BUas

Classifying Emotions in TV Shows

by

Victor Oorthuis

Kajetan Newes

Louie Daans

Content Intelligence Agency

Our Client

Content
Intelligence
Agency

Our client helps media groups gain comprehensive understanding of their films, series, TV shows, and videos through detailed analysis.

Using AI, we assisted them with the complex task of emotion classification within TV shows

Building an advanced Emotion Classification Pipeline using NLP

Added Value for the client

Content
Intelligence
Agency

Targeted Advertising (more effective ad placement)

Content Recommendation (suggest content based on user emotional preferences)

Audience Segmentation (understand audience segments based on emotional structures)

Our Approach

Transcribe → Translate → Classify Emotion

Transcribe & Translate

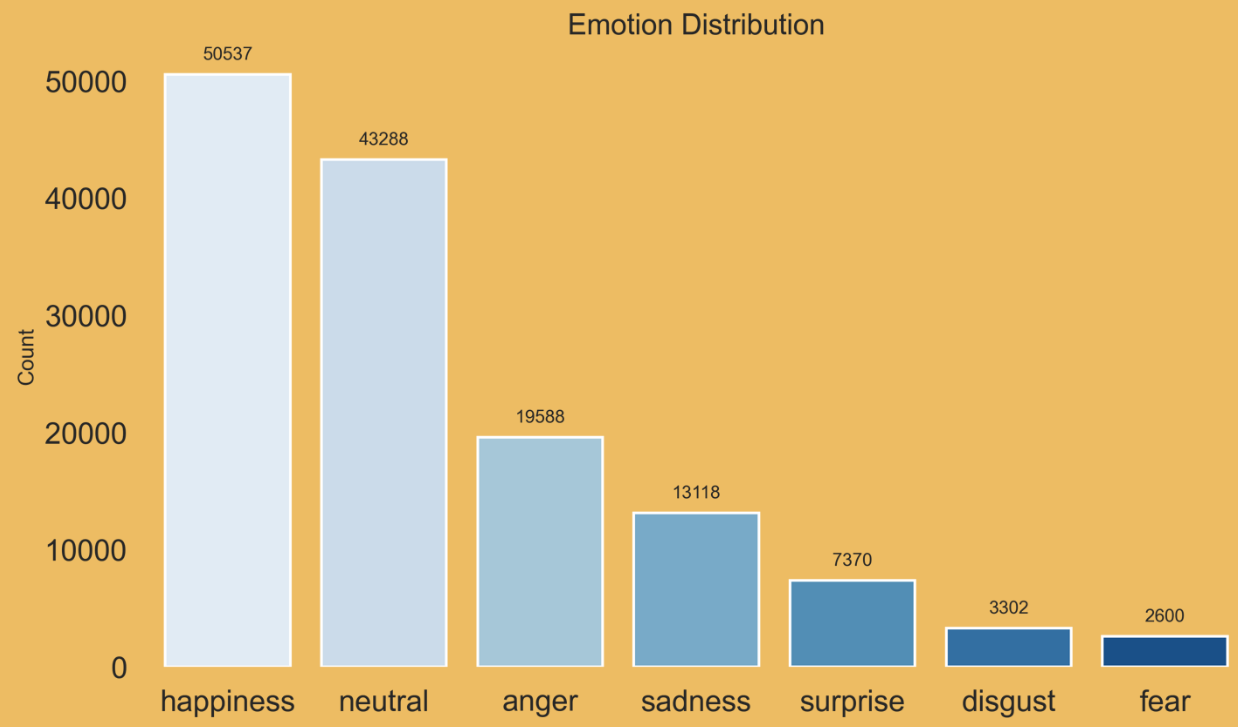
- Whisper transcription

	Whisper	AssemblyAI
WER	27	29

- Machine translation
- French -> English

Data Characteristics

- 140k examples
- Balanced to ~ 18000
- Imperfect test
- Different styles
- "Today was a great day!" - happiness



Data Preprocessing

- Data augmentation – synonyms
- TextAttack

- Feature extraction
- POS, TF-IDF, Embeddings, Sentiment score

- Relabel test set (Manual)

Implemented Models

- Linear Regression
- Naive Bayes
- LSTM
- RNN
- Transformers
  - DistilBERT
  - DistilRoBERTa
  - ALBERT

Implemented Models

- First approach on online data, no transcription test set

Accuracy	Precision	Recall	F1 Score
0.85	0.85	0.85	0.85

Implemented Models

- Second approach on transcription test set (results)

Accuracy	Precision	Recall	F1 Score
0.59	0.62	0.59	0.58

Key Iterations

1. Trying a different model

- DistilRoBERTa → ALBERT
- Did not give any meaningful improvements

Accuracy	Precision	Recall	F1 Score
0.57	0.66	0.57	0.59

Key Iterations

2. Manual validation data annotation

- Re-annotated the emotions in the validation data
- Meaningful improvements in score

Accuracy	Precision	Recall	F1 Score
0.66	0.78	0.66	0.68

Key Iterations

3. Adding more data

- Added more data to the training set
- Let us (almost) reach the client requirements

Accuracy	Precision	Recall	F1 Score
0.75	0.75	0.75	0.74

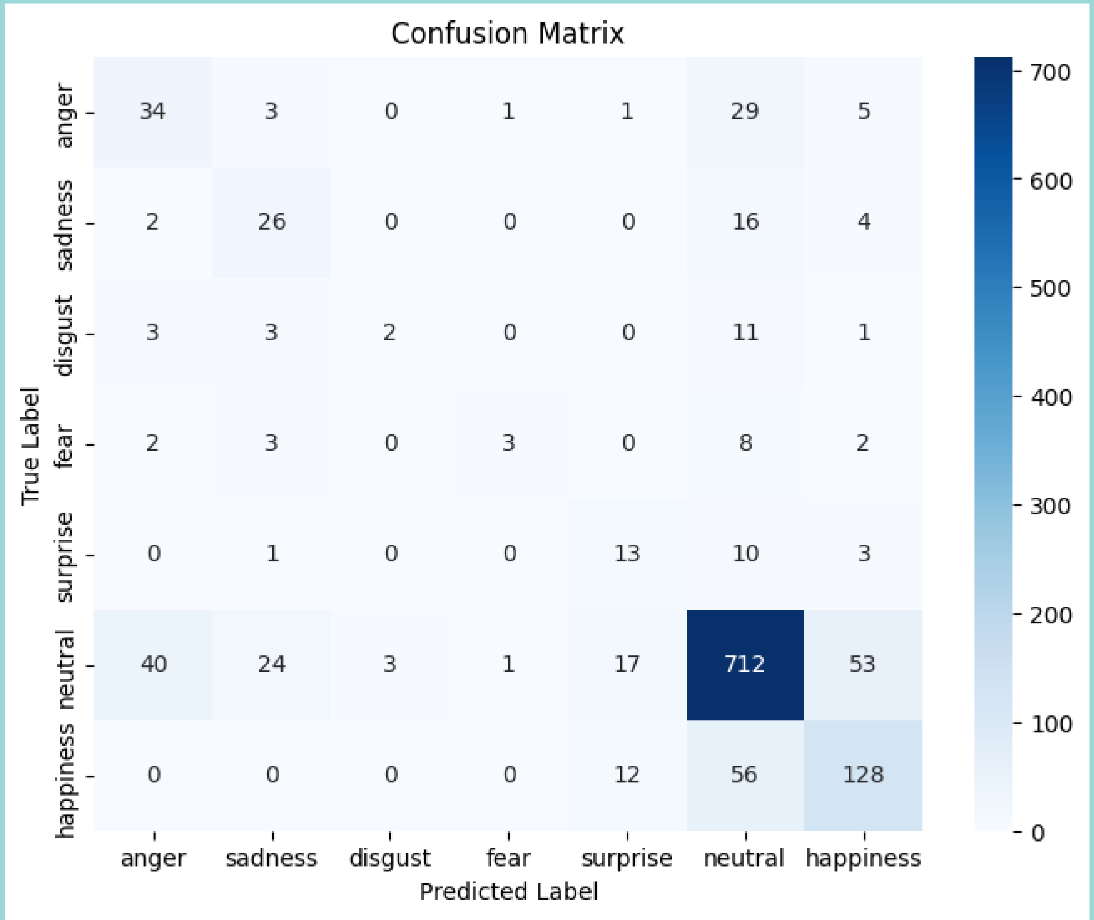
Best Model

- DistilRoBERTa-based transformer model
- F1 Score: 0.74 on our validation transcript
- Carbon Emitted: Approximately 0.2 kg

Model Performance

- Strengths and Weaknesses
  - Good at predicting neutral and happiness
  - Balanced accuracy, precision, and recall
  - Does not do too well with disgust, fear, surprise

Error Analysis



Error Analysis

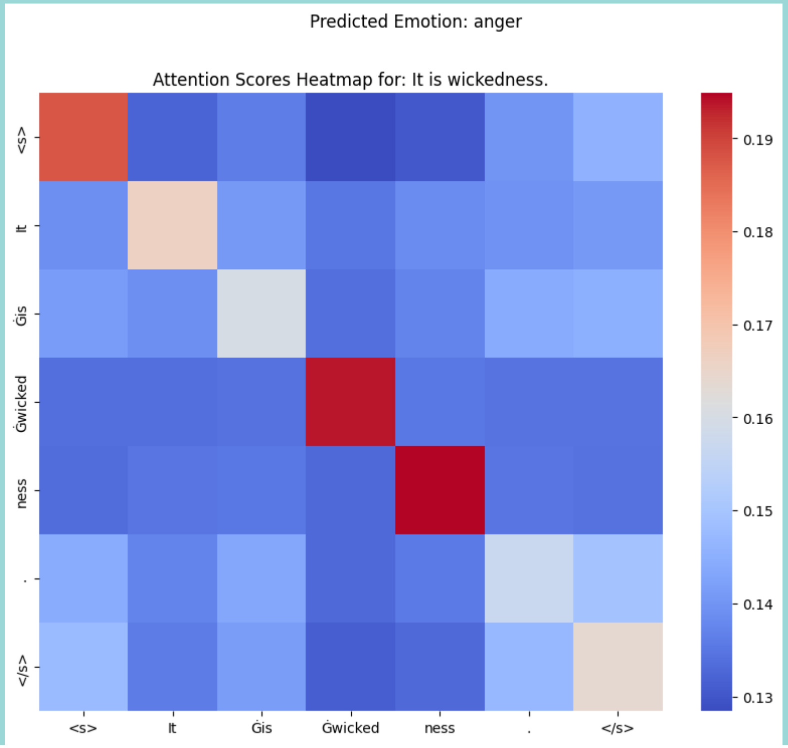
They all make fun of me → classified as happiness

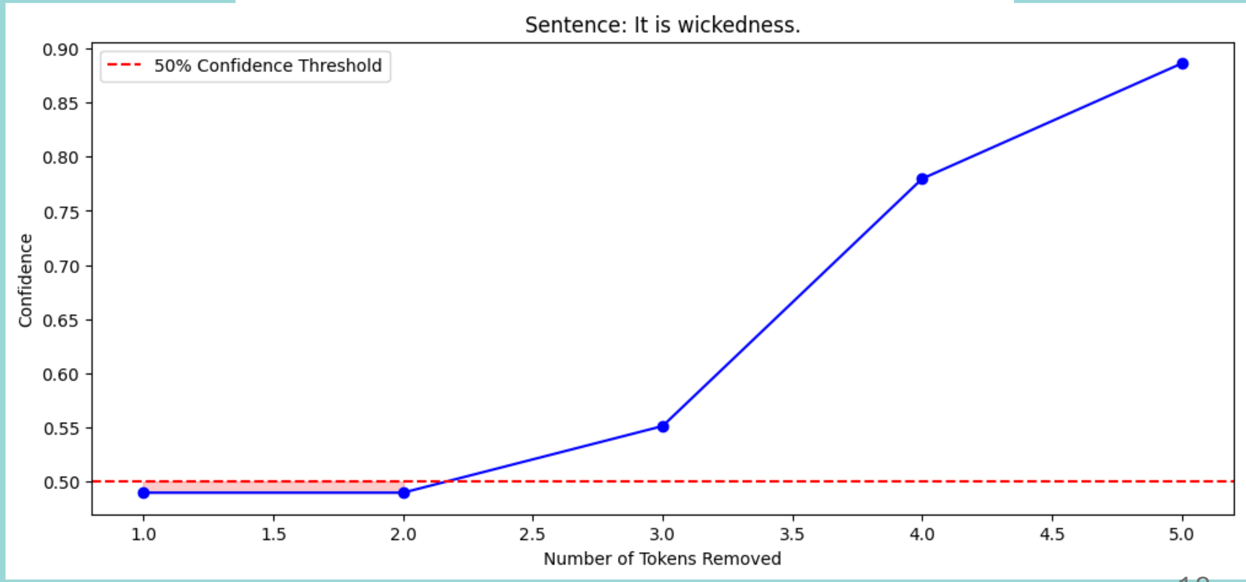
It already smells, it smells like the croissant → classified as disgust

! and ? → Present in all sentences misclassified as surprise

Explainable AI (XAI)

- Strong self-attention
- Lack of robustness





Limitation 1

Single sentences can lack context.

Distance (sec.)	Previous Sentence	Sentence		Emotion
30	' You were excellent. '	' Really. '	→	Neutral / Surprise Happy?

Limitation 2

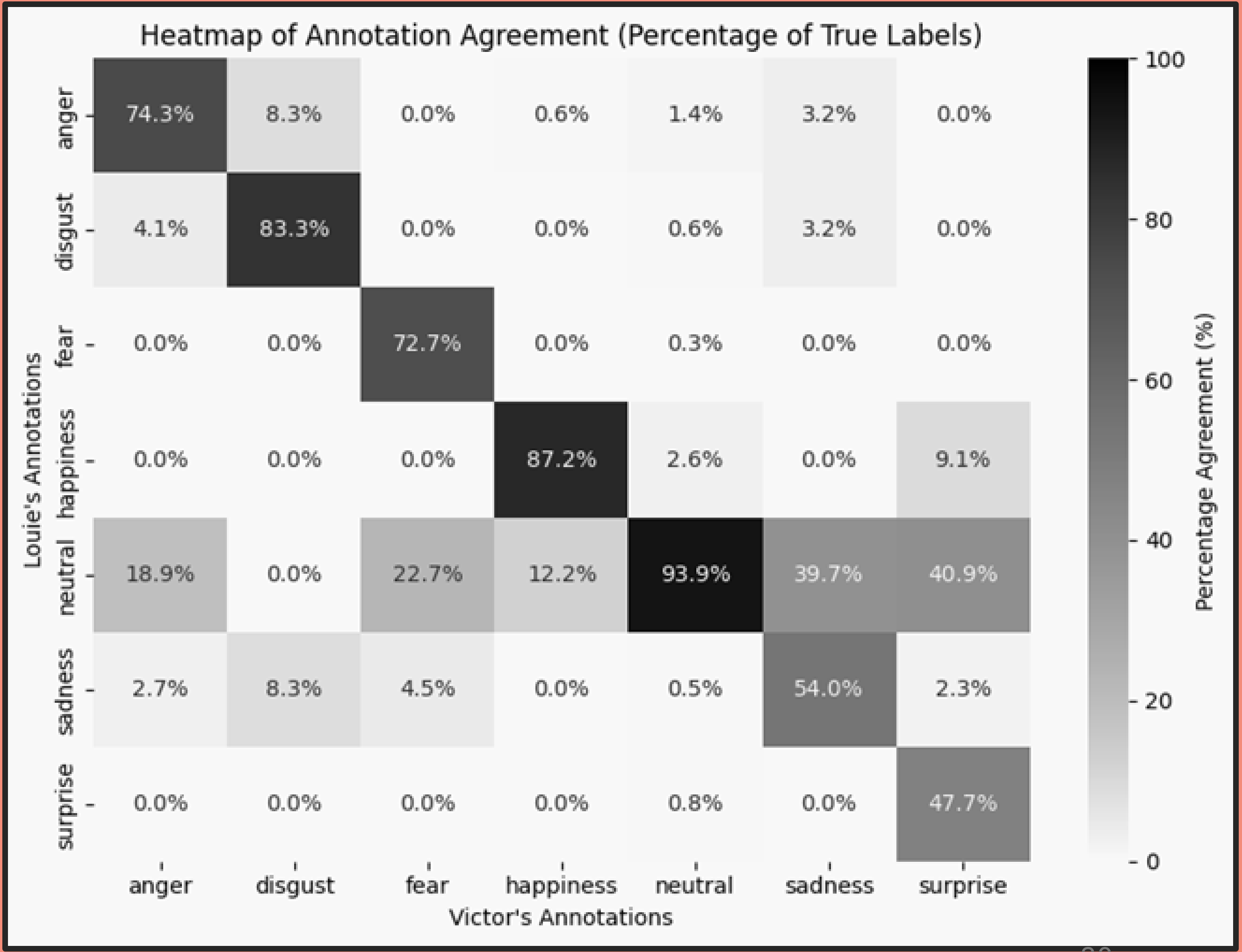
Even for humans, classifying emotions is complicated.

Victor and Louie, each annoted the same transcription mannualy

F1 score: 0.869

Accuracy: 0.873

Human Error Rate: 13%



Limitation 3

Translation isn’t always good.

Ex. 3

Sentence	Pipeline's Translation	Correct Translation
' C[illegible]s. '	' I[illegible]e. '	' Here we go[illegible] problems. '

Limitation 4

Transcription text isn't as expressive as text found online.

Book text

Tweets

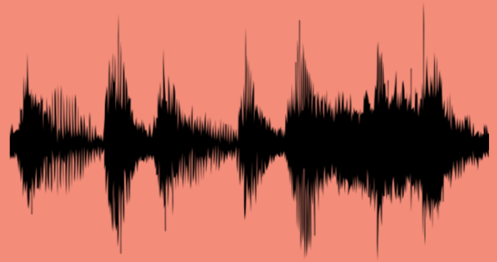
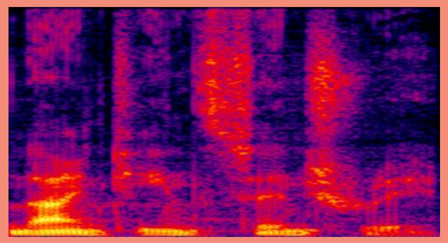
Other online text data

≠

Transcription data

Next Steps and Recommendations

Including extra features.

Audio →  →  → Extract features

Next Steps and Recommendations

Including extra features.

Video →

(computationally expensive)

Next Steps and Recommendations

Including extra features.

Time-positioning?

Person tagging?

Discussion group tagging?

Next Steps and Recommendations

More manual annotation support

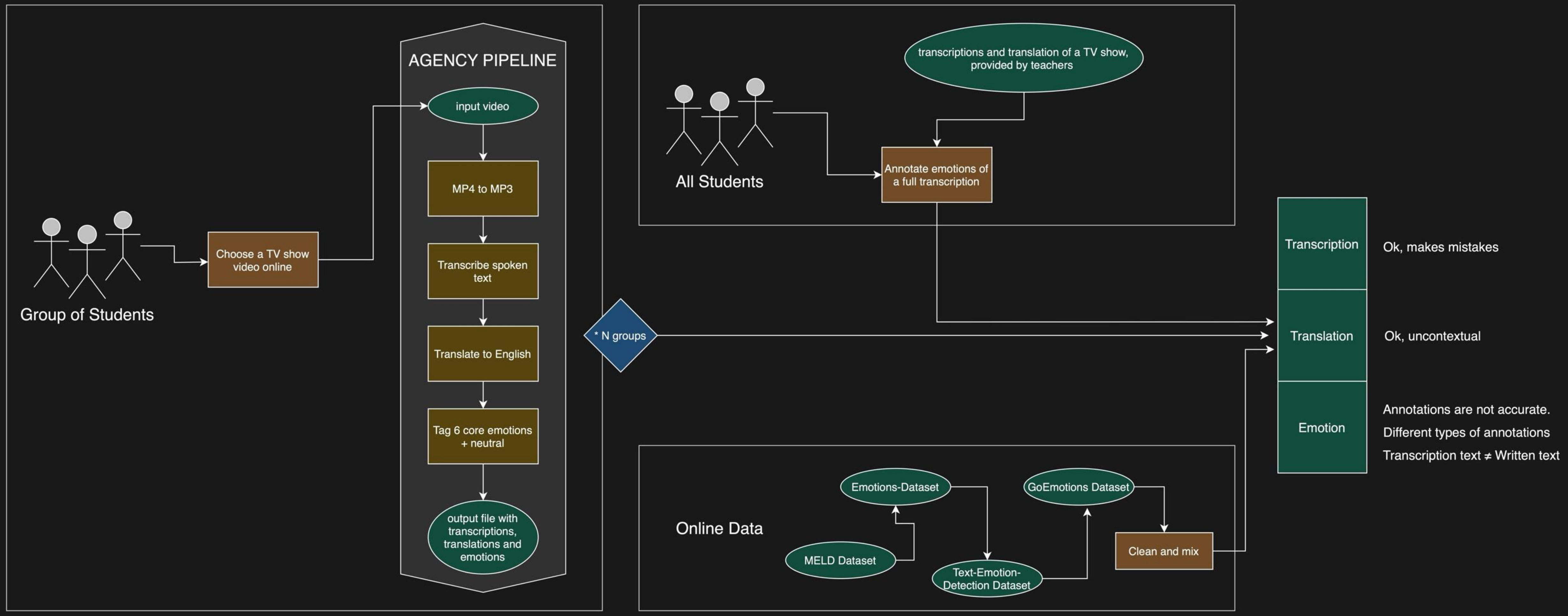
- Focus on manualy annotating transcriptions (one type of source)
- Annotation workspace supported by audio and other tagging features
- Guidelines focused on the client needs (peak emotion moments)

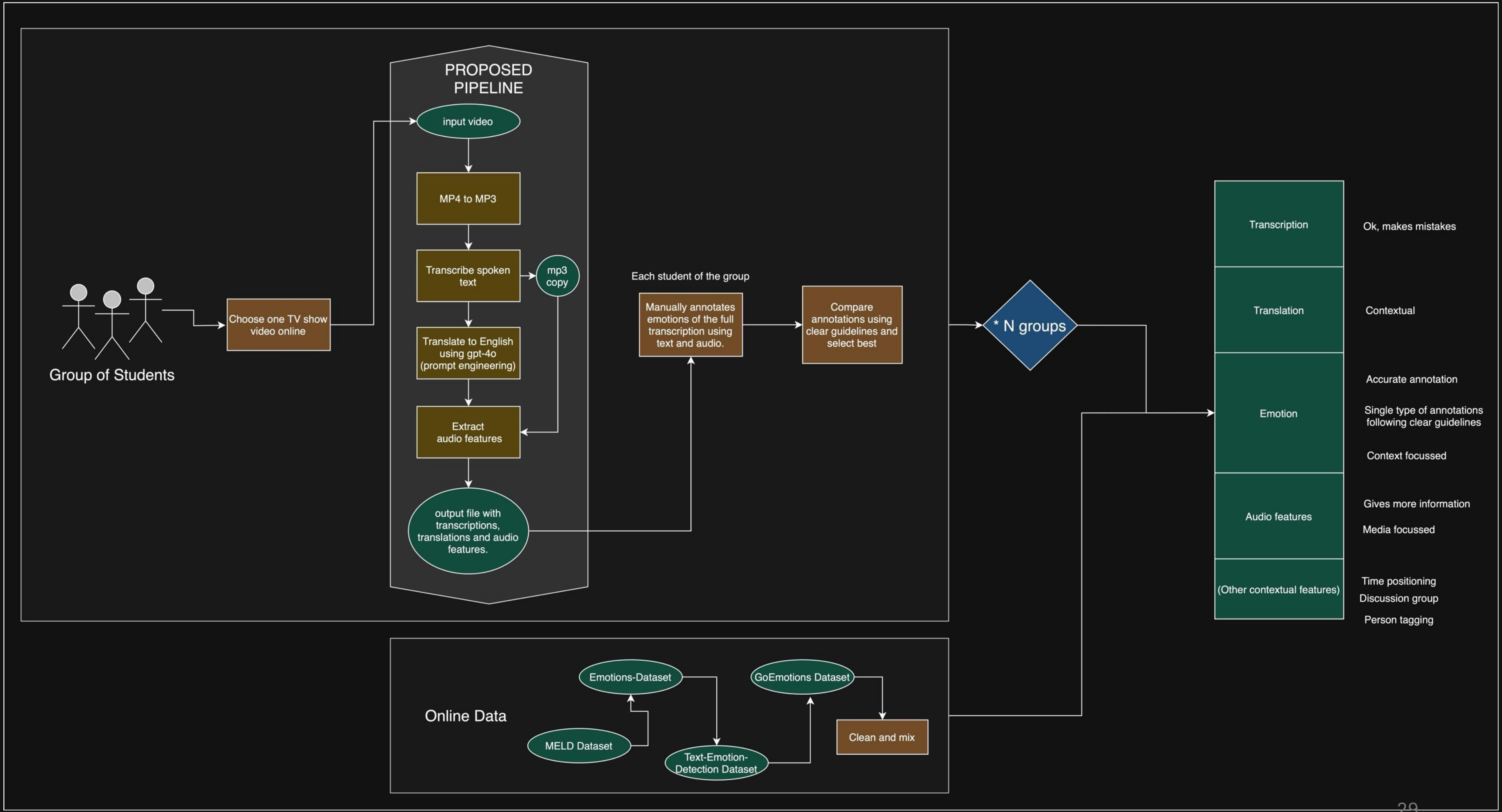
Ethical considerations

- Bias in data (cultural differences)

- Potential for misuse and manipulation (targeted advertisement intruisive)

- Ecology (more data and features requires more energy)


AGENCY PIPELINE
Group of Students
Choose a TV show video online
input video
MP4 to MP3
Transcribe spoken text
Translate to English
Tag 6 core emotions + neutral
output file with transcriptions, translations and emotions
* N groups
All Students
transcriptions and translation of a TV show, provided by teachers
Annotate emotions of a full transcription
Online Data
Emotions-Dataset
GoEmotions Dataset
MELD Dataset
Text-Emotion-Detection Dataset
Clean and mix
Transcription
Translation
Emotion
Ok, makes mistakes
Ok, uncontextual
Annotations are not accurate.
Different types of annotations
Transcription text ≠ Written text


PROPOSED PIPELINE
input video
MP4 to MP3
Transcribe spoken text
mp3 copy
Translate to English using gpt-4o (prompt engineering)
Extract audio features
output file with transcriptions, translations and audio features.
Group of Students
Choose one TV show video online
Each student of the group
Manually annotates emotions of the full transcription using text and audio.
Compare annotations using clear guidelines and select best
* N groups
Transcription
Ok, makes mistakes
Translation
Contextual
Emotion
Accurate annotation
Single type of annotations following clear guidelines
Context focussed
Audio features
Gives more information
Media focussed
(Other contextual features)
Time positioning
Discussion group
Person tagging
Online Data
Emotions-Dataset
GoEmotions Dataset
MELD Dataset
Text-Emotion-Detection Dataset
Clean and mix