It's not Sexually Suggestive; It's Educative | Separating Sex Education from Suggestive Content on TikTok videos

Enfa George Mihai Surdeanu

Computational Language Understanding Lab at the University of Arizona

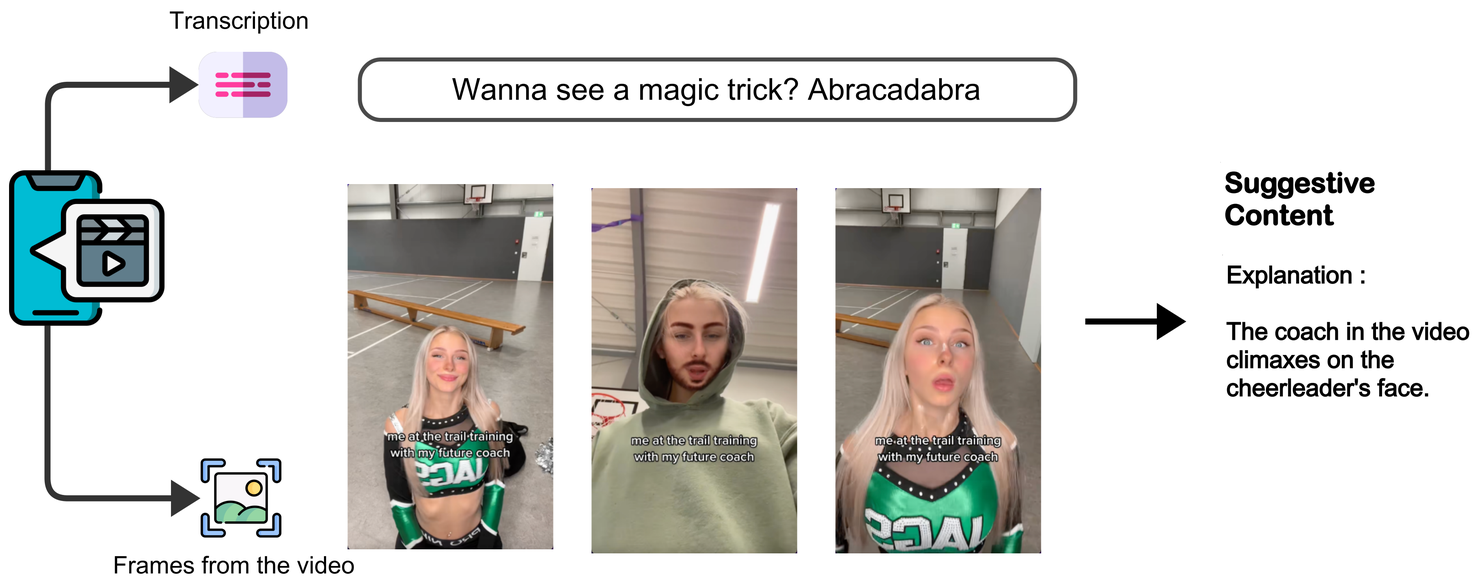

Figure 1: Two examples from the dataset. Credit: Icons from Flaticon.com

Short Summary

We introduce SexTok, a multi-modal dataset composed of 1000 TikTok videos labeled as *sexually suggestive (from the annotator's point of view), sex-educational content, or neither*. Our dataset contains video URLs, perceived gender expression, and audio transcription. To validate its importance, we explore two transformer-based models for classifying the videos. Our preliminary results suggest that the task of distinguishing between these types of videos is learnable but challenging. We invite further work on this task.

Motivation

- Sexual media content influences children's attitudes and contributes to the formation of adversarial beliefs about sex.
- TikTok offers a promising avenue for providing comprehensive and accessible sexual health information in a convenient, private, and inclusive manner.
- Incorrect classification or incorrect mass reporting results in video/account bans. LGBTQ+ sex-ed content creators are especially vulnerable to this.

Dataset

We created a new Tiktok account and manually browsed and collected 1000 Tiktok videos. The videos were downloaded without watermarks to minimize visual distractions. OpenAI Whisper was used to transcribe the audio. We share some interesting dataset statistics below.

Parameter	Sugg	Edu	Others	Total
Mean	16.46	231.18	82.18	98.83
Median	14.00	171.50	31.00	33.00
Std	14.33	220.81	126.37	156.08

Table 1: Mean, Median, and Standard Deviation of words present in video transcripts. Words were tokenized using the NLTK package. Sugg stands for Suggestive, and Edu stands for educative. Suggestive videos tend to be significantly shorter than the other classes.

Parameter	Sugg	Edu	Others	Total
Mean	8.96	66.41	39.99	39.06
Median	7.86	50.80	28.30	23.16
Std	3.82	56.92	37.88	42.90

Table 2: Mean, Median, and Standard Deviation of videos in the dataset in seconds. Sugg stands for Suggestive, and Edu stands for educative. Suggestive videos tend to be significantly shorter than the other classes.

Experiment Results

Text Classification : Fine-tuned *bert-base-multilingual-cased*
Video Classification : Fine-tuned *MCG-NJU/videomae-base*

Group	Suggestive	Educative	Others
Majority	0.00	0.00	0.60
All Text	0.30 ±0.14	0.83 ±0.01	0.80 ±0.02
Non-empty Text	0.38 ±0.03	**0.84** ±0.01	**0.81** ±0.02
Video	**0.55** ±0.02	0.63 ±0.13	0.72 ±0.15

Table 3: Overall F1 of each class label with the average and standard deviation of three random runs. Text-based classification gives a higher F1 for educative content when transcription is present, but suggestive content is detected best in videos where educative content is misclassified higher.

Discussion

Common error types found are as follows:

- Text classification
  - Audio unrelated to class label
  - Use of context clues and euphemism
  - No or partial transcription
- For video classification
  - Educative videos mistakenly classified as "others" and vice versa had the same format: a person looking at the camera and speaking.
  - Majority of misclassified suggestive videos featured fully or mostly clothed people in most or all of the video frames.

The results highlight the complexity and multimodal nature of the task. Accurately categorizing necessitates a nuanced understanding of contextual cues, subjectivity, evolving language, and robust algorithmic solutions.

The dataset and related code can be found at *https://github.com/enfageorge/SexTok*

Acknowledgements

This work was partially funded by the LGBTQ+ Grad Student Research Funds by The Institute for LGBTQ Studies at the University Of Arizona. We deeply appreciate the invaluable contributions of Shreya Nupur Shakya throughout this work.