Content Intelligence Agency

Emotion in TV Shows

Created by: Maks Burchard, Maciej Czerniak, Szymon Chirowski

Date: 31.10.2025

Today's Agenda

1. Building the Foundation
2. The Modeling
3. Ensuring Trust
4. The Final Pipeline

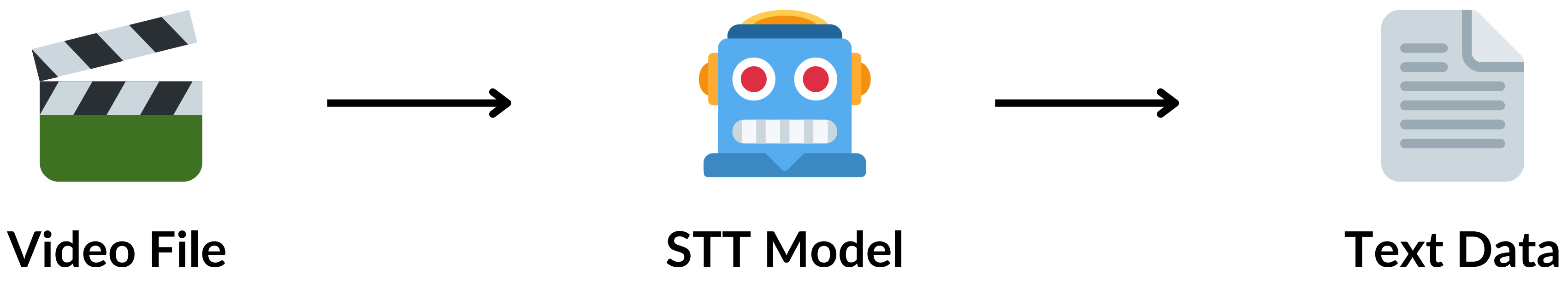
The Problem:

Analyzing emotion in media using LLMs is costly, and unsafe. The Content Intelligence Agency needs a local, cheaper solution.

Our Mission:

To build a robust, reproducible pipeline that transforms a raw video file into accurately labeled emotional data.

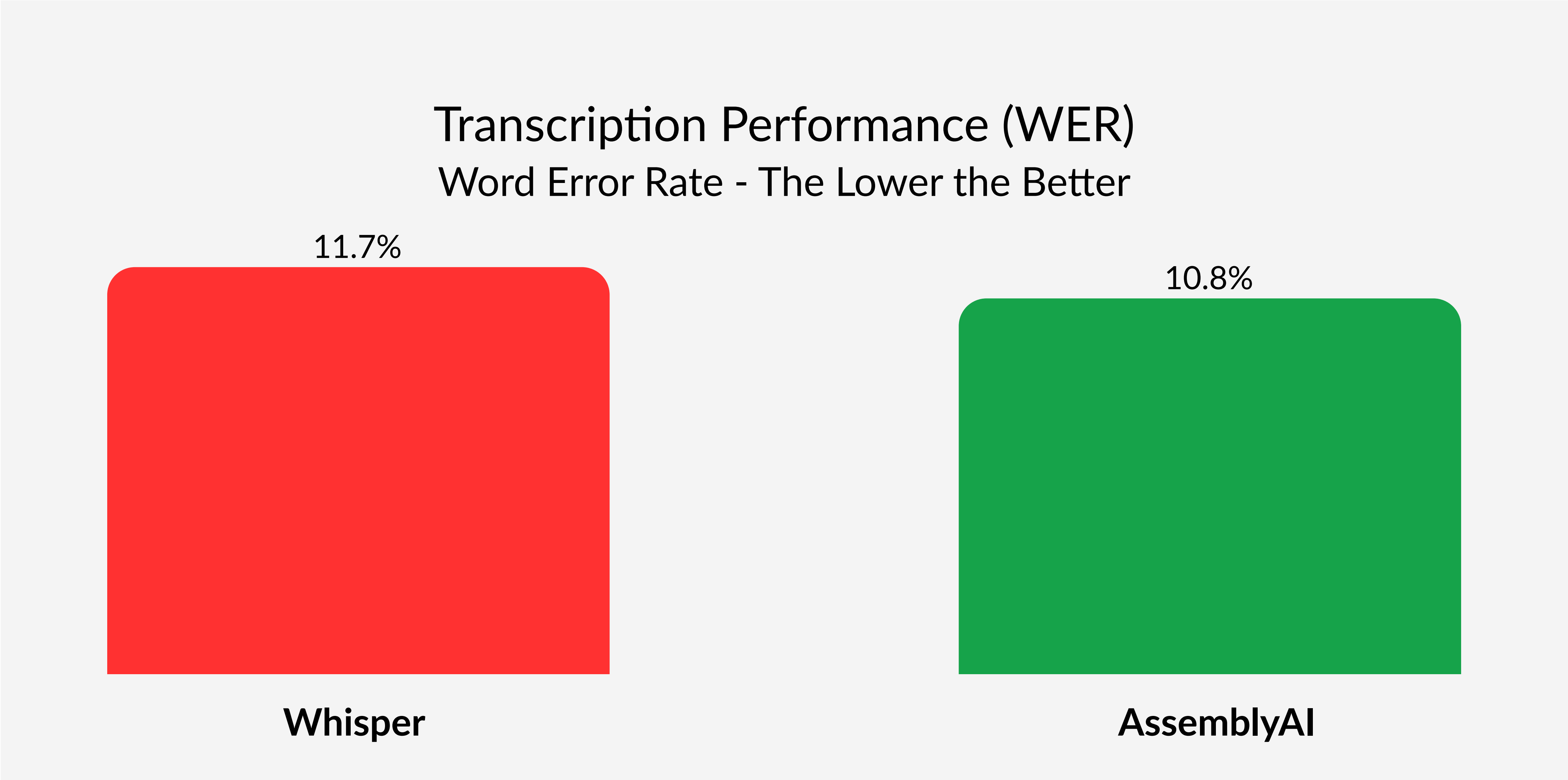
Step 1: From Speech to Text



Methodology

- Transcribe using **AssemblyAI** and **OpenAI's Whisper**.
- Evaluate quality to select the best foundation for our pipeline.

The Verdict: Data Quality is Everything

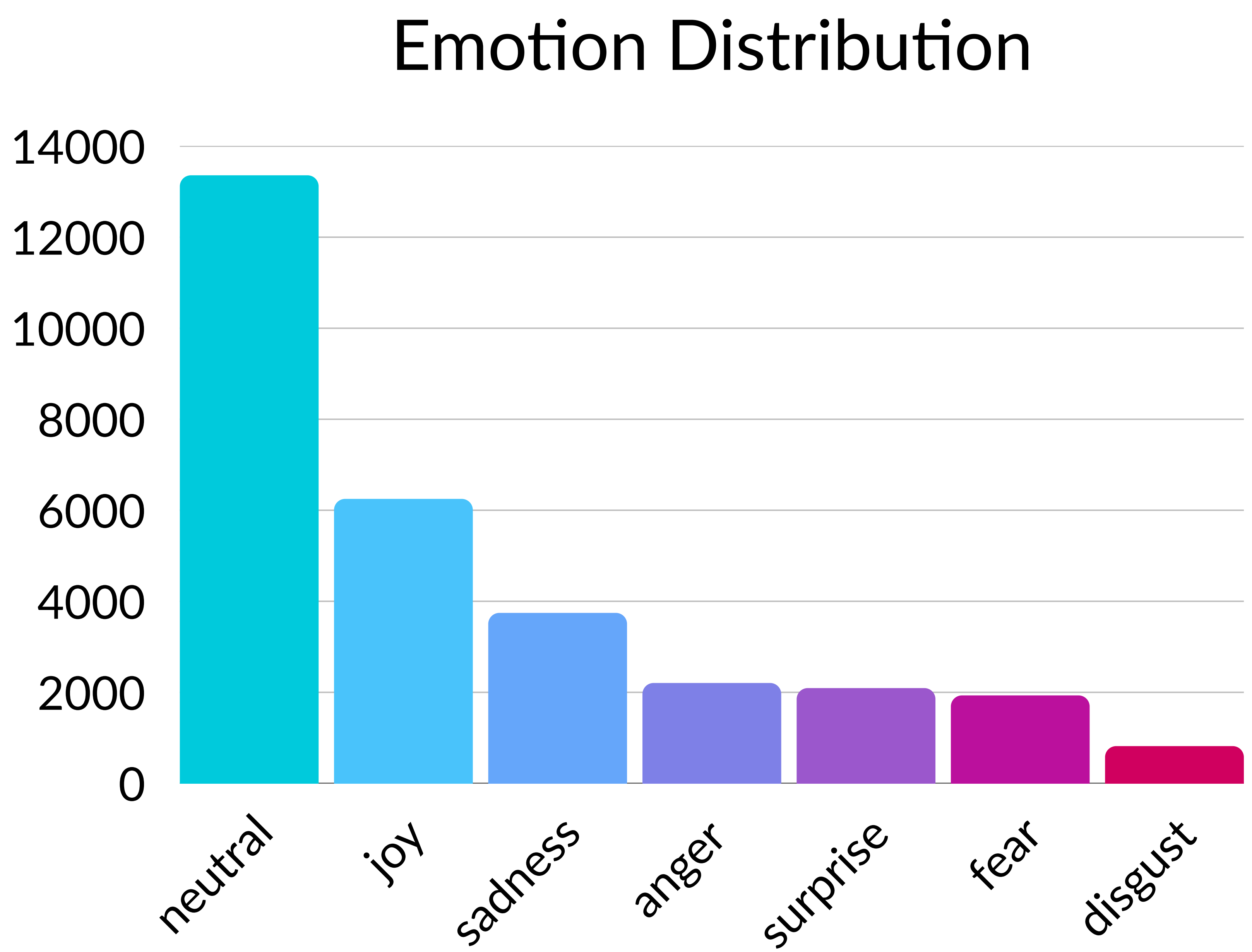


Decision: AssemblyAI was chosen for its higher accuracy, and lower cost of run providing a more reliable foundation.

Data

Summary:

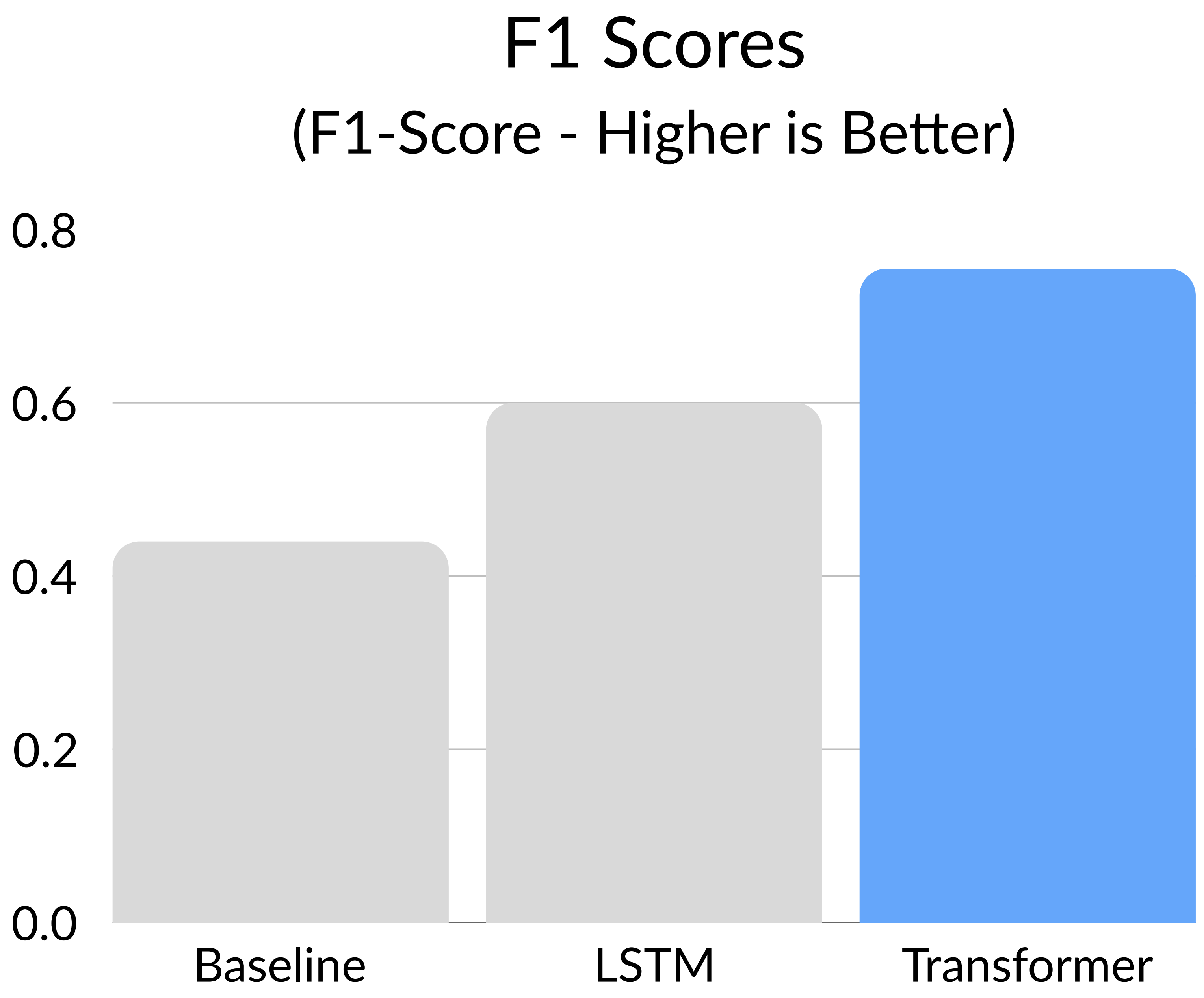
- **Source:** Transcripts of other TV shows run through CIA's pipeline
- **Volume:** 30,000 rows (44% majority class)
- **Language:** Multi-language with English translation
- Moderately imbalanced
- Imbalance solved with back-translation



Step 2: Modelling

Summary:

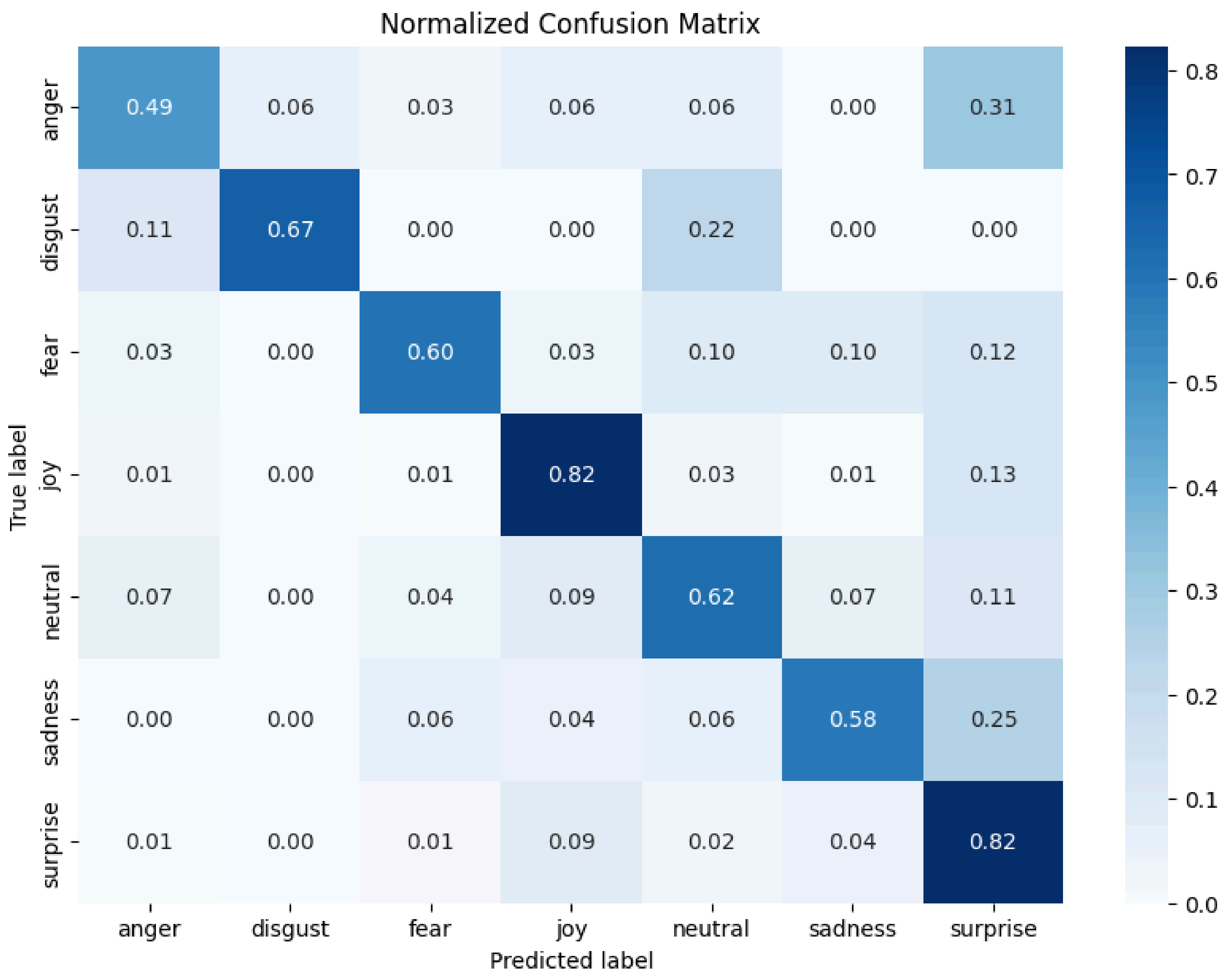
- We checked following model architectures:
  - Traditional ML (Logistic Regression, SVM & Naïve Bayes)
  - RNNs and LSTMs
  - Transformers (BERT, RoBERTa)



The winner: RoBERTa

Summary:

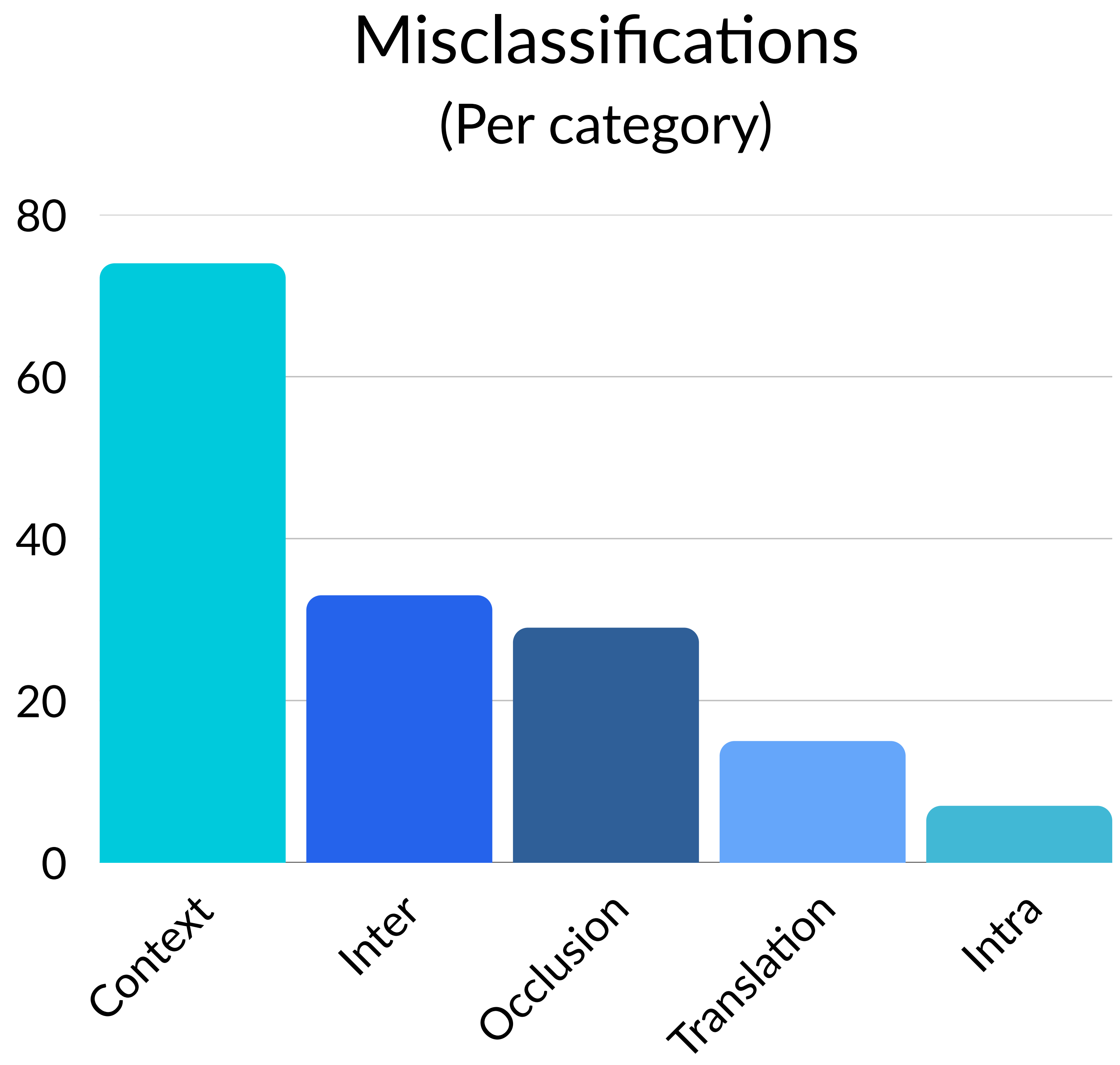
- Base model: RoBERTa-base
- SentencePiece tokenizer
- Highest F1 score 0.755



Error Analysis

Summary:

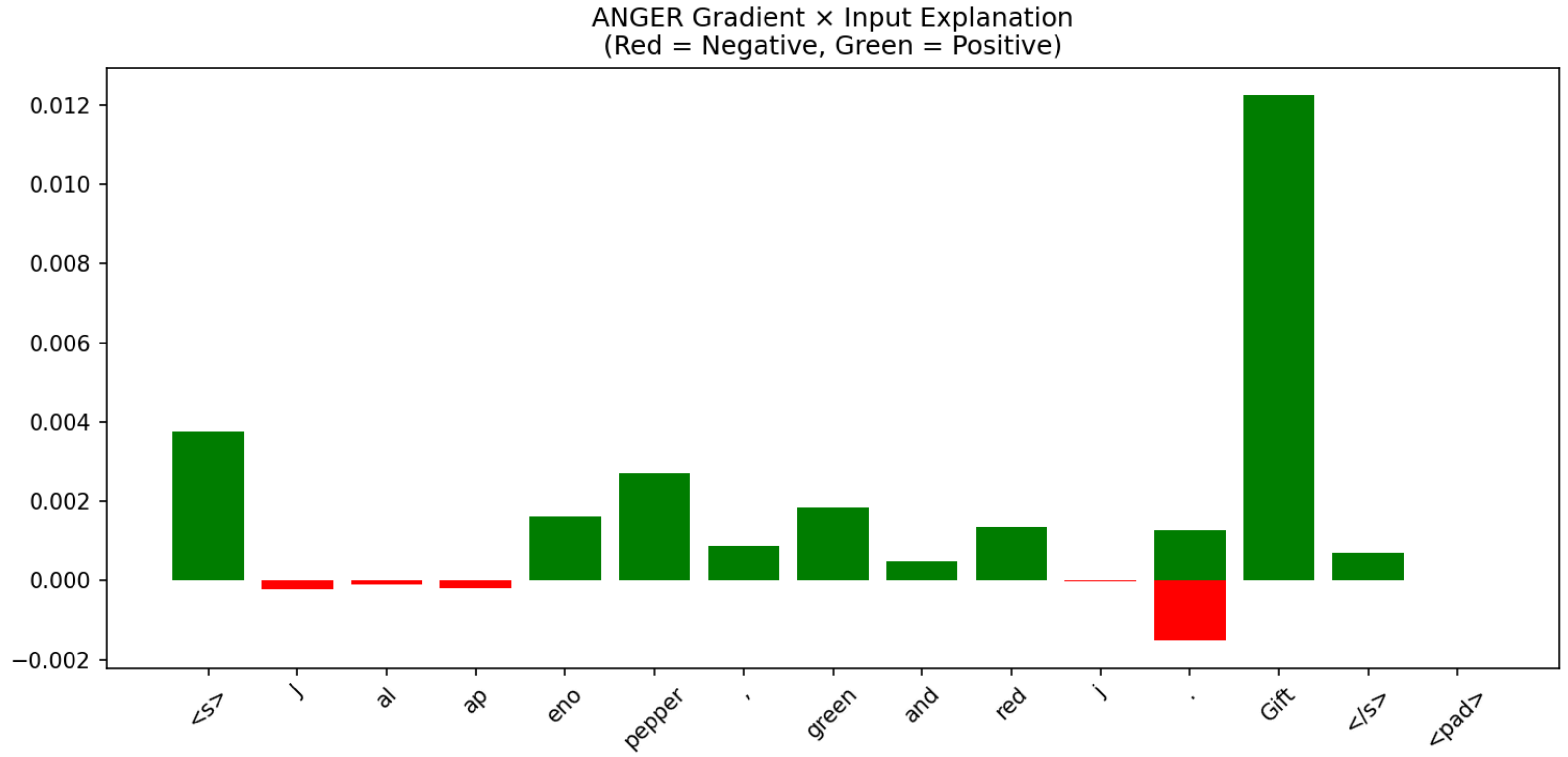
- Went through each mistake, classified casue
- Common mistakes:
  - No context
  - Ambiguous, even with context
  - Singular tokens



Explainable AI

Summary:

- Visualise output
- Understand the black box
- Check if expectations are seen



Limitations & Biases

Biases	Limitations
Bias inheritance from RoBERTa and AI-labeled data	No context leads to short sentences being misclassified
Translation & augmentation noise may alter emotional tone	Model sometimes overly bases predictions on singular tokens
Resource & interpretability limits costly and hard to fully explain	Some struggles with some emotions

Step 4: Final Pipeline

INPUT

Video → AssemblyAI Transcription → Autocorrection with LLM → Translation → RoBERTa Classification → Labelled Data

OUTPUT

Input

Raw video file or YouTube link

Output

A structured data file with timestamps, sentences, and emotion labels.

Conclusions & Future Work

Achievement: Successfully built a pipeline that classifies emotions with a 0.755 F1-score.

Recommendation: Adopt our Transformer-based pipeline. A comprehensive Model Card is provided.

Future Work: Fine-tune on genre-specific data, gathering more Polish language data & training directly on Polish language.

Thank you for your attention!

Questions?