

# EDTT: a Clinically-Grounded and Multimodal Feature Annotated Dataset of Pro-Eating Disorder Content on TikTok

EDEN SHAVEET\*, Cornell Tech, USA

THALIA VIRANDA, Cornell Tech, USA

TANZEEM CHOUDHURY, Cornell Tech, USA

TikTok has emerged as a key venue for the circulation of pro-eating disorder (pro-ED) content. Identifying such content remains challenging as it is often deliberately elusive and conveyed through multimodal cues that go beyond text. Existing datasets of ED-related TikTok posts rely primarily on hashtags, descriptions, or titles for identification, leading to potentially irrelevant or non-representative samples. To address this gap, we introduce the first clinically grounded dataset of pro-ED TikTok posts annotated for multimodal features, curated through text-based retrieval and recommendation-based surfacing. We identified 58 pro-ED hashtags and terms from prior research and retrieved videos using two approaches: (1) API retrieval: Using TikTok’s Research API, we extracted up to 10 videos per term per month from July 2024-June 2025 (3,511 total). We screened a randomized sample of 60 videos, retaining those that reflected eating disorder psychopathology based on the Eating Disorder Examination (EDE) subscales (e.g., food avoidance, weight/shape concern) or diagnostic items (e.g., compensatory behaviors). (2) Inductive “For You Page” (FYP) retrieval: To capture elusive or textless content, we created two new TikTok accounts. Over a week, we used each account to search 3-5 hashtags daily and fully view/“like” 10 videos per term to train the recommendation algorithm. In subsequent days, we reviewed For You page (FYP) feeds until 30 qualifying videos per account (60 total) were identified using the same EDE criteria. Across both methods, we collected and annotated 120 videos for EDE features and pro-ED medium (visual, audio, text) using consensus coding between co-authors. After removing videos taken down by TikTok and nine excluded for lack of consensus, 85 remained for analysis, which include the following fields: Video ID, Video Description, Transcript, On Screen Text, Summary, EDE Subscale Items, EDE Diagnostic Items, Video Elements, Audio Elements, Audio Tempo, Audio Valence, Audio Arousal, and Harmonic Ratio. The most commonly expressed ED psychopathology includes preoccupation with shape or weight and references to eating avoidance. Other common content characteristics includes the use of on-screen text for scenario re-enactment, melancholic music, elusive symbolism via emojis or coded language, and “what I eat in a day” trends. This resource will be continually updated and lays the groundwork for developing more robust detection models.

## Conference Venue

This abstract is prepared for presentation at the **47th Annual Meeting & Scientific Sessions of the Society of Behavioral Medicine (SBM)**, held **April 22–25, 2026 in Chicago, IL, USA**.

---

\*Corresponding author.

---

Authors’ Contact Information: Eden Shaveet, [ems486@cornell.edu](mailto:ems486@cornell.edu), Cornell Tech, Information Science, New York, NY, USA; Thalia Viranda, [tv74@cornell.edu](mailto:tv74@cornell.edu), Cornell Tech, Information Science, New York, NY, USA; Tanzeem Choudhury, [tanzeem.choudhury@cornell.edu](mailto:tanzeem.choudhury@cornell.edu), Cornell Tech, Information Science, New York, NY, USA.