
Flashy Consumption and Hidden Feelings on TikTok: Can Large Language Models Identify Teenager’s Emotions and Stress?

Sloka Chava *
Fulton Science Academy
Alpharetta, GA 30009
slokachava@outlook.com

Sahasra Chava
Fulton Science Academy
Alpharetta, GA 30009
siyasahasra@gmail.com

Abstract

Flashy and conspicuous consumption, where in, a latest *haul* or purchase is posted on social media through short-form videos is growing popular. Some of these TikToks showcase makeup purchases (for example, from *Sephora*) or clothing purchases (for example, from *Lululemon*), in hundreds of dollars. While teenagers watching this flashy consumption can experience joy, excitement and vicarious pleasure, they can also potentially experience jealousy, peer pressure and be overwhelmed. In order to understand the emotional responses and potential stress from watching flashy consumption, we first curate 100 TikToks with almost 20 million likes and 100,000 comments. We annotate each TikTok video to identify the emotion experienced, a rating for the intensity of the felt emotion and the reasoning for the emotional response. We generate Large Language Model (LLM) responses using a zero-shot video prompting for a SOTA open-source model (VideoLLaMA2) and a proprietary model (Google Gemini Pro 1.5). We assess the LLM performance using model-based metrics (BERT and ROUGE scores) and a survey of teenagers. We find that VideoLLaMA2 can describe the scene by scene happenings in the video but it doesn’t follow the instructions. Gemini Pro 1.5 does follow the instructions but the performance on identifying the emotion (based on the precision and recall) are low. In addition, the lexical similarity (using ROUGE score) for the explanation are both quite low. However, the semantic similarity (using BERT score) of explanations for the emotional response that teenagers may experience after watching the video are relatively high. Our survey identifies 73.1% of the respondents agree with the LLM explanation but find it verbose.

1 Introduction

The U.S. Surgeon General’s recent advisory on social media and youth mental health highlights the impact of social media on mental health. Up to 95% of youth ages 13–17 report using a social media platform, with more than a third saying they use social media “almost constantly.”² Conspicuous or flashy consumption is a major theme in many TikToks. Many of these TikTok *hauls* feature consumption in the hundreds of dollars and many a time, more than \$1000. These may cause social comparison and strong emotional response in teenagers, especially adolescent girls [Meier and Gray, 2014].

*Both authors contributed equally to the project.

²<https://www.hhs.gov/about/news/2023/05/23/surgeon-general-issues-new-advisory-about-effects-social-media-use-has-youth-mental-health.html>

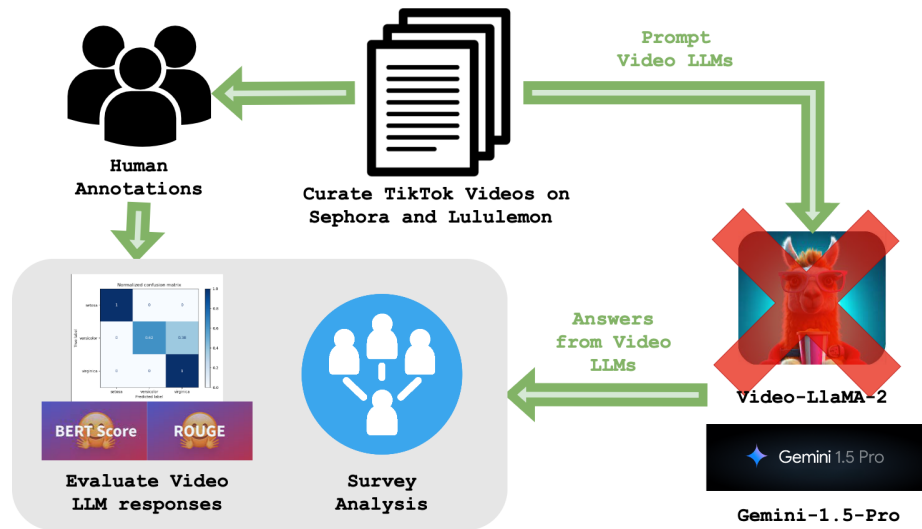


Figure 1: Pipeline to identify and evaluate emotions and stress using large video language models

TikTok and other social media algorithms are typically optimized to increase user engagement and can lead users down rabbit holes further increasing exposure to content that may impact their mental health³. Teens exposed to a barrage of viral *haul* videos may get the impression that happiness is dependent on material things and (a lot of) expensive makeup is normal or necessary (even when it's not affordable to them). This may also lead to lifestyle inflation and unrealistic and inaccurate perception of wealth. These TikToks can normalize conspicuous consumption and subconsciously influence teens to change their spending habits to match what they are seeing online. Emotions drive behaviour [Barlow et al., 2010] and the link may be amplified for teenagers. If an LLM can accurately detect the emotional response to a TikTok, policy makers may be able to ask social media companies to flag TikToks with potential negative emotional responses.

Demszky et al. [2023] explore the use of LLMs in psychology and argue that LLMs could change research in the field of psychology but could also be misunderstood because of biases. Yang et al. [2023] analyze the mental health and emotional reasoning ability of ChatGPT. They find that the LLM showed strong in-context learning ability but has a significant gap with advanced task-specific methods. LLMs that can take video input are recently released: Gemini Pro 1.5 from Google and VideoLLaMA2 that has SOTA open source performance [Zhang et al., 2023]. Motivated by these, our research questions are

- RQ1: Can language models identify emotions and stress from flashy consumption in short-form videos? If so, what are the common emotions that are experienced by teenagers?
- RQ2: Are there differences in the performance of open source and proprietary models in zero-shot video understanding of short-form videos?
- RQ3: Who does the LLM response align with more? A teenager who has taken a Psychology class or one that didn't.

Our main contributions include curating and annotation 100 TikToks with cumulative 20 million likes and 100,000 comments. Along with this, we identify the common emotions experienced by teenagers by watching these TikToks highlighting flashy consumption. For instance, the most common emotion that LLM is correctly identifying is *desire*. VideoLLaMA2 understands the videos but doesn't follow instructions. Gemini Pro 1.5 follows instructions well. Its accuracy score is low, but it aligns more with a teenager who has taken a Psychology class than one who hasn't. Despite being prompted to act as a teenager, it shows low lexical similarity but high semantic similarity.

³<https://www.wsj.com/articles/TikTok-algorithm-video-investigation-11626877477>

2 Data Collection and Pre-processing

We collected short-form videos from TikTok focusing on two specific categories: makeup and clothing. These two hashtags were chosen due to their high volume of content and millions of views. The 100 TikTok videos of Sephora and Lululemon hauls were chosen based (1) hashtags of *Sephora* (*#sephorahauls*) and *Lululemon* (*#lululemonhaul*) (2) At least 10,000 likes for Sephora Hauls and 1000 likes for Lululemon (3) Posted during June 1, 2023 and May 31, 2024 and are under 120 seconds

We use the yt-dlp package [yt dlp, 2024] to download all videos that meet the criteria. We (the authors) manually annotate the videos for the emotional response experienced, the intensity of the response and the reasoning.

3 Prompting

Plutchik [1980] proposed a psychoevolutionary classification approach to identify eight basic emotions. Scherer [2005], Scherer et al. [2013] propose Geneva emotional wheel, a theoretically derived and empirically tested instrument to measure emotional reactions to objects, events, and situations. Mohammad and Turney [2010] create a crowd sourced moderate-sized emotion lexicon for common Words and Phrases using Mechanical Turk and Zad et al. [2021] correct the biases in it. We adapt these methodologies and include a list of shopping and TikTok relevant emotions in our prompt.

The human annotators and both Gemini Pro 1.5 and VideoLLaMA2 were prompted with the below prompt. The human annotators each independently annotated all the 100 TikToks.

In this task you will see a video. Assume the role of a teenager watching this video. List the various emotions, if any, and the intensity of emotions that they are likely to experience watching the video. Intensity should be ranked on a 1-5 scale with 1 being very weak and 5 being very strong. Explain the reasoning. Choose emotions from Excitement, Joy, Inspiration, Confidence, Disappointment, Disapproval, Frustration, Jealousy, Pressured, Overwhelmed, Anxiety, Resentment, Insecure, Sadness, Irritation, Desire, Indifference, Surprise. None, one or two simultaneous emotions are allowed. For each of the emotions identified, the answer should be listed as emotion, intensity, reasoning.

We used Google Colab Pro for running the experiments with VideoLLaMA2 and we used Gemini API calls to generate responses for Gemini Pro 1.5 model.

4 Evaluation

VideoLLaMA2 only identified the products. A typical response is

"The video shows a person unboxing and reviewing different beauty and skincare products. They showcase each product, including a bottle of hair treatment, a container of face wash, a bottle of body lotion, a bottle of hand sanitizer, and a bottle of perfume. The person also demonstrates how to use some of the products, such as the hair treatment and the face wash. Overall, the video provides a detailed look at these items and their packaging."

Gemini correctly responded to the instructions and hence we report the model based scores and evaluation for Gemini only. One of the human annotators took a Psychology class and the other did not. So, we report the evaluation scores separately for the LLM.

The average rating for the TikToks by an annotator that has taken a high school Psychology class is 3.41 with a standard deviation of 0.67 where as the average rating is 4.09 with a standard deviation of 0.84. Gemini Pro 1.5 is within this range with an average rating of 3.90 and a lower standard deviation 0.42. We only consider the emotion with the highest rating for evaluation.

Results in Table 1 indicate that Gemini's response is closer to the annotator who has taken a Psychology class and much less aligned with the annotator who hasn't taken a Psychology class [Santurkar et al., 2023]. ROUGE score shows that the lexical similarity for the explanation are low

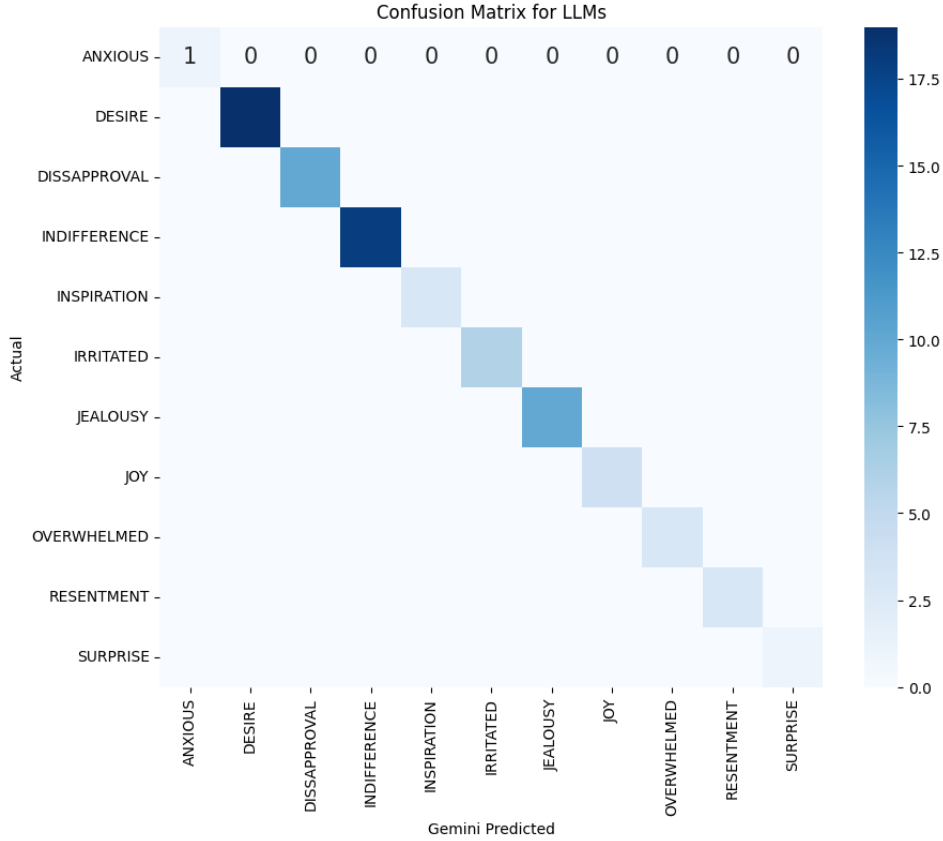


Figure 2: Confusion Matrix for Emotion Prediction

even though we prompted Gemini to assume the role of a teenager. However, the semantic similarity (using BERT score) is relatively high.

Task	Annotator	Accuracy			F1 Score			
Emotion	Pysch	0.2307			0.0356			
Classification	No-Pysch	0.0769			0.0110			
		BERT Score			ROUGE Score			
		P	R	F1	1	2	L	Lsum
Emotion	Pysch	0.8545	0.8740	0.8640	0.1235	0.0229	0.0970	0.0981
Explanation	No-Pysch	0.8583	0.8697	0.8639	0.1275	0.0105	0.0973	0.0972

Table 1: Evaluation of Emotion and Reasoning for Emotion

4.1 Results of Human Evaluations: Survey

We sent a survey with one Sephora Haul and one Lululemon video to 17 teenagers. 13 responded. We asked them to rate whether they strongly agree/ agree / neutral / disagree / strongly disagree with the LLM responses. We find that on average, 73.1% agree or strongly agree with the explanations provided by the LLM. 53.8% felt that the explanations provided by the LLM are verbose.

5 Limitations

We are unaware if Gemini has been previously trained on TikTok videos. Gemini refused to respond to some videos due to safety filters. Both annotators are female teenagers, and their annotations may potentially reflect their biases. Gemini evaluation code was written by the authors but had to be executed by an adult. The survey size is small and doesn't span socioeconomic statuses and cultural differences that influence the context in which teenagers view these videos.

References

- David H. Barlow, Kristen K. Ellard, Christopher P. Fairholme, Todd J. Farchione, Christina L. Boisseau, Laura B. Allen, and Jill Ehrenreich-May. 145 Understanding Behaviors 2: Emotion-Driven Behaviors. In *Unified Protocol for Transdiagnostic Treatment of Emotional Disorders: Workbook*. Oxford University Press, 12 2010. ISBN 9780199772674. doi: 10.1093/med:psych/9780199772674.003.0010. URL <https://doi.org/10.1093/med:psych/9780199772674.003.0010>.
- Dorottya Demszky, David Yang, David S. Yeager, et al. Using large language models in psychology. *Nature Reviews Psychology*, 2:688–701, 2023. doi: 10.1038/s44159-023-00241-5. URL <https://doi.org/10.1038/s44159-023-00241-5>.
- Evelyn P. Meier and James Gray. Facebook photo activity associated with body image disturbance in adolescent girls. *Cyberpsychology, Behavior, and Social Networking*, 17(4):199–206, 2014. doi: 10.1089/cyber.2013.0305. URL <https://doi.org/10.1089/cyber.2013.0305>. PMID: 24237288.
- Saif Mohammad and Peter Turney. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 26–34, Los Angeles, CA, June 2010. Association for Computational Linguistics. URL <https://aclanthology.org/W10-0204>.
- Robert Plutchik. A general psychoevolutionary theory of emotion. In Robert Plutchik and Henry Kellerman, editors, *Emotion: Theory, research and experience, Theories of emotion*, volume 1, pages 3–33. Academic Press, New York, 1980.
- Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cino Lee, Percy Liang, and Tatsunori Hashimoto. Whose opinions do language models reflect?, 2023.
- Klaus R. Scherer. What are emotions? and how can they be measured? *Social Science Information*, 44(4):693–727, 2005. doi: 10.1177/0539018405058216.
- Klaus R. Scherer, Vera Shuman, Johnny R. J. Fontaine, and Cristina Soriano. The grid meets the wheel: Assessing emotional feeling via self-report. In Johnny R. J. Fontaine, Klaus R. Scherer, and Cristina Soriano, editors, *Components of Emotional Meaning: A sourcebook*, pages 281–298. Oxford University Press, Oxford, 2013.
- Kailai Yang, Shaoxiong Ji, Tianlin Zhang, Qianqian Xie, Ziyang Kuang, and Sophia Ananiadou. Towards interpretable mental health analysis with large language models, 2023.
- yt dlp. yt-dlp: a feature-rich command-line audio/video downloader. <https://github.com/yt-dlp/yt-dlp>, 2024.
- Samira Zad, Joshuan Jimenez, and Mark Finlayson. Hell hath no fury? correcting bias in the NRC emotion lexicon. In Aida Mostafazadeh Davani, Douwe Kiela, Mathias Lambert, Bertie Vidgen, Vinodkumar Prabhakaran, and Zeerak Waseem, editors, *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*, pages 102–113, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.woah-1.11. URL <https://aclanthology.org/2021.woah-1.11>.

Hang Zhang, Xin Li, and Lidong Bing. Video-LLaMA: An instruction-tuned audio-visual language model for video understanding. In Yansong Feng and Els Lefever, editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 543–553, Singapore, December 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.emnlp-demo.49. URL <https://aclanthology.org/2023.emnlp-demo.49>.