Title: Toxicity detection using video captions vs. transcripts appendix for “Toxic politics and TikTok engagement in the 2024 U.S. election”
Authors: Ahana Biswas (1), Alireza Javadian Sabet (1), Yu-Ru Lin (1)
Date: August 20th, 2025
Note: The material contained herein is supplementary to the article named in the title and published in the Harvard Kennedy School (HKS) Misinformation Review.

# Appendix E: Toxicity detection using video captions vs. transcripts

We selected 300 (100 for each partisan alignment) videos randomly and ran the toxicity classification on only video captions. The results are compared to toxicity scores using full video transcripts (see Appendix H for manual validation). As shown in Figure E1, using only captions is not enough to capture the toxicity (including subtypes) levels in the video.

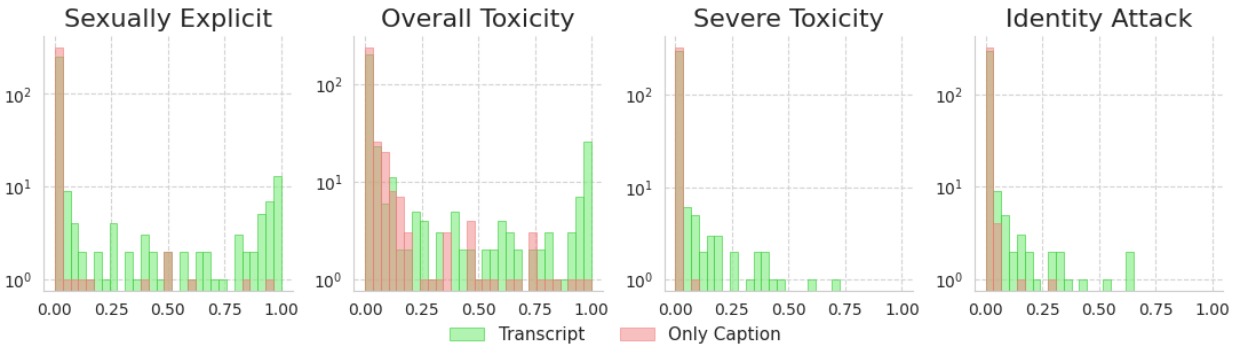


***Figure E1. Toxicity distribution using only captions vs. video transcripts.*** *Using only captions failed to predict toxic content levels present in the video.*