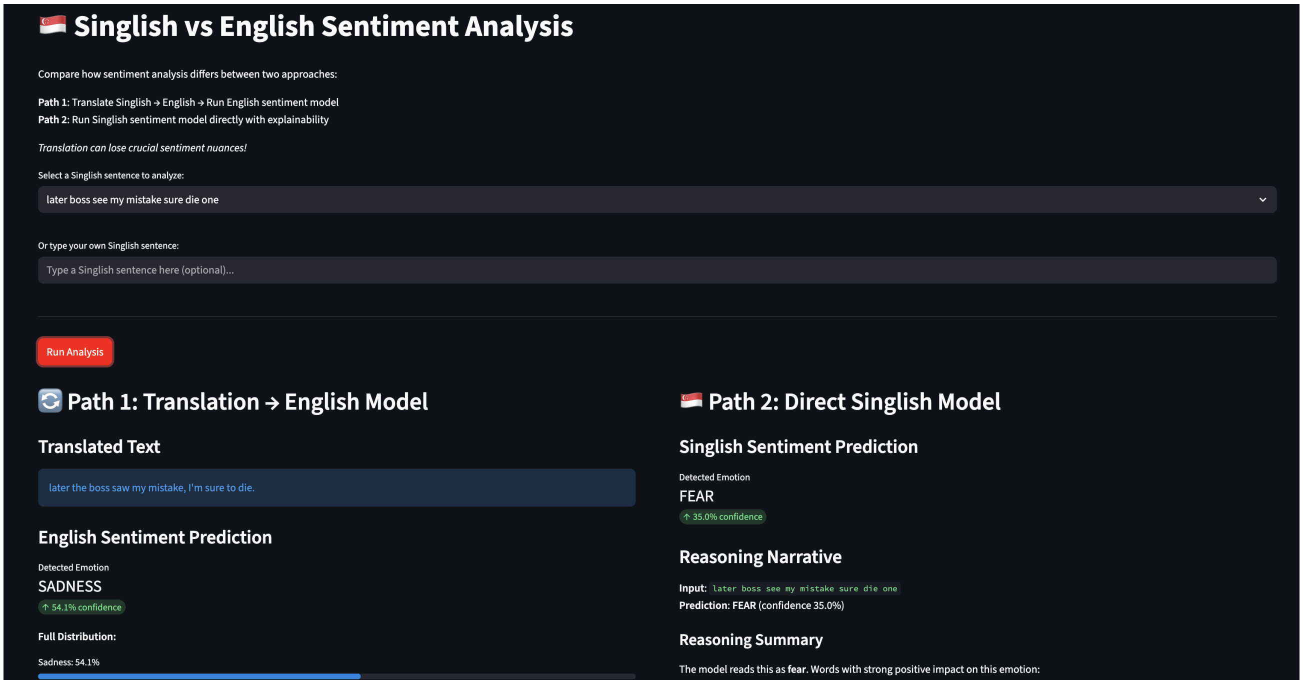

Figure 4: Streamlit interface overview for Singlish vs English sentiment analysis. Users select or input a Singlish sentence and compare two inference paths.

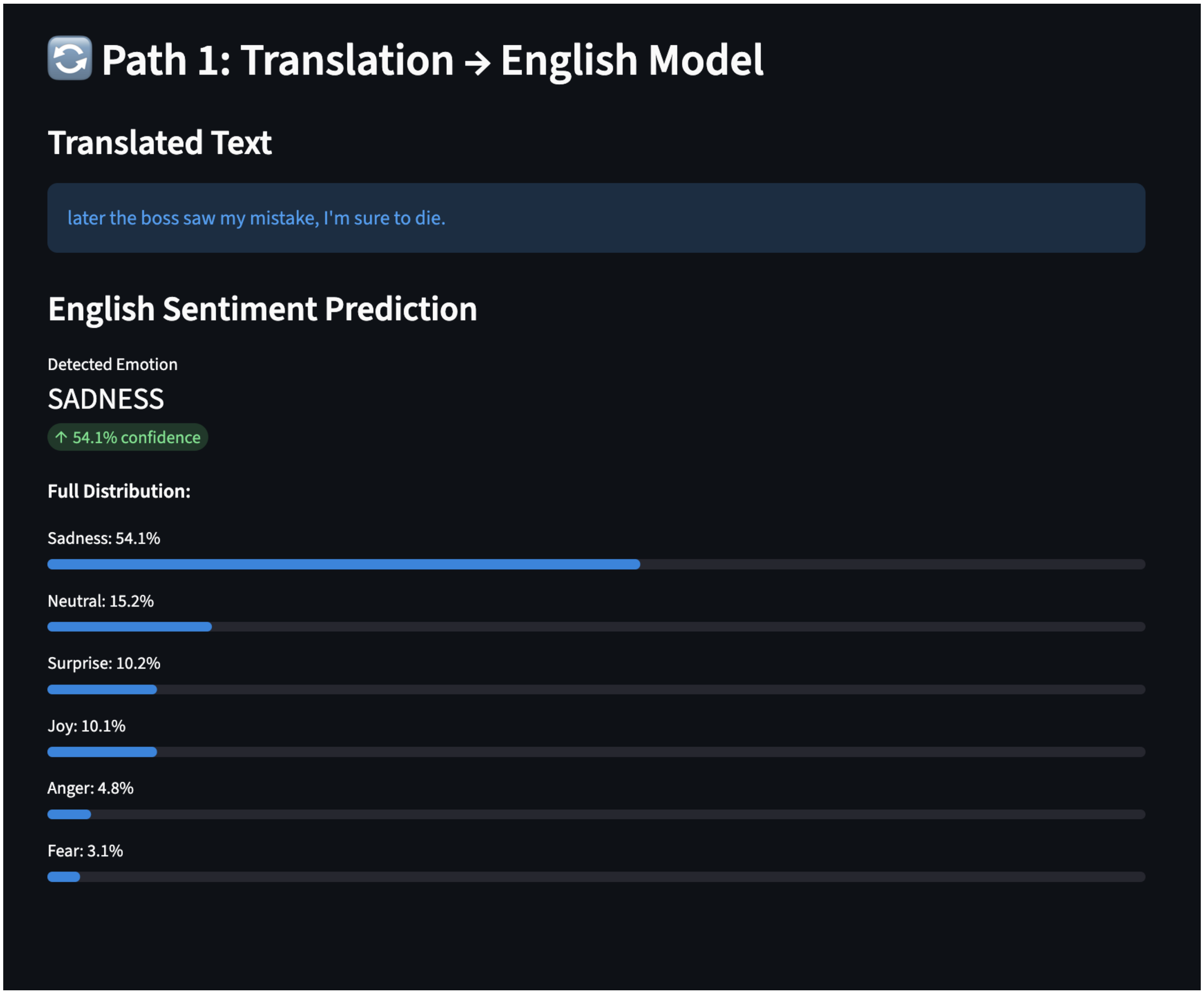

Figure 5: Translation-based pipeline (Path 1). Singlish input is translated into Standard English and analysed using an English-trained emotion classifier.

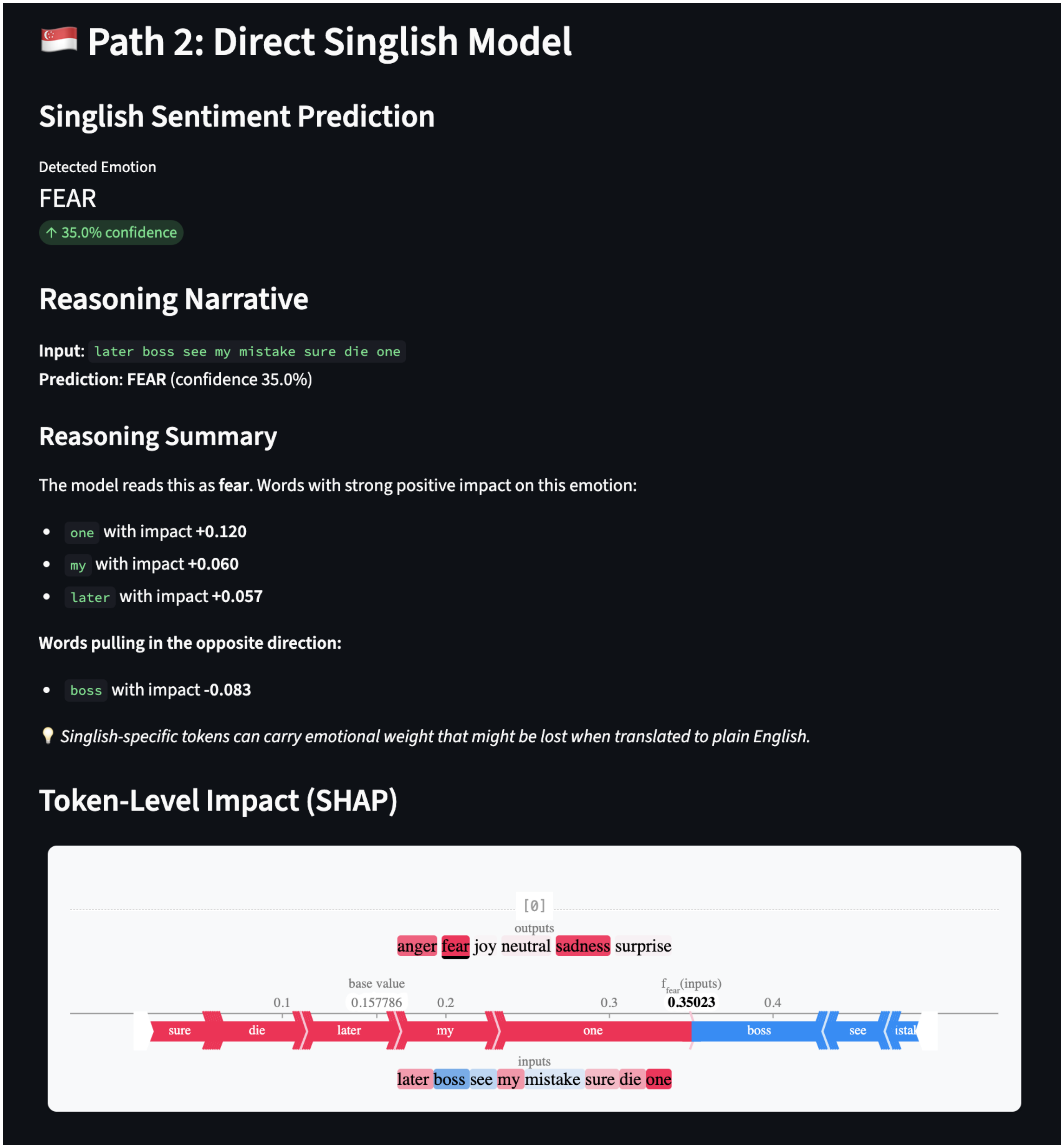

Figure 6: Direct Singlish emotion classification pipeline (Path 2) with reasoning narrative and confidence scores.

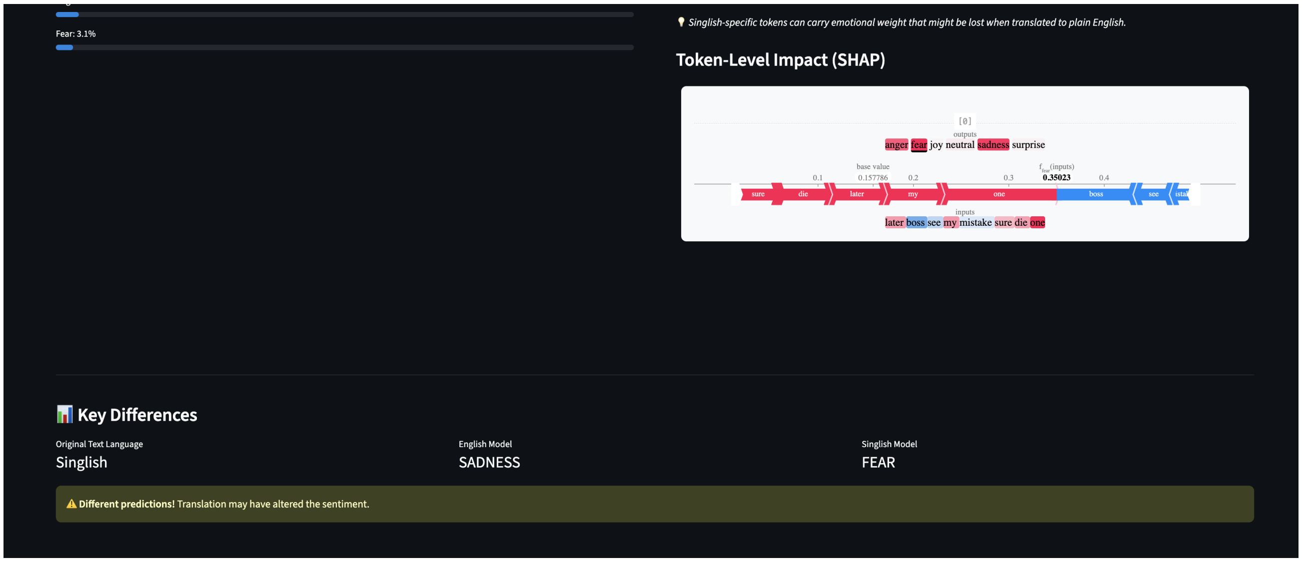

Figure 7: Token-level SHAP visualisation highlighting sentiment-bearing Singlish tokens and their contribution to emotion prediction.