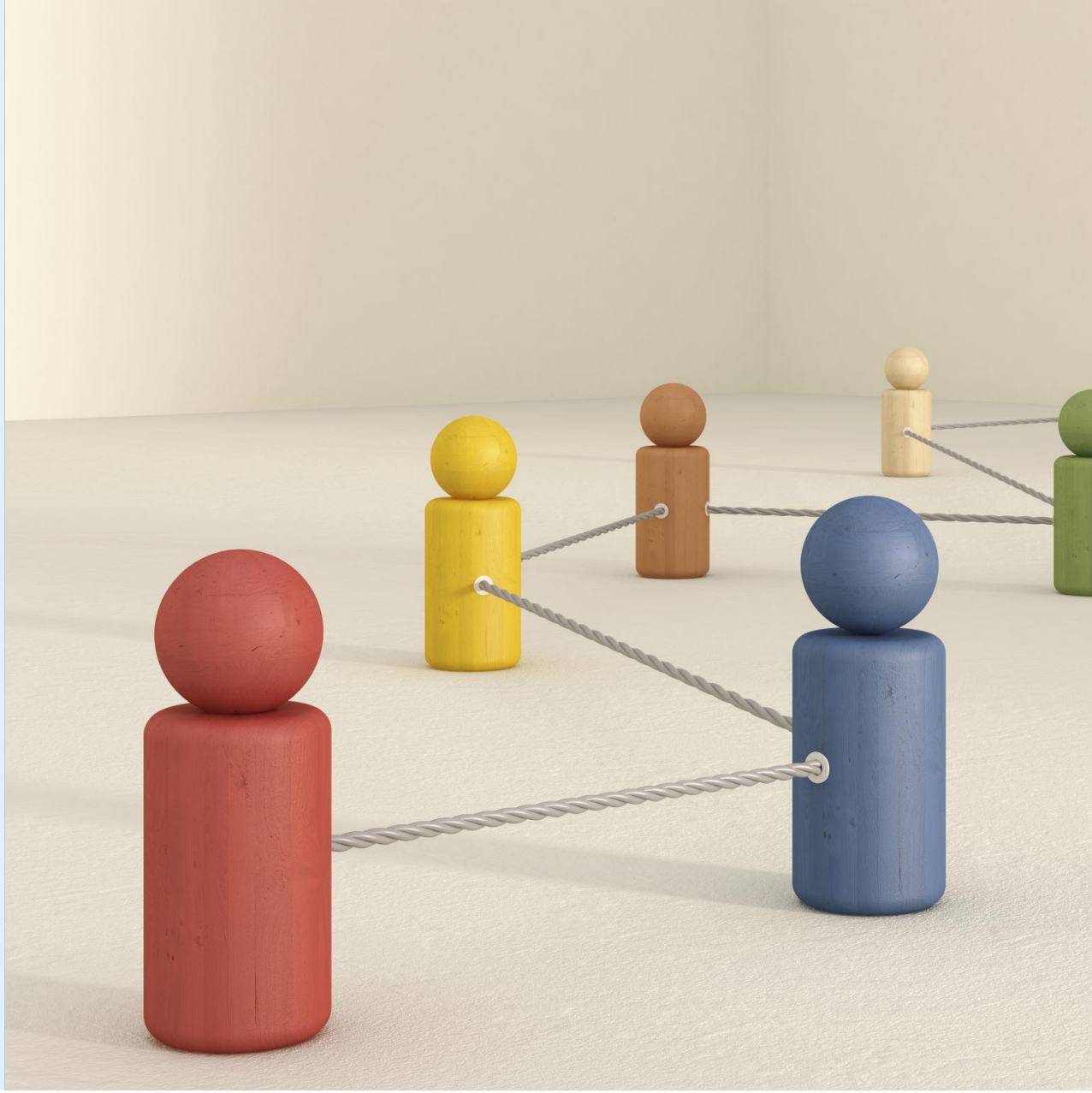
Emotion Classifier

Group 10, French

Block C Year 2 2024-2025

INDEX

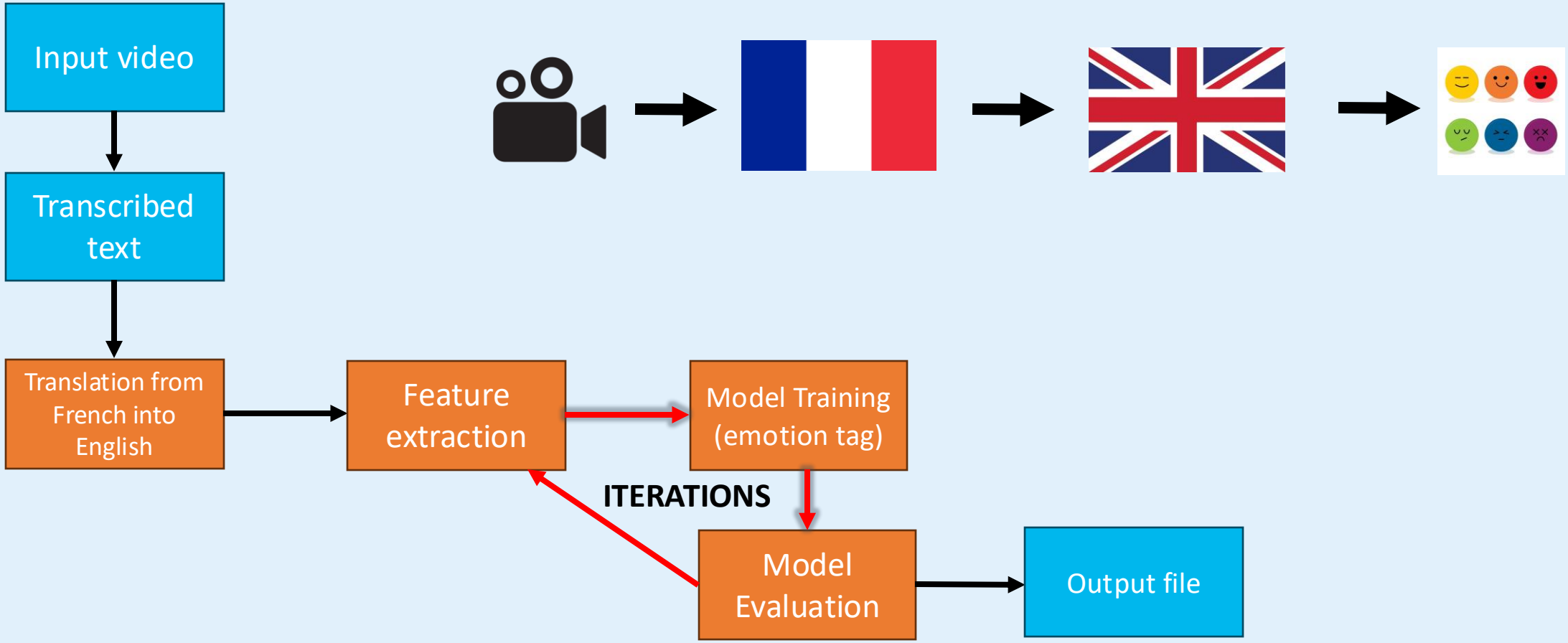
- Introduction
- Data
- Model Implementation
- Model Performance
- Ethical Considerations
- Challenges and Limitations
- Next Steps and Recommendations



Introduction

- **What is emotion classification?**
- **Why is it important?**
- **Objective:** Build a pipeline that classifies the emotions expressed in a transcribed video.
- **Client:** Content Intelligence Agency

Pipeline Overview



Data

- Training Data Sources:
- **Emotion Detection Dataset (Translated):** ~12,600 phrases from the TV show 'Friends', translated from English to French. Original labels were adapted to our 7 emotion classes.
  - **Synthetic Dataset (ChatGPT):** 3,000 generated French phrases specifically for 'sadness', 'disgust', and 'surprise' to improve class balance.
  - **Google AI GoEmotions (Translated):** ~46,000 Reddit comments, translated from English to French, with pre-existing emotion labels.

Data

- **Data Processing:**
  - **Concatenation:** All three datasets were combined into a single training corpus.
  - **Data Augmentation:** TextAttack was used to generate additional instances for underrepresented emotion classes (those with < 9000 instances) to further balance the dataset.
  - **Label Encoding:** Emotion labels were converted into numerical representations for the model.
  - **Tokenization:** Text sentences were processed into tokens using the CamemBERT tokenizer, preparing them for input into the model.

Data

- **Training/Validation Split:**
  - **Validation Set:** 20% of the combined training data was held out as a validation set to monitor model performance during training and prevent overfitting.
- **Testing Data:**
  - **Separately Curated:** A distinct test set was created from transcribed dialogue of two French reality TV shows: 'La Ville' and 'Les Marseillais en Australie'. This provides a more realistic evaluation on the model's intended use case.

Data

- **Limitations**
  - Reliance on Translated English Data
  - Potential Loss of Emotional Nuance
  - Variations in Original Emotion Labels
  - Domain Mismatch with Target Content
  - Underrepresentation of Dialects and Formal/Informal Language
  - Impact of Synthetic Data

Model Implementation

- **Core Architectural Foundation: CamemBERT Transformer:**
  - The emotion classification model is based on the Transformer architecture, a state-of-the-art neural network framework particularly effective in natural language understanding tasks.

- **Transfer Learning via Pre-trained Weights:**
  - Rather than training a model from initial parameters, we leveraged the pre-trained weights of the `camembert-base` model from Hugging Face Transformers.

Model Performance

- **Overall Performance**
  - Our emotion classifier correctly identified the emotion in **about 63%** of the French sentences in our final test.
- **Strengths**
  - **Good at Identifying Happiness:** The model is very good at recognizing when a sentence expresses happiness. It correctly identifies happy sentences about **88% of the time**.
  - **Reliable When It Says "Neutral":** When the model predicts "neutral," it's usually correct (about **78% of the time**). So, when it's confident a sentence is neutral, it's likely right.
  - **Decent at Recognizing Surprise:** The model also performs reasonably well in identifying surprise.

Model Results

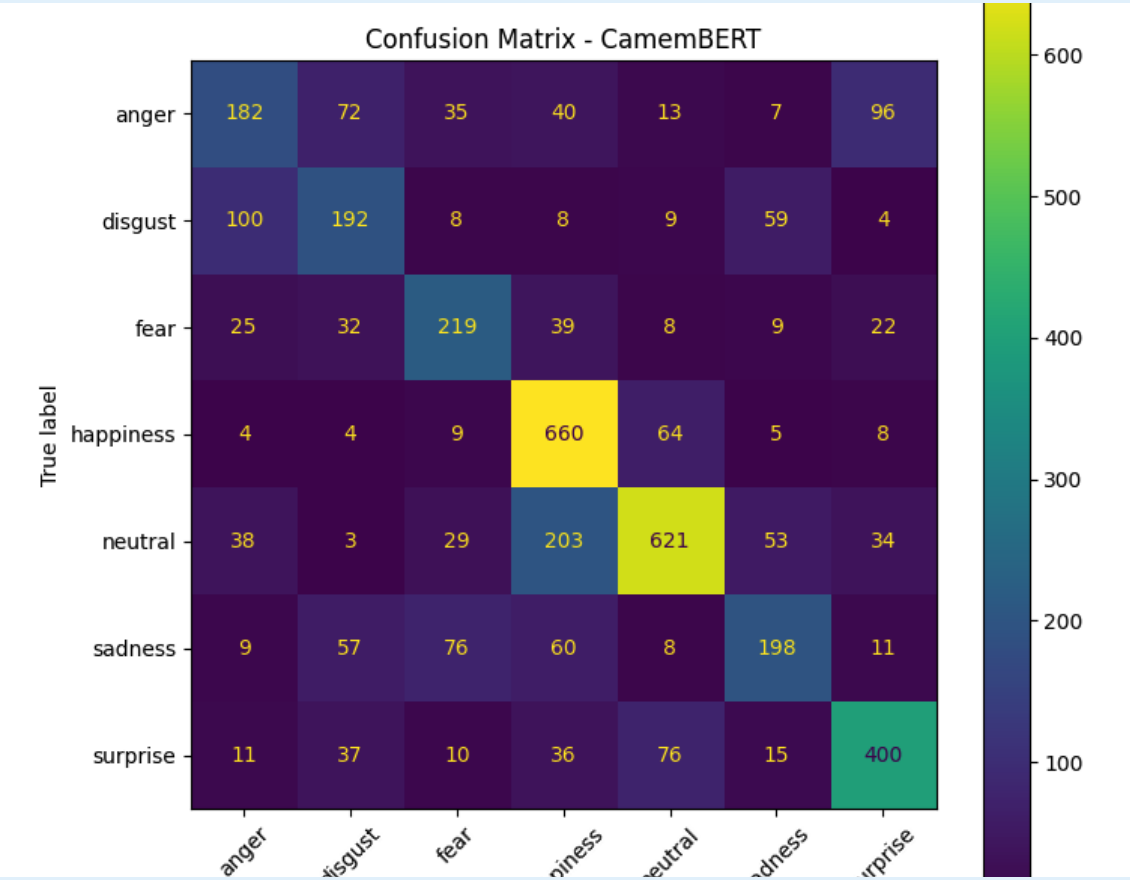
- **Weaknesses**
  - **Confusion Between "Neutral" and "Happiness"**
  - **Difficulty with "Anger"**
  - **Over-reliance on Keywords for "Disgust"**
  - **Over-attention to Punctuation**
  - **Reliance on Common, Non-Emotional Words**

Model Results

- **Quick stats**
- For our project, **recall for emotions like anger and sadness is particularly important**

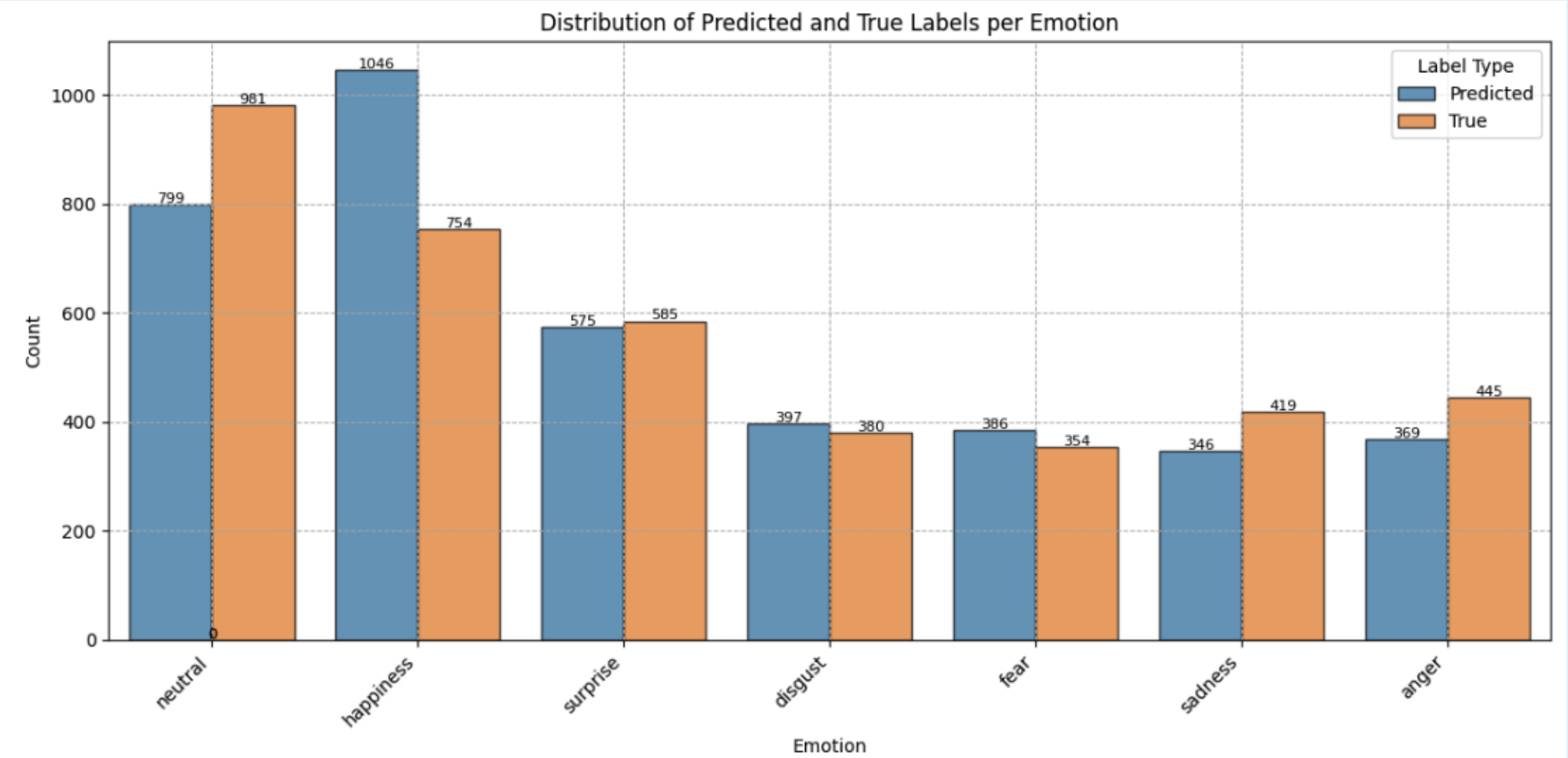
Emotion	Precision	Recall	F1 score
neutral	0.7772	0.6330	0.7002
happiness	0.7008	0.8753	0.7833
surprise	0.6846	0.6974	0.6909
anger	0.3834	0.4090	0.3958
sadness	0.5766	0.4726	0.5176
disgust	0.4842	0.5053	0.4942
fear	0.6211	0.5989	0.6098
Overall	**0.6353**	**0.6309**	**0.6264**

Model Results



- This shows which emotions the model is most often confusing with each other. You can see the darker squares along the diagonal, which represent correct predictions, but also off-diagonal squares showing common misclassifications (like neutral/happiness and disgust/anger). The 'Confusion Matrix' shows which emotions are being mixed up.

Model Results



Ethical Considerations

- Discuss the ethical considerations that influenced your choice and how that relates to the metrics discussed

Limitations of the pipeline

- What are the possible limitations for implementing this pipeline?

Next Steps

- Further model fine-tuning to improve F1 score.
- Generate more balanced data for underrepresented emotion classes.
- Potential to integrate with spoken French datasets for improved real-world applicability.
- Plan to expand to multi-turn dialogue and emotion trends across time.

Thank you for your attention.

Group 10, French

Anastasiia Mokhonko, Daria Vlăduţu,
Felipe Davila