Watching What We Eat: Information Quality and Body Image in Diet-Related YouTube Videos

Maddalena Ghiotti
KTH Royal Institute of Technology
Stockholm, Sweden
ghiottimaddalena@gmail.com

Daniela Paolotti
ISI Foundation
Turin, Italy
daniela.paolotti@isi.it

Yelena Mejova
ISI Foundation
Turin, Italy
yelenamejova@acm.org

Abstract

The widespread use of social media—particularly image- and video-based platforms—has turned them into key sources of both normative and informational content related to health and diet. This may contribute to the development of disordered eating behaviors or, potentially, eating disorders. This study uses mixed-methods analysis applied to 3129 YouTube videos about diet and weight loss in order to quantify the level of risk of low-quality information and heightened focus on the body image. We adapt three quality measurement frameworks from the literature – PRHISM, HONcode and SMEC – to the online video context, perform manual annotation of a sample of the data, and design an LLM content characterization pipeline to score the videos on the quality of their content and the focus on the human body. Surprisingly, we find that videos around personal storytelling and mindset & motivation are associated with higher-quality content, whereas supplement reviews (arguably more medically sensitive ones) are not. Further, the body-related mentions of weight measurement and negative body image are associated with an increased viewership, whereas the mentions of positive body image are associated with an increased engagement rate in terms of likes and comments, but not viewership. Worryingly, we find a cluster of videos categorized as "music" which promote the dietary supplements Mitolyn and the injectable weight loss drug Mounjaro. As video-based platforms grow in popularity, particularly among younger audiences, studies such as the one presented here are essential for developing empirically grounded tools to enhance the detection of harmful content and inform more effective moderation practices.

CCS Concepts

• **Information systems** → **Social networks**; • **Applied computing** → **Consumer health**; • **Computing methodologies** → Natural language processing.

Keywords

Information Quality; Eating Disorders; Diet; Body image; Social Media; Video Platforms; YouTube

ACM Reference Format:
Maddalena Ghiotti, Daniela Paolotti, and Yelena Mejova. 2026. Watching What We Eat: Information Quality and Body Image in Diet-Related YouTube Videos. In *International Conference on Information Technology for Social Good (GoodIT '26), September 02–04, 2026, Pisa, Italy.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3794786.3830741



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
GoodIT '26, Pisa, Italy
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2483-1/26/09
https://doi.org/10.1145/3794786.3830741

1 Introduction

Social media platforms have become central for information consumption and entertainment, particularly among younger people, with video-based platforms such as TikTok, YouTube, and Instagram experiencing rapid growth in both usage and influence. These environments increasingly shape how users learn about body norms, health behaviors, and lifestyle practices. Visual and short-form content, often delivered through algorithmically curated feeds, amplifies exposure to idealized body images and normative narratives around health and attractiveness. Diet-related content represents a particularly prominent and influential category within social media ecosystems, yet content analyses consistently reveal a lack of scientific rigor in online content, with popular videos often promoting rapid weight-loss strategies, restrictive diets, or anecdotal advice without expert validation [3, 32] and prioritizing aesthetic appeal over informational accuracy [37].

Beyond issues of quality, there is substantial evidence linking exposure to such content with negative health outcomes. Studies have documented associations between diet-related social media use and increased body dissatisfaction, disordered eating behaviors, and harmful coping strategies [29]. Content genres such as *fitspiration* and *thinspiration* further exacerbate these risks by normalizing extreme body ideals and unhealthy practices [27, 33]. Emerging evidence suggests that younger users are especially susceptible to these dynamics, as they encounter repeated and often passive exposure to body-focused media through personalized recommendation systems [8, 21]. These findings underscore the dual challenge posed by diet-related social media content: its widespread reach and its potential to misinform and harm users. However, most current research focuses on the most popular content produced in a short period [3, 26, 32] or that already being recommended by the algorithm [37], instead of surveying the broader ecosystem of the diet-related content on these platforms.

In this study, we aim to systematically assess the quality and potential risks of diet-related content on YouTube, the most popular video platform worldwide. To address the limitations of prior work, which has often relied on small-scale manual annotation and focused on popular content, we propose a scalable methodology combining automated transcript analysis with structured evaluation frameworks. Specifically, we leverage large language models (LLMs) to annotate video transcripts according to the quality criteria and risk indicators developed and validated in the communications science literature. We formulate three research questions:

RQ1: What is the information quality of diet-related content on YouTube and the extent of its focus on the body shape?

RQ2: What video content attributes are associated with lower information quality and higher body focus?

RQ3: To what extent does the quality and body focus relate to the user engagement with the content?

By integrating computational methods with validated assessment standards, this study contributes to a more comprehensive understanding of the informational landscape surrounding diet and health on social media.

2 Related Work

The evaluation of information quality has been a longstanding concern across multiple domains, leading to the development of standardized frameworks, survey instruments, and annotation methodologies. In the context of health information, tools such as the Health On the Net (HONcode) [4] and the more recent Principles for Health-related Information on Social Media (PRHISM) [10] provide structured criteria for assessing credibility, transparency, and evidence-based communication. These frameworks have been applied in empirical studies to evaluate online content, revealing widespread deficiencies in quality and adherence to best practices [9]. Surveys have also played a key role in understanding how users seek and evaluate information, with large-scale studies demonstrating high reliance on digital sources and limited verification with healthcare professionals [13]. Complementary research highlights users' difficulties in assessing credibility, particularly in environments characterized by information overload and heterogeneous sources [13, 19, 21]. Instead of surveying content consumers about their perceptions, content analysis methods often employ detailed coding schemes to operationalize relevant dimensions such as accuracy, completeness, and source authority [5, 15, 30, 31]. While these approaches provide nuanced insights, they are typically labor-intensive and difficult to scale, motivating the exploration of automated techniques that can extend quality assessment to larger and more diverse datasets.

A substantial body of work has focused specifically on evaluating the quality of health and diet-related content on social media platforms, with increasing attention to video-based media. Early studies of YouTube identified significant gaps in the accuracy and reliability of nutrition-related videos, particularly among those with high view counts [3]. Subsequent research has expanded this analysis to other platforms, revealing consistent patterns of low-quality information and a lack of expert involvement. For instance, analyses of TikTok content demonstrate that highly engaging diet-related videos often prioritize visual appeal and personal narratives over evidence-based guidance [37]. Yet other content may promote conditions acknowledged to be dangerous to health, such as that circulating within pro-anorexia communities [26]. These studies frequently rely on manual annotation using predefined criteria, although some have begun to incorporate semi-automated approaches to handle larger datasets. In parallel, research has examined the potential harms associated with exposure to such content, linking it to body dissatisfaction, negative affect, and disordered eating behaviors [16]. The convergence of findings across platforms and methodologies underscores the systemic nature of the problem. However, the quality assessment of these platforms has focused on the most popular content, neglecting the "long tail" of the posted videos that are available to the consumer. Further, the studies alternate between small samples of popular content examined manually (e.g. [37]) and shallow larger samples on which sentiment analysis tools are run ([26]). In this work, we use a hybrid data collection approach to sample both longitudinal posting activity around the topic of dieting, as well as to capture the most popular actors, and use the latest LLM technology to deeply examine their quality.

Recent advances in large language models (LLMs) have opened new possibilities for the automated annotation and analysis of social media data. Unlike traditional machine learning approaches, which require task-specific training data and often struggle with contextual nuance, LLMs demonstrate strong performance in zero-shot and few-shot settings. The prompting strategies employ the decomposition of tasks into interpretable sub-dimensions, use of rubric-based scoring, and iterative refinement to improve consistency and reliability [18, 20, 25, 34]. For example, models such as GPT-4 have been successfully applied to the evaluation of medical content, showing high agreement with expert assessments [17, 18]. These capabilities enable the adaptation of established evaluation frameworks into prompt-based annotation schemes, allowing LLMs to assess multiple dimensions of content quality and risk simultaneously. In the context of social media, this is particularly valuable for analyzing transcript data derived from videos, which can be lengthy, unstructured, and linguistically diverse [7]. LLMs can capture implicit meanings, identify misleading claims, and evaluate tone and framing, outperforming conventional NLP pipelines that rely on surface-level features [28, 36]. Moreover, their flexibility allows researchers to integrate multiple annotation tasks within a single workflow, improving efficiency and consistency. As a result, LLM-based approaches, which we also use in this work, represent a significant advancement over traditional methods, offering both scalability and depth in the analysis of complex, real-world data.

3 Data & Methods

3.1 Data Collection

We collect data using two distinct procedures to obtain samples of both the videos uploaded to the platform on a daily basis (which we refer to as the daily sample) and the videos that gained popularity during the study period (the popular sample).

Daily sample. Between January 8 and June 7, 2025, we perform a continuous collection via the YouTube Data API [1] by running daily retrievals of 500 most recent diet and weight-loss videos uploaded the previous day. We combine search keywords (*diet, dieting, diets, weight loss, weightloss, fasting, nutrition*) in OR logic, and follow the methodology of [22] to filter the results by the US region and English language (interestingly, the API still returned videos posted by channels identifying from outside of the US, as we show in the Results section). Specifically, we filter the videos using several criteria: English as the default audio language, duration between one and sixty minutes, and at least one keyword match in both the title and transcript (or at least three transcript matches if the title contained none). For each selected video, we collect metadata at upload time and then track it periodically — daily for the first week, then weekly up to three months — capturing evolving engagement statistics and a sample of comments. Finally, because the Data

API does not provide video transcripts, we extract them using the YouTube Transcript API [2], which retrieves both manual and auto-generated captions. We automatically exclude any video lacking a transcript, map category IDs to human-readable names, drop records with malformed fields or missing engagement statistics. This approach yielded 2870 videos with complete metadata.

Engagement. The engagement metrics associated with the collected videos include the number of views, likes, and comments. These statistics are captured at the time of the video's collection. Thus, we consider the latest most complete available statistic we find available during the periodic re-collection, which is 70 days from upload for the daily sample.Further, we compute the engagement rate, which is the sum of the number of likes and comments, normalized by the number of views.

Popular sample. To ensure that highly popular videos were not missed by the above daily search, we conduct a supplementary retrospective collection between November 10 and December 1, 2025. We target the same five-month window and query the API, but order results by view count across sliding time windows of one day, ten days, and one month, in batches of 50 (the API maximum per request). We then take the union of the three results. After the filtering steps as above, this popular sample yielded 384 videos. Combined with the daily sample, the final dataset consists of 3129 unique videos.

3.2 Content Characterization

We summarize the topical characteristics of the sampled videos using three resources. First, we examine the list of topics YouTube API provides for each video. Second, we use Non-negative Matrix Factorization (NMF) to extract a set of topics from the video title, description, and transcript. Following Cheng et al. [6], we pre-process the text by removing URLs, numbers, and emojis, converting to lowercase, removing timestamps and text enclosed in square brackets, and other stop words. We use a combination of the topic coherence measure and manual selection of the topics to select the final number of topics, $k = 13$ (with coherence at 0.642). Finally, we assign meaningful titles to the topics, for further analysis. A third source of information is that about the channel: the YouTube API provides a field describing the topical specialty of the posting channel. We also manually classify the channel owner of a subset of videos (see the section below) following [22] into institution, commercial, individual, or other, as well as into broad topical categories: doctor, health, science, news, opinion, lifestyle, or other.

Further, we compute a range of surface-level statistical features from the title, description, and transcript text to quantitatively characterize video content. These include basic length-based metrics such as title length (in characters), description length, and average word length in the transcript, which capture verbosity and linguistic complexity. Additional features quantify stylistic elements, such as the ratio of uppercase letters, frequency of exclamation marks, and presence of emojis, reflecting emphasis or sensationalism in presentation. The analysis also considers structural indicators such as the number of hashtags, mentions, and external links in the description, which may signal engagement strategies or informational richness. Together, these surface statistics provide a standardized, non-semantic representation of textual characteristics. We explore the association between these, and the topical variables obtained above, with the quality and body scores, described below.

3.3 Annotation for Information Quality and Body-related Content

Guided by the previous literature, we utilize previously validated questionnaires in order to measure the quality of the information in each video and the extent to which the videos refer to the human body. Specifically, for the former, we adopt relevant measures from the Principles for Health-related Information on Social Media (PRHISM) questionnaire [10], the Health on the Net Foundation Code of Conduct (HONcode) [4], and the Social Media Evaluation Checklist (SMEC) [31]. The final list of dimensions is shown in Table 1. We score each principle on a scale from 0 to 4, or mark it as not applicable, if video lacks enough information for rating. The overall quality score of a video is an average of the non-empty scores, rescaled to a range of 0 to 100 (with a higher score meaning higher quality).

To measure the extent to which the content refers to the human body, we use the body-related variables developed by Munro et al. [24], listed in Table 2. The variables are scored and aggregated similarly to those for information quality, resulting in a score ranging from 0 to 100 (with a higher score meaning more focus on the body).

We created a validation set by manually annotating 50 videos, with 25 randomly selected from the daily sample and 25 from the popular one. While one author annotated all 50 sampled videos, four external annotators labeled a subset of 20 videos. On these 20 videos, a consensus between the two annotators was reached for every score, and the scores for all videos were adjusted to achieve a consistent interpretation of the principles. The annotation involved watching the full video, the duration of which ranged from 65 to 3194 sec, with an average of about 12 minutes. During annotation, we also included a question assessing the extent of AI use in the video production (from 0 when no AI tools were visibly used, to 4 when entire video was AI-generated), and two questions designed to characterize the channel which posted the video adopted from Mejova and Tizzani [22] (described in subsection above). Finally, we annotated for any specific brands that were mentioned in the video and validate the topics resulted from the NMF analysis.

Next, we develop an automatic annotation pipeline using OpenAI's GPT-4.1 model to estimate each of the criteria in the two scores, given the video title, description, and transcript. We explore zero- and few-shot alternatives, prompting for the chain-of-thought explanations, as well as two temperature settings. Validated on the manually annotated subset, the performance of the different models is statistically indistinguishable, thus we use the simplest version: zero-shot prompting without the chain-of-thought explanations with temperature 0.1. Compared to the manually annotated scores, it achieved a correlation of Spearman's $\rho = 0.669$ for the information quality and $\rho = 0.752$ for the body-related content. We truncate long transcripts by retaining the concatenation of the first and last 1,500 words (3,000 words in total). This model is then used to annotate the entire dataset. In the aims of reproducibility, we make the data, annotations and code used for this study available online[1] for the research community.

[1] https://github.com/MaddalenaGhiotti/YouTubeWeightLossProject

Table 1: Information Quality Principles. Each principle includes citation to the questionnaire from which it was adapted: PRHISM [10], HONcode [4], and SMEC [31].

Principle	Description
Authorship (PRHISM)	The speaker clearly presents her/his identity and relevant qualifications or experience (e.g., nutritionist, personal trainer, medical professional, or personal experience).
Authoritative (PRHISM)	Information is within the speaker's scope of expertise.
Action-oriented (PRHISM)	Information is presented clearly and concisely, providing enough context to help viewers make informed decisions, while acknowledging that multiple approaches may exist and explaining possible alternatives.
Attribution (PRHISM)	Information includes citations and hyperlinks (either verbally, in graphics, or in the video description) to the original source of information.
Balance and justifiability (PRHISM)	Information is balanced and unbiased, avoiding exaggeration of benefits, minimization of risks, or promotion of 'miracle' solutions.
Risks and benefits (PRHISM)	Information clearly outlines the potential benefits and any associated risks or side effects of the products, behaviors, or practices being discussed.
Complementary information (PRHISM)	Information is complementary and not designed to replace the relationship between individuals and health professionals. Information includes statements encouraging individuals to discuss choices with a relevant health professional.
Referrals and support (PRHISM)	The speaker directs viewers to additional reliable resources, official guidelines, or support services when relevant (e.g., eating disorder help lines).
Readability and comprehensibility (PRHISM)	Information is presented in plain, everyday language, avoiding jargon, technical terms, abbreviations, or uncommon words, or providing clear explanations when such terms are necessary.
Acknowledgment of uncertainty (HONcode)	Video openly acknowledges the limitations of the evidence it presents and the fact that knowledge may evolve over time.
Separation of interests (HONcode and SMEC)	Information is separate from financial, political, or ideological messages, including avoiding promotion of stereotypical attitudes and discriminatory actions about clinical or social issues.
Data (HONcode)	Creator shares data to support statements.

Table 2: Select body-related aspects adapted from Munro et al. [24].

Principle	Description
Weight measurement	Content includes references to body-weight measurement, whether as absolute values or as mentions of weight lost or gained. (ex. 'What I eat to lose 16 kg'.)
Mention of calories	Content includes mentions of the number of calories. (ex. 'This is one pint of ice cream, 1000 calories. And this is 8 cups of protein ice cream, 385 calories'.)
Referencing body image	Directly referencing body (ex. Man filming his shirtless body in the mirror at different timepoints. 'This is my body one day of dieting. One week of dieting. Two weeks of dieting'.)
Negative portrayal of body image	Content negatively refers to or negatively portrays body image. Body image is how we think and feel about ourselves physically (including our perceived sexual attractiveness) and how we believe others see us. (ex. A man starts working out to look more like the muscular men on TV his partner was admiring.)
Positive portrayal of body image	Content positively refers to or positively portrays body image. Body image is how we think and feel about ourselves physically (including our perceived sexual attractiveness) and how we believe others see us. (ex. 'This outfit looks so good on me, I don't care what you say, this dress looks so good on me')
Comparison	Content includes comparisons of the body across at least two time points, eventually using body-checking techniques to illustrate changes in weight or appearance, for example, Day 1 of a diet plan vs. Day 20. (ex. 'Diet journey. 83 kg59·6 kg'. Photos of a woman at different timepoints, with the weight at the time in text.)

4 Results

4.1 Categories of Diet-related Content

We begin by examining the sources of diet information we have captured in the samples. Concerning the categories of channel owner type and channel topical category identified by hand during annotation, we find that by far most accounts (58%) are individual ones, and only 8% are identifiably commercial, and only 1 account institutional (the rest were mostly removed channels which were difficult to characterize). Table 3 shows the distribution of most frequent channel categories (provided by the API, for the whole dataset), along with the number of channels and videos. As expected, *Health*, *Lifestyle*, and *Physical fitness* top the list. *Food* channels contribute proportionally most to the dataset (on average 1.93 videos per channel), however *Health* and *Lifestyle* also have relatively high posting rate (at around 1.74), indicating channels repeatedly returning to the topic.

When we compare the video categories provided by YouTube (Figure 1, left) to those we extracted from the video content (right), we find that the platform ones are dominated by generic labels, mainly *People & Blogs*, followed by *Entertainment* and *Education*. Instead, our NMF topics provide a fine-grained view of the data collection (note that a video may be assigned several NMF topics).

Table 3: Top channel topic categories (up to a large drop in videos), and their number of channels (N_c) and videos (N_v).

	N_c	N_v	v/c		N_c	N_v	v/c
Health	1569	2711	1.73	Knowledge	27	43	1.59
Lifestyle	1170	2044	1.75	Pet	15	25	1.67
Phys. Fitness	376	541	1.44	Film	12	12	1.00
Food	297	572	1.93	Hobby	10	18	1.80
Society	145	206	1.42	Fashion	9	13	1.44
Ent-ment	104	130	1.25	Politics	8	9	1.13
TV program	88	135	1.53				

Figure 2: Distribution of the quality and body scores, as well as the mean (μ) and variance (σ^2).

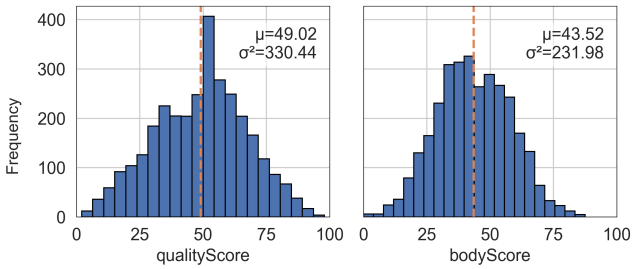
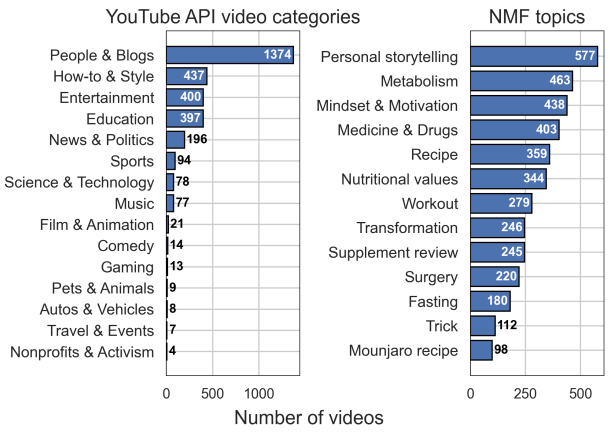


Figure 1: Number of videos across the categories provided by YouTube (left) and the topics extracted using NMF (right). Note, a video can have only one YouTube category, but can have multiple NMF topics.



We find that *personal storytelling* (disclosures of personal and emotional nature around the successes and challenges, mainly with weight loss) covers the most videos, followed by content focusing on the *metabolism* (how calories are burned, blood sugar levels, and digestion), and on *mindset & motivation* (stress and emotion management, body satisfaction, weight loss habit maintenance). The smallest topic is also the most peculiar: it concerns the recipes meant to replicate the effects of the prescription drug Mounjaro (see the case study below). In the subsequent analyses of information quality and user engagement, we consider the NMF topics.

4.2 Information Quality and Body-related Content

Next, we turn to the estimates of the information quality ("quality score") and the extent to which body-related content appears in the video ("body score"). Figure 2 presents the distribution of the scores in the dataset. We find that few videos achieve high quality scores, with an average of around 49 out of 100. Similarly, the body score averages at 44 out of 100. Note that the two scores correlated only slightly (Pearson $r = 0.054$, $p = 0.0024$), indicating that there are low-quality videos at both sides of the body score spectrum. Manual inspection revealed that the lowest-quality videos were those promoting "easy" diet "tricks" ("coffee loophole", "detox drinks", and weight-loss challenges). Those scoring low on body score focus strictly on the food, those with high body scores often mention the body explicitly ("How To Starve Yourself to 8% Body Fat").

Further, Figure 3 shows the distribution of the quality and body scores by the topic and the country of the posting channel. We find that the quality score is highest on average for channels in the categories of *Knowledge*, *Politics* and *Society*, and the lowest in the categories of *Music* (concerning which, see the case study in the section below). Although the data collection queries requested videos in the "US region", 31% of channels of the returned videos identified themselves as another country (UK at 9.5%, India 7.2%, Canada 5.4%, Australia 2.4%, and most of the rest can be seen in the figure). We find that the quality score is the highest for content from channels in New Zealand, the US and UK, Australia, Germany, and Canada. It is lowest for channels self-identifying as in Brazil, Dominican Republic, Mexico, and Peru (see Case Study subsection below on more regarding some Central and South American channels).

To understand what characteristics of the videos are most correlated with the quality and body scores, we perform univariate correlation analyses, shown in Table 4 (because of the complex relationships between the variables, we report the univariate analysis instead of the multivariate regressions, which require exclusion of variables due to multicollinearity). We log-transform highly skewed variables, and use Spearman's rank correlation coefficient (ρ_s), which is a non-parametric measure of the strength and direction of a monotonic relationship between two variables which may be skewed or non-normal.

Concerning the quality score (left half of the table), the largest positive effect size is observed for the video duration, followed by the number of mentions in the description. The lengths of content provide a contradictory picture: the length of the title and the average length of the words used in the text (title, description, and transcript) are negatively correlated with the quality score, while length of the channel description – positively. This may point to the practice of video title (but not description) "stuffing" with content for maximizing retrieval or recommendation [38]. Surprisingly, the number of subscribers a channel has is negatively related to the quality score, pointing to the popularity of those posting such content.

Figure 3: The distribution of the quality and body scores across (left) channel categories and (right) channel countries.

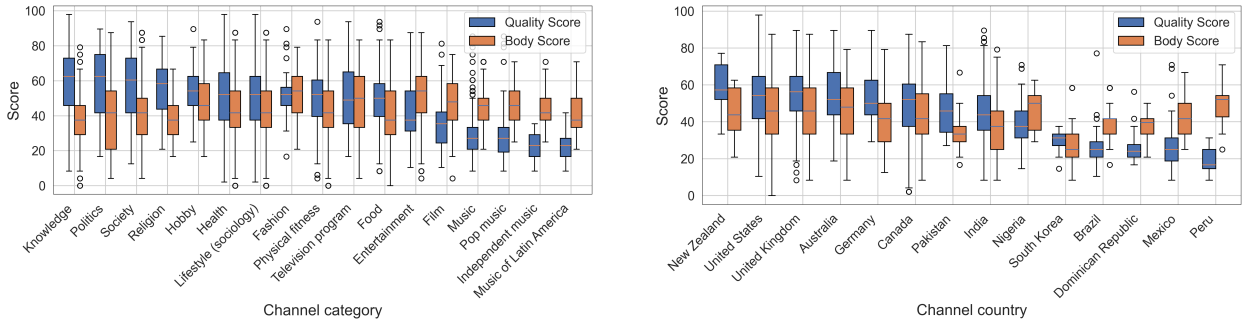


Table 4: Spearman's correlation coefficient (ρ_s) between the *quality score* (left) or *body score* (right) and descriptive variables of the video (or their logarithmic transformations) and the topic of the video. Additionally, corresponding p-values (p) and Bonferroni-adjusted p-values (p_{adj}) intervals are presented. Within each of the two sections of the table, variables are reported in descending order of the absolute values of the coefficients. Confid. levels: $p_{adj} < 0.001$ *, $p_{adj} < 0.01$ **, $p_{adj} < 0.05$ *.**

Quality Score				**Body Score**			
	ρ_s	p	p_{adj}		ρ_s	p	p_{adj}
duration (secs) (log)	0.413	3.034e-129	***	avg word length (char)	-0.316	2.210e-73	***
# mentions in description (log)	0.248	3.381e-45	***	duration (secs) (log)	0.311	2.707e-71	***
title length (char)	-0.241	1.711e-42	***	# mentions in description (log)	0.143	9.724e-16	***
avg word length (char)	-0.234	2.633e-40	***	title&desc uppercase ratio	0.119	2.060e-11	***
channel description length	0.224	8.116e-37	***	title&desc exclamations ratio	0.074	3.292e-05	***
# links in description (log)	0.152	1.485e-17	***	# links in description (log)	0.059	9.274e-04	*
channel video count (log)	-0.149	6.433e-17	***	# hashtags in description (log)	-0.026	0.142	
channel age (days)	0.143	9.163e-16	***	description length (log)	0.018	0.312	
channel subscriber count (log)	-0.138	8.192e-15	***	title length (char)	0.018	0.313	
title&desc uppercase ratio	-0.128	6.305e-13	***	channel video count (log)	-0.007	0.692	
title&desc emoji ratio	-0.123	4.856e-12	***	channel age (days)	-0.006	0.737	
# hashtags in description (log)	-0.045	0.011		channel description length	0.005	0.763	
title&desc exclamations ratio	-0.037	0.041		channel subscriber count (log)	0.004	0.826	
description length (log)	-0.015	0.414		title&desc emoji ratio	-9.320e-04	0.958	
Medicine & Drugs	0.444	1.804e-151	***	Personal storytelling	0.429	5.611e-140	***
Mindset & Motivation	0.265	1.357e-51	***	Transformation	0.212	3.372e-33	***
Personal storytelling	0.246	1.591e-44	***	Metabolism	-0.211	9.073e-33	***
Surgery	0.207	1.471e-31	***	Mindset & Motivation	0.200	1.485e-29	***
Trick	-0.181	1.797e-24	***	Workout	0.142	1.529e-15	***
Nutritional values	0.163	3.664e-20	***	Nutritional values	0.139	5.754e-15	***
Supplement review	-0.159	3.547e-19	***	Fasting	0.137	1.231e-14	***
Mounjaro recipe	-0.145	3.617e-16	***	Supplement review	-0.119	2.096e-11	***
Fasting	0.091	3.484e-07	***	Surgery	0.118	2.969e-11	***
Workout	0.087	1.031e-06	***	Recipe	-0.085	1.791e-06	***
Recipe	-0.061	5.804e-04	*	Medicine & Drugs	-0.046	0.010	
Transformation	-0.049	0.007		Trick	-0.035	0.049	
Metabolism	-0.006	0.720		Mounjaro recipe	0.013	0.451	

Considering the topic-related variables, several statistically significant associations are also observed. The strongest positive effect size is associated with *Medicine & Drugs*, followed by *Mindset & Motivation*. This result is consistent with expectations: the former topic is more closely related to medical content, which is more likely to be communicated by professionals or experts, while the latter typically involves motivational narratives rather than prescriptive claims or technical explanations, thereby reducing the likelihood of low quality. Conversely, strong negative correlations are observed for *Trick*, *Supplement review*, and *Mounjaro recipe*. These topics are more likely to involve non-expert creators presenting scientific claims or health-related advice, increasing the risk of

Table 5: Spearman's correlation coefficient (ρ_s) between individual principle scores and the view count (left) and engagement rate (right), together with the corresponding p-values (p) and Bonferroni-adjusted p-values (p_{adj}) intervals. Below these, same is listed for significant correlations of extracted video topics. Scores are reported in descending order of the absolute values of the coefficients, and only results with $p_{adj} < 0.05$ are shown. Confidence levels: $p_{adj} < 0.001$ *, $p_{adj} < 0.01$ **, $p_{adj} < 0.05$ *.**

Views				**Engagement Rate**			
	ρ_s	p	p_{adj}		ρ_s	p	p_{adj}
Separation of interests	-0.202	5.618e-26	***	Weight measurement	0.149	1.915e-14	***
Negative B.I.	0.151	4.670e-15	***	Complementary information	0.133	7.513e-12	***
Weight measurement	0.139	7.242e-13	***	Authorship	0.124	2.252e-10	***
Comparison	0.128	3.606e-11	***	Readability and comprehens.	0.117	2.217e-09	***
Referencing B.I.	0.137	1.535e-12	***	Positive B.I.	0.114	4.455e-09	***
Balance and justifiability	-0.115	3.033e-09	***	Referrals and support	0.087	9.268e-06	***
Risks and benefits	-0.100	2.154e-07	***	Action-oriented	0.082	2.813e-05	***
Authorship	0.068	4.765e-04	**	Data	-0.079	5.390e-05	***
Mention of calories	0.065	8.748e-04	*	Comparison	0.066	6.916e-04	*
				Mention of calories	0.060	0.002	*
Personal storytelling	0.208	2.491e-27	***	Personal storytelling	0.319	7.751e-63	***
Mindset & Motivation	-0.207	5.187e-27	***	Metabolism	-0.126	1.156e-10	***
Metabolism	-0.141	2.825e-13	***	Medicine & Drugs	-0.118	1.551e-09	***
Recipe	0.131	1.030e-11	***	Transformation	-0.116	3.052e-09	***
Mounjaro recipe	0.123	2.258e-10	***	Nutritional values	0.112	9.305e-09	***
Nutritional values	0.109	1.885e-08	***	Recipe	0.078	7.003e-05	**
Supplement review	0.072	1.927e-04	**	Workout	0.074	1.586e-04	**
				Fasting	0.061	0.002	*

lower-quality information. This is aligned with the aforementioned manual inspection.

Note that during the manual annotation of the 50-video subsample, we find a significant correlation between the level of AI use and the *quality score* (not shown in this table). The estimated coefficient is $\rho_s = -0.466$, with a p-value of approximately 0.001, indicating that higher levels of AI-assisted content generation are associated with lower video quality.

Turning to the body score (right half of the table), a smaller number of variables, particularly non-semantic ones, show statistically significant associations. We find that the videos using shorter words are more likely to have body-related aspects, as well as those with higher duration. The presence of uppercase and exclamation marks in the title and description are also associated with body-related mentions, indicating there are stylistic peculiarities to such content. The strongest positive correlation is observed for the topic of *Personal storytelling*. It is followed by *Transformation*, *Workout*, and, somewhat unexpectedly, *Mindset & Motivation*. Negative, although moderate, correlations are observed for topics such as *Metabolism*, *Recipe*, and *Supplement review*. This suggests that, although videos aligned with the *Supplement review* topic may exhibit lower informational quality in previous analyses, their primary focus is less explicitly centered on physical appearance and body-related content. Finally, in contrast to the findings for the quality score, the level of AI use in video production, manually annotated for the 50-video subsample does not exhibit a statistically significant correlation with the body score.

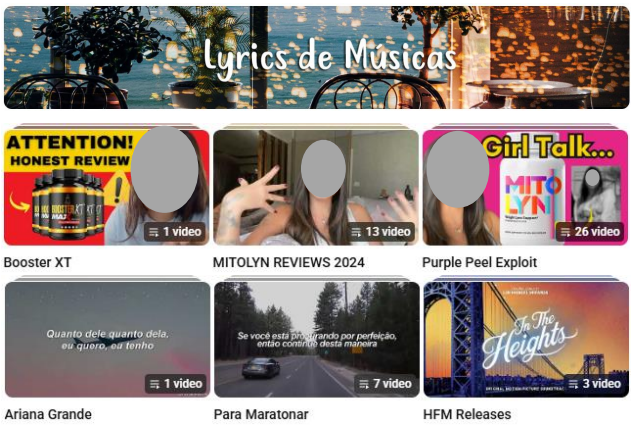
Although no AI-detection tools were used in this study, we can report the assessment of the annotators, who reported 11 out of 50 videos (22%) were categorized by the labelers as having AI tools were used somewhere in the video's production. Out of these, in 8 videos, AI tools were used to modify or generate a substantial portion of the video, significantly influencing its content, although the video was not entirely AI-generated.

4.3 User Engagement

Recall that we can measure engagement in two ways: the number of views the videos have received and the engagement rate, which is the sum of the number of likes and comments, normalized by the number of views. Due to the difference in collection timings, we perform this analysis on the daily collection only. We find that the correlation between the quality score and both metrics is low: $\rho_s = -0.053$ with the view count and 0.059 with the engagement rate. The relationship is slightly stronger for the body score, at 0.157 with the view count and 0.112 with the engagement rate. Table 5 shows the significant correlations between individual principles (the principles that comprise the quality and body scores) and the number of views the video receives (left) or the engagement rate (right). We also include the correlations with the presence of specific topics.

Nearly all of the principles comprising the body score, e.g. *weight measurement*, *comparison* and *negative body image*, are mainly positively correlated with the view count, indicating a viewer preference to the content focused on the human body. We find a similar pattern with engagement rate, however there it is the *positive body image* that is positively associated with the engagement rate. Unfortunately, we find several important signals of quality negatively related to views: the declaration of the *separation of interests*, *balance*

Figure 4: Example of a channel that first posted music videos, and later drug reviews.



and justifiability and the disclaimer of *risks and benefits*. However, being clear about the video's *authorship*, providing *complementary information* and *referrals*, and higher *readability and comprehensibility*, as well as being *action-oriented* are all principles associated with greater engagement rate.

Considering the video topics, those around *personal storytelling* and *recipe* (including *Mounjaro recipe*, see case study below) are positively associated with the views, whereas those mentioning *mindset & motivation* or the *metabolism* are associated with fewer views. The engagement rate is also highest for the *personal storytelling* videos, however the mentions of *metabolism*, *medicines & drugs* and *transformation* are associated with a lower engagement rate. These results point to the theme of *personal storytelling* as the most engaging to the platform's audience.

4.4 Case Study: Mitolyn and Mounjaro

During the topic discovery, we noted the appearance of two pharmaceuticals: Mitolyn, a dietary supplement (a "proprietary blend of 6 exotic nutrients and plants", according to the company's website [23]) and Mounjaro (generic drug name: tirzepatide), a gastric inhibitory polypeptide analog and a GLP-1 receptor agonist that is used for managing type 2 diabetes [11]. The videos mentioning these products occur unusually often in content posted by channels self-categorizing as "music". Among the 28 channels that uploaded videos associated with Mitolyn, the majority came from Central or South America, with Brazil (7 channels) and Mexico (6 channels) dominating the ranking. They are followed by the Dominican Republic, Peru, Chile, Ecuador, Venezuela, and Colombia.

Upon manual inspection, we found that several of these channels were originally created to post music content (and some still appear to do so at the time of writing), but were later repurposed to share videos promoting pharmaceuticals. Notably, the same three women appear in multiple videos across different channels, suggesting the presence of a single producer behind this content. An example of such a channel can be seen in Figure 4. At the time of writing, the medical "review" videos on this channel regularly garner over 1K views.

5 Discussion & Conclusions

This study offers two key contributions to advancing the understanding of the quality of health-related content shared on social media. First, we offer a sample of diet-related YouTube videos that combine the daily activity of the platform's posters with the most popular content over the same time period, capturing both posters' efforts in this domain and the most successful ones. Second, we adapt existing quality assessment rubrics from the literature to this specific domain, and, using these, develop and evaluate an LLM-based approach for applying them at scale to thousands of videos.

We show that our dataset displays a wide range of quality scores, with an average of about 49 out of 100, indicating much room for improvement. Interestingly, our results indicate that information quality and the degree of body emphasis represent independent (largely uncorrelated) dimensions within diet-related discourse, and each will require targeted attention from researchers. For instance, whereas we find that the mentions of weight measurement and negative body image are associated with the view count of the videos, the mentions of the positive body image are associated with an increased engagement in terms of likes and comments per view. The fact that the consumption of negative body-related messaging is more prominent than that of positive one, is especially worrying, since passive social media use has been shown to be related to body dissatisfaction and eating concerns (especially in young women) [35], while the consumption of body positivity content may not have protective effects [29].

Although dieting involves both medical and practical dimensions, the dominant theme in our dataset is personal storytelling—reflected not only in the volume of videos posted, but also in its strong positive link with views and the engagement rate. This illustrates how personal accounts (including highly popular "influencer" profiles) can undermine or complicate the official messages that health authorities share on the platform. Worryingly, we find that a channel's number of subscribers is negatively related to the quality score of its diet-related content. Efforts are already underway to offer influencers training and support so they can produce higher-quality health-related content [14]; however, additional work is required to evaluate the reach and impact of these initiatives.

An especially egregious example of platform manipulation can be seen in our case study of Mitolyn/Mounjaro promotion by accounts first established as music channels. Such subscriber "bait-and-switch" has been found on, for instance, Twitter, with the "repurposed" accounts often used to promote political misinformation [12]. A more extensive review is needed to determine whether some countries are more prone to hosting this type of content—knowledge that is important for both YouTube's quality assurance team and public health authorities. Note that these videos were identified using NMF topic discovery, a relatively lower-cost solution compared to LLM-based annotation. Further, the manual annotation effort in this work has revealed a worrying trend: the labelers detected AI use in 22% of the annotated videos, and the level of AI use was negatively related to the quality score, indicating its association with problematic content. Nonetheless, the role of AI in health-related communication will inevitably expand, and its impact needs to be carefully tracked by the research community.

This study has several limitations. Because of resource constraints, data collection was restricted to the U.S. Although some videos from other countries were included (due to the API's definition of "region"), the sample primarily reflects content seen by U.S. users and likely does not generalize to other countries. Targeted audits in other Central and South American countries are needed to assess the quality of their diet-related content. The LLM-based scorer also underperforms human evaluators, introducing potential error. Although we apply Bonferroni correction for multiple hypothesis testing, many observed correlations are small, suggesting no single strong determinant of diet-related content quality. In the future, including the visual information (alongside the voice transcripts) would improve the detection of body-related focus. We encourage further development of rubrics and metrics for evaluating and auditing health information on multimedia platforms.

References

[1] [n. d.]. YouTube API Reference. https://developers.google.com/youtube/v3/docs?hl=en. Accessed: 2025-11-25. (Cited on 2)

[2] [n. d.]. YouTube Transcript API. https://github.com/jdepoix/youtube-transcript-api. Accessed: 2025-11-25. (Cited on 3)

[3] C. H. Basch, I. C. H. Fung, A. Menafro, C. Mo, and J. Yin. 2017. An exploratory assessment of weight loss videos on YouTube™. *Public Health* 151 (Oct. 2017), 31–38. doi:10.1016/j.puhe.2017.06.016 (Cited on 1, 2)

[4] Celia Boyer, Mark Selby, J-R Scherrer, and Ron D Appel. 1998. The health on the net code of conduct for medical and health websites. *Computers in biology and medicine* 28, 5 (1998), 603–610. (Cited on 2, 3, 4)

[5] Deborah Charnock. 1998. *The DISCERN handbook: quality criteria for consumer health information on treatment choices*. Radcliffe Medical, Abingdon. OCLC: 59581385. (Cited on 2)

[6] Keyi Cheng, Stefan Inzer, Adrian Leung, Xiaoxian Shen, Michael Perlmutter, Michael Lindstrom, Joyce Chew, Todd Presner, and Deanna Needell. 2022. Multi-scale Hybridized Topic Modeling: A Pipeline for Analyzing Unstructured Text Datasets via Topic Modeling. https://arxiv.org/abs/2211.13496v1 (Cited on 3)

[7] Peggy Chi, Senpo Hu, Lei Shi, Tanya Kraljic, Justin Secor, Tao Dong, Irfan Essa, and Mike Cleron. 2025. WatchWithMe: LLM-Based Interactive Guided Watching of Review Videos. In *ACM Conference on Conversational User Interfaces (CUI '25)*. New York, NY, USA, 1–15. doi:10.1145/3719160.3736624 (Cited on 2)

[8] Caitlin Davey, Emily Newman, Joanna Hare, David Fluck, and Thang Sieu Han. 2024. Risk of Instagram dieting trends on eating behaviour and body satisfaction in women of different age and Body Mass Index. *Journal of Technology in Behavioral Science* (2024), 1–10. (Cited on 1)

[9] Emily Denniss, Rebecca Lindberg, Laura E. Marchese, and Sarah A. McNaughton. 2024. #Fail: the quality and accuracy of nutrition-related information by influential Australian Instagram accounts. *The Int J of Behavioral Nutrition and Physical Activity* 21, 1 (2024). doi:10.1186/s12966-024-01565-y (Cited on 2)

[10] Emily Denniss, Rebecca Lindberg, and Sarah A. McNaughton. 2022. Development of Principles for Health-Related Information on Social Media: Delphi Study. *Journal of Medical Internet Research* 24, 9 (Sept. 2022), e37337. doi:10.2196/37337 (Cited on 2, 3, 4)

[11] Drugs.com. [n. d.]. Mounjaro. https://www.drugs.com/mounjaro.html. Accessed: 2026-04-15. (Cited on 8)

[12] Tuğrulcan Elmas, Rebekah Overdorf, and Karl Aberer. 2023. Misleading repurposing on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 17. 209–220. (Cited on 8)

[13] Philippine Fassier, Anne-Sophie Chhim, Valentina A Andreeva, Serge Hercberg, Paule Latino-Martel, Camille Pouchieu, and Mathilde Touvier. 2016. Seeking health-and nutrition-related information on the Internet in a large population of French adults: results of the NutriNet-Santé study. *British Journal of Nutrition* 115, 11 (2016), 2039–2046. (Cited on 2)

[14] Wendy Glauser. 2026. Influencing the Influencers: How Health Experts Are Partnering With Content Creators to Fight Misinformation Online. *J Med Internet Res* 28 (27 Feb 2026). doi:10.2196/93450 (Cited on 8)

[15] Rocío Guardiola-Wanden-Berghe, Josefa D. Gil-Pérez, Javier Sanz-Valero, and Carmina Wanden-Berghe. 2011. Evaluating the quality of websites relating to diet and eating disorders. *Health Information & Libraries Journal* 28, 4 (2011), 294–301. doi:10.1111/j.1471-1842.2011.00961.x (Cited on 2)

[16] Flavio Jeronimo and Eliana Veiga Carraca. 2022. Effects of fitspiration content on body image: a systematic review. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity* 27, 8 (2022), 3017–3035. (Cited on 2)

[17] Katikapalli S. Kalyan. 2023. A Survey of GPT-3 Family Large Language Models Including ChatGPT and GPT-4. doi:10.2139/ssrn.4593895 (Cited on 2)

[18] Mahmoud Khalil, Fatma Mohamed, and Abdulhadi Shoufan. 2025. Evaluating the quality of medical content on YouTube using large language models. *Scientific Reports* 15, 1 (March 2025), 9906. doi:10.1038/s41598-025-94208-6 (Cited on 2)

[19] Megan Kreft, Brittany Smith, Daniella Hopwood, and Renee Blaauw. 2023. The use of social media as a source of nutrition information. *South African Journal of Clinical Nutrition* 36, 4 (Dec. 2023), 162–168. doi:10.1080/16070658.2023.2175518 Publisher: NISC (Pty) Ltd. (Cited on 2)

[20] Leonardo La Rocca. 2025. Multimodal LLMs vs LLMs vs RoBERTa: evaluating AI performance in detecting conspiracies on YouTube. *Politecnico di Milano, Doctoral Thesis* (April 2025). (Cited on 2)

[21] Joanne Mayoh and Ian Jones. 2021. Young people's experiences of engaging with fitspiration on Instagram: Gendered perspective. *Journal of Medical Internet Research* 23, 10 (2021). doi:10.2196/17811 Place: Canada Publisher: JMIR Publications. (Cited on 1, 2)

[22] Yelena Mejova and Michele Tizzani. 2025. Vaccine Hesitancy on YouTube: A Competition Between Health and Politics. In *2025 10th International Digital Public Health Conference (DPH)*. IEEE, 1–8. (Cited on 2, 3)

[23] Mitolyn. [n. d.]. Official Webpage. https://mitolyn.com/. Accessed: 2026-04-15. (Cited on 8)

[24] Emily Munro, Gabriella Wells, Rigel Paciente, Nicole Wickens, Daniel Ta, Joelie Mandzufas, Karen Lombardi, and Alix Woolard. 2024. Diet culture on TikTok: a descriptive content analysis. *Public Health Nutrition* 27, 1 (Jan. 2024), e169. doi:10.1017/S1368980024001381 (Cited on 3, 4)

[25] Harsha Nori, Nicholas King, Scott Mayer McKinney, Dean Carignan, and Eric Horvitz. 2023. Capabilities of GPT-4 on Medical Challenge Problems. doi:10.48550/arXiv.2303.13375 arXiv:2303.13375 [cs]. (Cited on 2)

[26] Atte Oksanen, David Garcia, Anu Sirola, Matti Näsi, Markus Kaakinen, Teo Keipi, and Pekka Räsänen. 2015. Pro-Anorexia and Anti-Pro-Anorexia Videos on YouTube: Sentiment Analysis of User Responses. *Journal of Medical Internet Research* 17, 11 (Nov. 2015). doi:10.2196/jmir.5007 (Cited on 1, 2)

[27] Rebecca L. Pearl and Rebecca M. Puhl. 2016. The distinct effects of internalizing weight bias: An experimental study. *Body Image* 17 (June 2016), 38–42. doi:10.1016/j.bodyim.2016.02.002 (Cited on 1)

[28] Les Rose, Susan Bewley, Mandy Payne, David Colquhoun, and Simon Perry. 2024. Using artificial intelligence (AI) to assess the prevalence of false or misleading health-related claims. *Royal Society Open Science* 11, 10 (2024). (Cited on 2)

[29] Christina M Sanzari, Sasha Gorrell, Lisa M Anderson, Erin E Reilly, Martha A Niemiec, Natalia C Orloff, Drew A Anderson, and Julia M Hormes. 2023. The impact of social media use on body image and disordered eating behaviors: Content matters more than duration of exposure. *Eating behaviors* 49 (2023), 101722. (Cited on 1, 8)

[30] William M Silberg, George D Lundberg, and Robert A Musacchio. 1997. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor—Let the reader and viewer beware. *Jama* 277, 15 (1997), 1244–1245. (Cited on 2)

[31] Kelly Squires, Alisha Brighton, Lisa Urquhart, Lucy Kocanda, and Susan Heaney. 2023. Informing online professional dietetics practice: The development and pilot testing of the Social Media Evaluation Checklist. *Nutrition & Dietetics* 80, 4 (2023), 351–361. doi:10.1111/1747-0080.12794 (Cited on 2, 3, 4)

[32] Hao Tang, Sungwoo Kim, Priscila E Laforet, Naa-Solo Tettey, and Corey H Basch. 2022. Loss of weight gained during the COVID-19 pandemic: Content analysis of YouTube videos. *JMIR Formative Research* 6, 2 (2022), e35164. (Cited on 1)

[33] Marika Tiggemann and Mia Zaccardo. 2015. "Exercise to be fit, not skinny": The effect of fitspiration imagery on women's body image. *Body Image* 15 (Sept. 2015), 61–67. doi:10.1016/j.bodyim.2015.06.003 (Cited on 1)

[34] Yuqing Wang, Yun Zhao, and Linda Petzold. 2023. Are Large Language Models Ready for Healthcare? A Comparative Study on Clinical Language Understanding. In *Proceedings of the 8th Machine Learning for Healthcare Conference*. PMLR, 804–823. (Cited on 2)

[35] Kangqiao Xiang and Fanchang Kong. 2024. Passive social networking sites use and disordered eating behaviors in adolescents: The roles of upward social comparison and body dissatisfaction and its sex differences. *Appetite* 198 (2024), 107360. doi:10.1016/j.appet.2024.107360 (Cited on 8)

[36] Liuyi Yang, Sinan Chen, Yangmei Xie, Zhiyi Zhu, Miao Zhang, Yue Zhang, and Jialong Li. 2025. LLM-Based Teacher Tone Recognition for Educational Scenarios. In *2025 7th International Conference on Software Engineering and Computer Science (CSECS)*. IEEE, 1–6. (Cited on 2)

[37] Michelle Zeng, Jacqueline Grgurevic, Rayan Diyab, and Rajshri Roy. 2025. #WhatIEatinaDay: The Quality, Accuracy, and Engagement of Nutrition Content on TikTok. *Nutrients* 17, 5 (2025), 781. doi:10.3390/nu17050781 (Cited on 1, 2)

[38] Herbert Zuze and Melius Weideman. 2013. Keyword stuffing and the big three search engines. *Online Information Review* 37, 2 (2013), 268–286. (Cited on 5)