

0100

Cheng Jia Peng
Jessica Wong Chai Fei
Yap Hiew Tong
Yap Ern Ying Ainsley

CommentSense

Measuring Engagement Beyond Likes



MONASH
University

L'ORÉAL
MALAYSIA



MUMTEC
MONASH UNIVERSITY MALAYSIA TECH CLUB

MD™
Premier Digital
Tech Institution



MONASH ANALYTICS CLUB

1. Project Overview

PROBLEM STATEMENT

- Brands and creators struggle to measure the quality and relevance of comments at scale.
- Although engagement metrics like shares, saves and etc. tell us how much people interact but not why or whether comments are meaningful for a business.



EXISTING SOLUTIONS & LIMITATIONS

1. Rule-based heuristics + Lexicon / sentiment tools:
misclassify slang, non-English comments, and sarcasm.
(Karulkar et al., n.d.)
2. Topic modeling:
needs manual interpretation and do not map neatly to product categories. (University of Michigan, 2020)

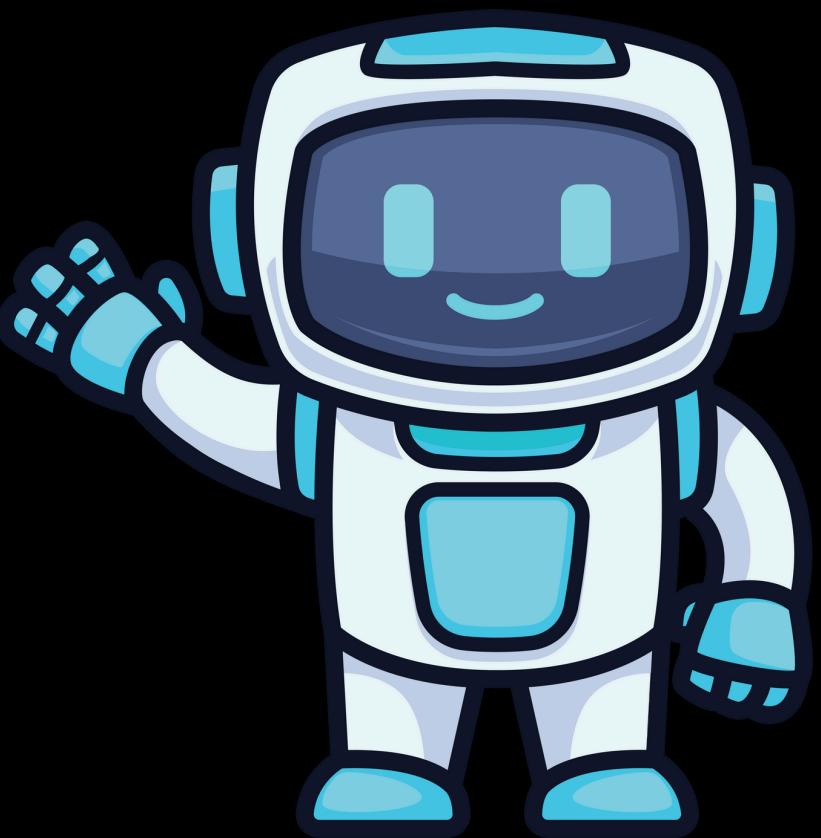
EXISTING SOLUTIONS & LIMITATIONS

3. Pre-trained Transformer classifier:
expensive to run at scale (Tay et al., 2021)

4. Off-the-shelf social listening tools:
They don't combine everything (Lama & Lama, 2025)

OPPORTUNITY STATEMENT

Build a solution with AI enhancement that filters, classifies and quantifies comment quality to give deeper insight into audience engagement.





2. Project Benefits

TARGET AUDIENCE

- 1. Marketing Teams**
- 2. Product Innovation Teams**
- 3. Brand Managers**
- 4. Corporate Communications**
- 5. Community Managers**

ADVANTAGES & UNIQUE VALUE PROPOSITION

1. Beyond Sentiment

- Combines relevance + sentiment + engagement + spam detection
→ QualityScore & Quality Ratio KPI.

2. Context-Aware Relevance

- Uses embeddings, not just keywords.
- Filters out off-topic noise → highlights beauty & L'Oréal-related insights.

3. Spam & Low-Quality Filtering

- Removes bots, links, and emoji spam.
- Delivers a clean view of authentic engagement.

ADVANTAGES & UNIQUE VALUE PROPOSITION

3. Category-Level Insights

- Tags by Skincare / Makeup / Fragrance / Other.
- Reveals which product lines drive meaningful conversations.

4. Simple, Explainable KPI

- Quality Ratio = % of relevant, positive, non-spam comments.
- Clear for executives, useful for benchmarking campaigns.

5. Scalable & Future-Proof

- Runs on English now, ready for multilingual (XLM-R).
- Flexible to integrate cloud GPUs and advanced NLP later.

3. Project Deliverables

0100

FINAL PRODUCT

Dashboard

L'ORÉAL

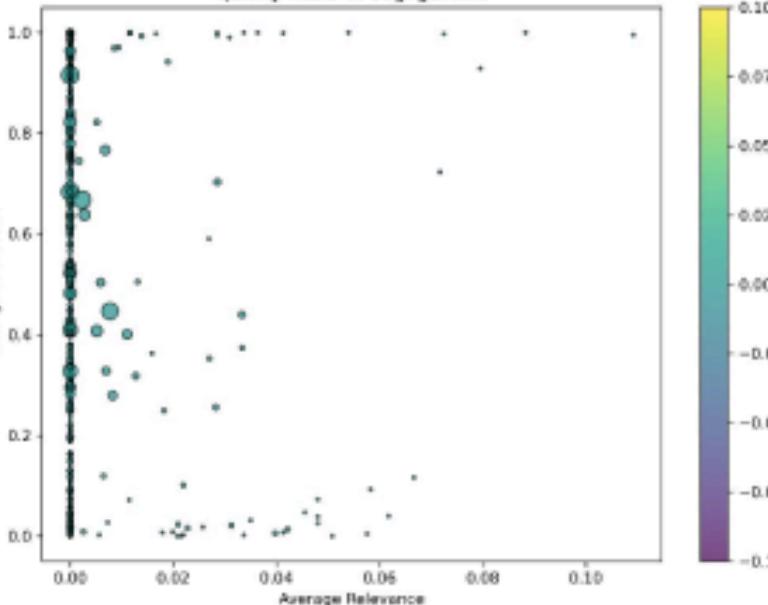

5M+
 Top Comments


0.77
 Quality Ratio

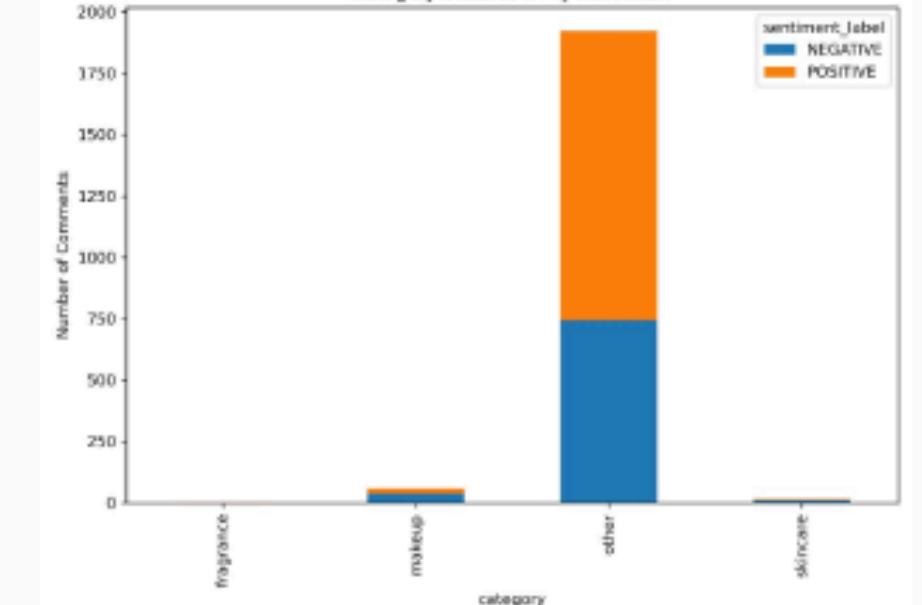

6.2+
 Spam Rate


2.4 (positive)
 Average Sentiment Score

Quality Score vs Engagement



Category Breakdown by Sentiment



category	sentiment_label	Number of Comments
fragrance	NEGATIVE	~10
makeup	NEGATIVE	~20
other	NEGATIVE	~750
other	POSITIVE	~1250
skincare	NEGATIVE	~10

Comment Table Preview

Comment ID	Video ID	Sentiment	Category	Quality Score	Spam ?
3403605	31993	Positive	skincare	High	✓
3957372	88966	Neutral	makeup	Low	✗
1272592	69091	Positive	other	Low	✗
4006296	86018	Negative	makeup	High	✓

DEPLOYMENT



Great article! Very informative.

2 hours ago

REPLY



I agree with all the points made in this post.

3 hours ago

REPLY



I found this very helpful, thank you!

4 hours ago

REPLY



I have a question about this topic.

5 hours ago

REPLY

- 1. Detect relevance**
- 2. Sentiment analysis**
- 3. Category tagging**
- 4. Spam detection**
- 5. Quality scoring → Quality ratio**

PROJECT DELIVERABLES

- Our Prototype:
 - Python AI pipeline + Streamlit dashboard
- Outputs:
 - Quality ratio
 - Sentiment mix
 - Category tags
 - Spam filter text



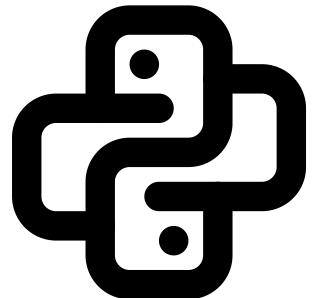
4.

Development Methodology

TOOLS AND TECHNOLOGIES

PYTHON

- Industry-standard for data science and AI development.
- Easy integration with APIs and web framework.



PANDAS

- Essential for data preprocessing, cleaning and structuring large comment datasets.
- Integrates seamlessly with Hugging Face Transformers.
- Makes it easy to transform raw comment data into structured insights.



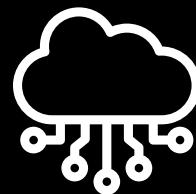
HUGGING FACE

- Provides state-of-the-art pre-trained NLP models without training from scratch.
- Supports tasks like sentiment analysis and processes multilingual comments.



LICENSING AND COSTS

(FUTURE IMPLEMENTATION)



AWS SUBSCRIPTION

Compute (EC2):

- Small CPU instance for lightweight tasks: RM115/month
- GPU instance: RM1150/month

Storage (S3 / RDS): RM45-90/month. depending on dataset size

Total:

- RM160-200/month (small-scale, CPU-only)
- RM1200-1500/month (with GPU for faster inference)

5. Future Plans and Possibilities

FUTURE PLANS & POSSIBILITIES

Multilingual Support:

Extend analysis to more languages (e.g., Malay, Mandarin) to match L'Oréal's diverse global markets

Real-Time Monitoring:

Integrate live comment streams from YouTube, Instagram or TikTok for instant campaign tracking

Voice & Video Analysis:

Expand from text comments to analysing video/audio sentiment to create a full social listening platform.

Overcoming Limitations

Computational Limits:

Running large models and datasets is slower without dedicated GPUs

Solution:

Deploy GPU-backend servers on AWS to accelerate model inference and scale to larger datasets

Overcoming Limitations

Perform Limitation:

Dataset is limited to YouTube comments while other social platforms not yet included.

Solution:

Use more time to develop connectors for Instagram, TikTok and many more.

References

- Karulkar, Y., Vora, D. T., Vaddepalli, S., & Thakur, Y. (n.d.). Are emojis the new words? A sentiment analysis of social media brand conversations. CSUSB ScholarWorks. <https://scholarworks.lib.csusb.edu/jitim/vol33/iss1/5/>
- Lama, D., & Lama, D. (2025, July 7). What is Social Listening? All You Need To Know. Sprinklr. <https://www.sprinklr.com/blog/social-listening/>
- Tay, Y., Dehghani, M., Rao, J., Fedus, W., Abnar, S., Chung, H. W., Narang, S., Yogatama, D., Vaswani, A., & Metzler, D. (2021, September 22). Scale Efficiently: Insights from Pre-training and Fine-tuning Transformers. arXiv.org. <https://arxiv.org/abs/2109.10686>
- University of Michigan. (2020). Topic Modeling and Interpretation. Arts Engine. <https://artsengine.engin.umich.edu/topic-modeling-and-interpretation/>



THANK YOU