Platform-Labeled Synthetic Content in the 2024 U.S. Presidential Election: Prevalence, Diffusion, and Visual Signatures on YouTube

Yongjun Zhang[1]

Stony Brook University, Stony Brook, NY 11794, USA
yongjun.zhang@stonybrook.edu

Abstract. Generative AI raised acute fears about synthetic political media in the 2024 U.S. presidential election, but prior measurement has relied on fact-check audits, deepfake-detector heuristics, or small experimental stimuli. We use YouTube's own "Altered or synthetic content" disclosure as a platform-grounded measurement anchor and study its prevalence, diffusion, and visual signatures at scale. From a scraped universe of 1,870,097 election-relevant YouTube videos, 50,099 (2.679%; Wilson 95% CI [2.656, 2.702]) carry the platform's synthetic-content label. For 20,803 of these we recover full metadata and longitudinal engagement; for 8,736 we sample 503,466 frames and produce visual annotations using Qwen3-VL-30B together with the `athec` image-feature library. Three findings stand out. First, the labeled corpus comes from a long tail of 8,375 producer channels (HHI 21.7 [20.9, 23.7]), yet audience attention is extremely concentrated: the view Gini is 0.947 [0.937, 0.957] and the top 1% of videos absorb 60.8% of views. Second, the upload peak falls in the post-election transition (January 5, 2025: 198 single-day uploads) rather than around Election Day. Third, within the labeled set, creator self-disclosure of AI use aligns strongly with foundation-model visual judgment (Cohen's $d = 0.74$), with self-disclosed videos exhibiting higher color saturation, warmer Lab chroma, and a stronger vertical-format bias. Visual AI-likeness is essentially uncorrelated with engagement (Pearson $r = -0.052$ with log views; 95% CI $[-0.073, -0.031]$, within a $|r| = 0.10$ equivalence bound). Platform disclosures, creator-issued labels, and VLM annotations together produce a measurement stack for synthetic political media that no single signal supplies alone.

Keywords: Generative AI · Synthetic media · YouTube · 2024 U.S. election · Vision-language models · Large-scale social media analytics.

1 Introduction

The 2024 U.S. presidential election was the first in which generative AI produced video content at large consumer scale. Synthetic candidate footage, AI-narrated newsreels, and stylized "deepfake" memes circulated widely, motivating familiar concerns about misinformation, disclosure, and the integrity of political communication [13,36,7]. Yet the empirical picture has remained thin. Published

measurements rely on curated samples of viral incidents [21], fact-check-anchored databases [10], or experimental stimuli of unknown ecological validity [34,3,16]. Two questions remain hard to answer at scale: *how much* synthetic political video is actually present on a major platform during an election, and *what does it look like*?

A productive way to make headway is to take seriously a signal the platform itself produces. Since March 2024, YouTube has required creators to disclose meaningfully altered or synthetically generated content, and the platform surfaces the resulting badge on the watch page [37]. The label is imperfect — it captures creator-acknowledged AI use rather than ground-truth provenance, and creators who fail to disclose escape it — but it has two properties that detector outputs and fact-check tallies do not. It is uniformly applied by the platform, and it is anchored in the platform's own policy enforcement [8,11].

This paper uses YouTube's synthetic-content disclosure as the anchor for a two-part empirical study. **Part 1** characterizes the population of platform-labeled synthetic videos: prevalence, daily upload distribution, producer concentration, audience reach. **Part 2** focuses on a downloaded image-analyzed subset and asks whether the visual content of labeled videos differs systematically from videos in which creators *self-disclose* AI use. We triangulate three labeling signals — platform-issued disclosure, creator-issued self-disclosure, and Qwen3-VL-30B foundation-model visual judgment — across a 503,466-frame sample.

We make three contributions. *First*, the first large-scale prevalence estimate of platform-labeled synthetic election videos on YouTube using a public disclosure signal rather than detector heuristics or fact-check tallies. *Second*, we document a striking diffusion structure: synthetic content peaks *after* Election Day, is uploaded by a long tail of small channels, and is consumed primarily through a tiny minority of viral cases. *Third*, we show that VLM-derived visual *AI-likeness* aligns strongly with creator self-disclosure within the labeled set, and that color saturation, Lab chroma, and vertical-format framing are the most informative low-level visual signatures. Throughout, we treat the VLM as an auditable stylometer rather than a deepfake provenance detector [38,12,28].

2 Background and Related Work

Synthetic media measurement and election effects. Pre-election commentary framed AI as a likely force multiplier for deception [13,7]; post-election retrospectives are more tempered [32,21]. Verdoliva [35] provides the canonical media-forensics overview; FaceForensics++ [30] and Korshunov and Marcel [22] established the benchmark-dataset tradition; AMMeBa [10] catalogs misinformation modalities. Experimental work shows that deepfakes produce *uncertainty* more than belief [34], that microtargeted deepfakes can shift attitudes [9], that deepfakes are credible but not more persuasive than other fake media [3], and that audiovisual deepfakes are often judged less credible than cheapfakes [16,15]. These instruments are fact-check- or stimulus-anchored; ours is anchored in platform-issued labels on the platform where consumption happens.

YouTube as a political-content platform. Ribeiro et al., the Hosseinmardi-Watts collaboration, Munger and Phillips, Chen et al., and Brown/Lai et al. establish YouTube as a central site for computational political-communication research and document extreme audience-reach concentration [29,18,19,25,6,23]. None of this work measures synthetic-media prevalence per se.

Foundation models and image features. VLMs increasingly function as zero-shot annotators in computational social science [12,33,38]; LLaVA, Qwen-VL, and Qwen3-VL scale the vision-language family to social-science-relevant capacity [24,1,28]. On the visual side, Joo et al., Casas & Webb Williams, and Peng established computer vision for political visual analysis [20,4,26], with the `athec` library packaging 142 image features for social research [27]. Foundation-model visual annotation has not previously been deployed against platform-labeled political video.

Attention concentration. The structural heavy-tail of attention on platforms is well documented [5,31,2]. We test whether platform-labeled synthetic content inherits the same skew.

3 Data and Pipeline

Corpus. We scraped 1,870,097 election-relevant YouTube videos from the YouTube Data API [14] using political and election keywords; the scrape spans July 28, 2024 through February 4, 2025. For each video we read the public watch-page disclosure region and recorded a binary field (1 if YouTube surfaced the *Altered or synthetic content* badge; 2 otherwise). Of the 1,870,097 scraped videos, 50,099 carry the badge. For 20,803 of these we recover complete API metadata; the metadata-rich subset spans 8,375 unique channels with topic-category labels indicating that 69.4% are tagged *politics* and 65.6% *society*.

Engagement panel. For each labeled video we collect repeated view, like, and comment counts via the YouTube Data API. The panel covers 1,766,696 video-day rows over the first 100 days after publication; the latest observation is available for 20,802 of 20,803 metadata-rich videos.

Frame sampling and visual annotation. We frame-sampled 8,736 labeled videos at 15-second intervals using `ffmpeg`, yielding 503,466 JPEG frames (median 8 sampled frames per video, mean 57.6). We annotated every frame with Qwen3-VL-30B-A3B-Instruct [28] in bfloat16 across four H200 GPUs, prompting for a one-sentence summary, a binary `ai_generated` flag, a confidence score, and a short rationale (greedy decoding, `max_new_tokens=96`, chunk size 64). We aggregate frame labels per video as the share of frames flagged AI-like:

$$A_i = \frac{1}{n_i} \sum_{j=1}^{n_i} a_{ij}, \quad a_{ij} \in \{0, 1\}.$$

We call A_i *video AI-likeness* and treat it as a foundation-model judgment of *visual style*, not a provenance claim. In parallel we extract 142 deterministic per-frame features using the `athec` library [27], covering five color spaces, colorfulness [17], blur and sharpness, saliency, and image geometry.

Table 1. Main analytic artifacts.

Artifact	Scale	Role
Scraped universe	1,870,097 videos	Denominator for prevalence
Platform-labeled synthetic	50,099 videos	Outcome population
Metadata-rich labeled	20,803 videos	Patterns and engagement
Engagement panel	1,766,696 obs.	Day-of-life dynamics
Frame-sampled subset	8,736 videos	VLM + CV (503,466 frames)
Self-disclosed AI subset	7,774 videos	Internal validity check

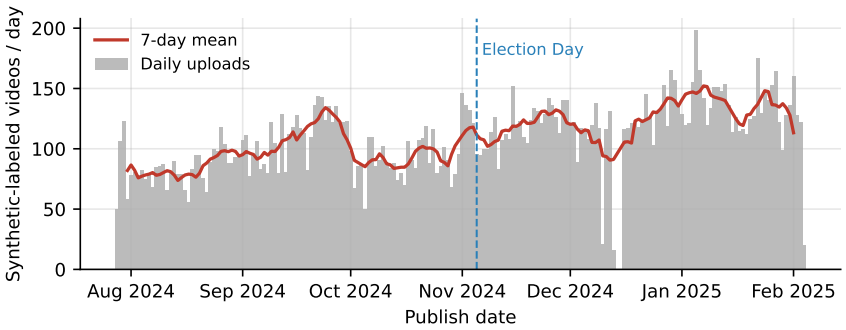

Fig. 1. Daily uploads of platform-labeled synthetic election videos. Red curve: 7-day centered moving average. Dashed line: Election Day (November 5, 2024). $n = 20{,}803$ with metadata.

Self-disclosure label. We identified creator self-disclosure by string-matching titles, descriptions, and tags against keywords ("AI", "AI-generated", "AI video", "deepfake", "synthetic"). The deduplicated set covers 7,774 videos overall and 1,132 within the image-analyzed subset.

Estimators. Prevalence is reported with Wilson 95% CIs. Channel HHI and view Gini use 95% bootstrap CIs ($B = 500$, seed = 20260514). Group comparisons use Welch t-tests and Cohen's d. Correlations use Pearson r with Fisher-z 95% CIs. We apply Holm-Bonferroni correction across the three primary significance tests; descriptive concentration metrics (HHI, Gini, rank-order timing) are reported separately. The near-zero engagement correlation is evaluated against a $|r| = 0.10$ equivalence bound.

4 Part 1: Patterns of Platform-Labeled Synthetic Videos

4.1 Prevalence and temporal dynamics

The platform's synthetic-content badge is rare but not negligible: **2.679%** of election-relevant videos in our scraped universe carry the label (Wilson 95% CI $[2.656, 2.702]$). Figure 1 shows daily upload counts. All ten of the top ten

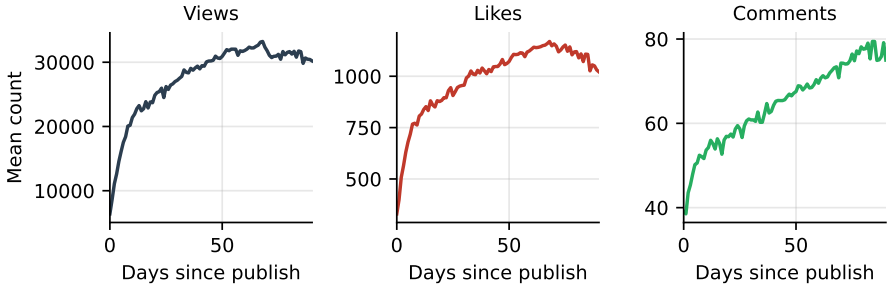

Fig. 2. Mean day-of-life engagement trajectory ($n = 20{,}803$ videos, 1,766,696 video-day observations across days 0–90).

single-day upload dates fall in the post-election window. January 5, 2025 is the absolute maximum (198 uploads), followed by January 22 (175), December 29 (165), and January 6 (165). Upload volume rises gradually through the campaign (337 videos in July 2024; 2,792 in October) and surges through January 2025 (4,254 in that single month). The transition window produces more synthetic-labeled content than the campaign closing period, suggesting that speculation and commentary frames are an unusually good fit for AI-video formats.

4.2 Channel concentration

The 20,803 labeled videos are distributed across **8,375 unique channels** with an upload-share Herfindahl-Hirschman index of HHI = 21.7 (95% bootstrap CI $[20.89, 23.70]$) — about two orders of magnitude below standard concentration thresholds. The top 10 channels account for 10.52% of labeled uploads, and the top 1% of channels (84 channels) for 28.5%. Content is produced overwhelmingly by smaller actors rather than a few high-volume factories: there is no short list of channels whose removal would meaningfully reduce total synthetic-content volume.

4.3 Engagement trajectories and attention concentration

Figure 2 shows the mean day-of-life trajectory. Views climb from 6,397 on the publish day to 18,402 on day 7, 27,187 on day 30, and 31,823 on day 60, plateauing thereafter. Roughly 60% of cumulative views accrue within the first 30 days, consistent with prior YouTube engagement-decay findings [29].

Attention is dramatically concentrated. The median platform-labeled synthetic video has 505 views, 9 likes, and 1 comment at the latest observation, while the mean view count is 24,703 and the maximum is 43,912,181. The view Gini coefficient is **0.947** (bootstrap 95% CI $[0.937, 0.957]$), with comparable Ginis for likes (0.957) and comments (0.942). Figure 3 shows the Lorenz curve: the top 1% of videos absorb **60.8%** of all views, the top 10% absorb 93.1%. Only 37.6%

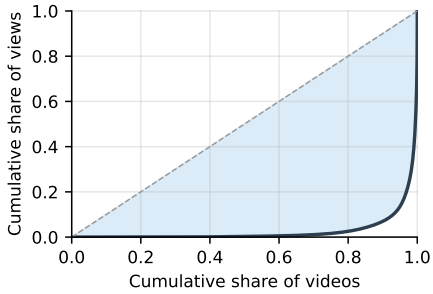

Fig. 3. Lorenz curve for views accruing to platform-labeled synthetic videos at the latest observation.

of labeled videos exceed 1,000 views; 15.5% exceed 10,000; 3.6% exceed 100,000; and just 0.4% reach one million views. Synthetic-content *prevalence* is a long-tail producer phenomenon, but synthetic-content *exposure* is dominated by a tiny minority of viral videos.

5 Part 2: Visual Signatures

For the 8,736 frame-sampled videos we ask two questions. First, does the VLM's visual judgment align with creator self-disclosure of AI use? Second, which low-level image features distinguish self-disclosed AI videos from other platform-labeled videos? We frame the comparison explicitly as a *convergent-validity* exercise, not a behavioral effect. We note up front that Qwen3-VL was trained on internet-scale data that may have included YouTube videos with disclosure language; we return to this in Section 7.

5.1 Convergent validity of VLM AI-likeness with self-disclosure

The video-level AI-likeness distribution is bimodal: 43.1% of labeled videos have $A_i < 0.2$ (few AI-like frames) and 32.0% have $A_i \geq 0.8$ (almost all frames AI-like). Self-disclosure aligns closely with the visual signal (Figure 4; Table 2). Among 1,132 self-disclosed videos the mean A_i is 0.706 (median 1.0); among the 7,604 platform-labeled but not self-disclosed videos it is 0.403 (median 0.25). The Cohen's d between the two groups is $+\mathbf{0.740}$ (Welch $t = 24.6$, uncorrected $p = 8.95 \times 10^{-113}$, Holm-corrected $p < 10^{-3}$). A robustness check on the politics-only subset ($n = 6{,}101$) yields a comparable effect ($d = +0.806$), confirming that the alignment is not driven by topic-mix differences.

A substantial share of *not*-self-disclosed platform-labeled videos — 28.0% — still register high VLM AI-likeness ($A_i \geq 0.8$), suggesting that visual analysis surfaces synthetic-style content that creators have not voluntarily flagged but the platform has.

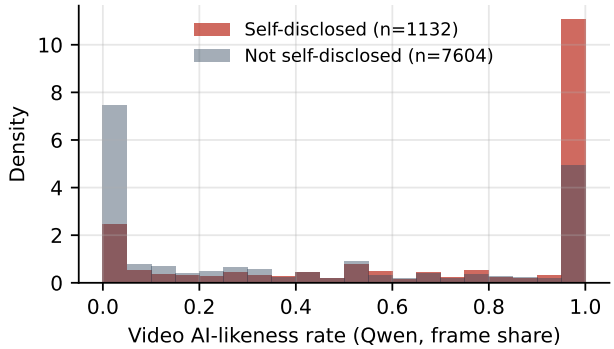

Fig. 4. Distribution of video-level AI-likeness rate A_i within the platform-labeled subset, by creator self-disclosure status.

Table 2. AI-likeness by self-disclosure status, image-analyzed subset ($n = 8{,}736$).

Group	Videos	Mean A_i	$\Pr(A_i \geq 0.5)$	$\Pr(A_i \geq 0.8)$
Self-disclosed AI	1,132	0.706	71.8%	58.8%
Not self-disclosed	7,604	0.403	40.0%	28.0%

5.2 Image-feature signatures

Table 3 compares low-level visual features between self-disclosed and non-self-disclosed platform-labeled videos. This is a descriptive ranking selected from 142 `athec` features; effect sizes should be read as exploratory rather than as a multiple-comparison-corrected hypothesis test. Self-disclosed AI videos are *more saturated*, *taller* (more vertical), and *less spatially salient* than other platform-labeled videos. Lab a and b shifts (less green, more yellow) and HSL/HSV saturation differences point to a perceptual style that is more vivid and warmer than typical political-discourse footage. Blur and sharpness measures are essentially indistinguishable: AI-generated political content in this corpus is not characterized by visible-artifact softness.

The same pattern emerges in continuous form. Pearson r between A_i and HSV saturation is $+0.129$ ($p < 10^{-33}$), with frame height $+0.148$, colorfulness $+0.103$, Lab a -0.062, and Lab b $+0.062$ (all $p < 10^{-8}$). The stylistic signature is small in magnitude (most d values are 0.05–0.15) and clearly limited to bulk visual style; we do not claim it can detect any individual AI-generated video.

5.3 Engagement and visual AI-likeness

Within the labeled subset, Pearson r between A_i and log latest view count is $\mathbf{-0.052}$ (95% CI $[-0.073, -0.031]$, Holm-corrected $p = 3.0 \times 10^{-6}$). The 95% CI

Table 3. Visual features by self-disclosure status, ranked by absolute Cohen's d ($n =$ 8,736). * $p < 0.01$; ** $p < 0.001$ (Welch). Descriptive ranking selected from 142 features.

Feature	Self-disclosed	Not disclosed	d	Sig.
Lab a (green–red)	133.21	134.27	−0.153	**
HSV saturation	95.52	90.32	+0.150	**
Lab b (blue–yellow)	131.14	129.93	+0.135	**
HSL saturation	91.78	87.03	+0.131	**
Frame height (px)	714.29	676.62	+0.089	*
Saliency total	0.380	0.392	−0.083	*
Brightness (HSV V)	118.56	121.74	−0.080	*
Luma (Rec. 709)	96.19	98.98	−0.076	*
Blur (Laplacian)	24.12	24.98	−0.066	—
Colorfulness [17]	53.68	52.35	+0.055	—

excludes $|r| > 0.10$, satisfying a small-effect equivalence bound. The log comment-count correlation is more negative ($r = -0.088$, 95% CI $[-0.109, -0.067]$) and falls outside the equivalence bound by a small margin; we note this asymmetry rather than pool. Self-disclosed videos do accumulate more attention than other labeled videos (mean views 59,539 vs. 22,680; median 668 vs. 494), but the gap is not driven by AI-likeness per se: at the same disclosure level, more visibly AI-like videos do not outperform less visibly AI-like ones.

6 Discussion

Three findings have direct implications for large-scale social media intelligence about generative AI. **Platform disclosures are a tractable measurement anchor.** The 2.679% prevalence estimate rests on a public, uniformly applied platform signal — not a detector output and not dependent on internal classifier choices. Our denominator sits at a different epistemic level from AMMeBa's fact-check corpus [10] and Kapoor and Narayanan's incident audit [21]; the three measurement designs are complementary.

Synthetic content is long-tail in production and heavy-tail in exposure. The 8,375 channels producing labeled content (HHI 21.7) are too dispersed for source-side moderation, but the view Gini of 0.947 means a tiny fraction of viral videos drives most exposure. A demand-side intervention strategy — rapid triage of high-momentum cases — can reach a disproportionately large share of viewing even as upload-side enforcement remains intractable.

Visual style is informative; engagement is not. The strongest internal-validity check we have aligns strongly with foundation-model judgment ($d = 0.74$), and a handful of low-level features (HSV/HSL saturation, Lab chroma, vertical-aspect bias) separate self-disclosed from non-disclosed labeled videos. Engagement counts, in contrast, are nearly uninformative. This is consistent with Vaccari and Chadwick's finding that synthetic political video produces *uncertainty* rather than belief [34] — uncertainty does not amplify into viral attention.

7 Limitations

Several limitations qualify these claims. The synthetic-content disclosure is creator-issued; non-disclosers escape the label, so 2.679% is a lower bound. The Qwen3-VL AI-likeness rate is a visual-style judgment, and high values can flag stylized but non-synthetic content (animation, filters). We do *not* claim it is a deepfake-detection benchmark; a human-coded validation block (≥ 200 frames, two coders, Cohen's κ) is the most important next step. Qwen3-VL was trained on internet-scale image-text pairs that may have included YouTube content with disclosure language, so the convergent-validity finding cannot rule out training-distribution overlap. Frames are exported every 15 seconds; brief synthetic segments may be missed. Engagement counts come from the public API and do not capture watch time or recommendation impressions. Non-English content is underrepresented in our scraping coverage; cross-platform generalization to TikTok or short-form X video is not warranted.

8 Conclusion

We provide the first large-scale measurement of platform-labeled synthetic political video on YouTube during the 2024 U.S. presidential election cycle, anchored to YouTube's own *Altered or synthetic content* disclosure. We document a 2.679% prevalence rate, a long-tail producer distribution paired with extreme audience concentration (view Gini 0.947), a post-election rather than pre-election volume peak, and a coherent visual signature of self-disclosed AI content (Cohen's $d = 0.74$ for AI-likeness; small but systematic shifts in HSV/Lab features and vertical framing). The combination of platform disclosures, creator-issued labels, and large-scale VLM annotation provides a principled measurement stack for the next election cycle and a starting point for monitoring generative political media on other platforms where comparable disclosure regimes are in place.

Acknowledgments. We thank Stony Brook University's HPC center for H200 GPU access used in the VLM pipeline.

Disclosure of Interests. The author declares no competing interests. This paper used Qwen3-VL-30B-A3B-Instruct as a visual-feature annotator on locally stored frames (H200 cluster, no off-campus data transfer) and Claude Code (Anthropic) for data analysis and draft preparation; only summary statistics and code snippets were sent to the Anthropic API. All analytic decisions, interpretations, and conclusions are the author's responsibility.

References

1. Bai, J., Bai, S., Yang, S., et al.: Qwen-VL: A versatile vision-language model for understanding, localization, text reading, and beyond (2023)

2. Bakshy, E., Messing, S., Adamic, L.A.: Exposure to ideologically diverse news and opinion on Facebook. Science **348**(6239), 1130–1132 (2015)
3. Barari, S., Lucas, C., Munger, K.: Political Deepfakes Are as Credible as Other Fake Media and (Sometimes) Real Media. The Journal of Politics **87**(2) (2025)
4. Casas, A., Webb Williams, N.: Images that Matter: Online Protests and the Mobilizing Role of Pictures. Political Research Quarterly **72**(2), 360–375 (2019)
5. Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.Y., Moon, S.: I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User-Generated Content Video System. In: Proc. ACM IMC '07. pp. 1–14 (2007)
6. Chen, A.Y., Nyhan, B., Reifler, J., Robertson, R.E., Wilson, C.: Subscriptions and external links help drive resentful users to alternative and extremist YouTube channels. Science Advances **9**(35) (2023)
7. Chesney, R., Citron, D.K.: Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. California Law Review **107**, 1753–1820 (2019)
8. Coalition for Content Provenance and Authenticity: C2PA Technical Specification v2.1 (2024)
9. Dobber, T., Metoui, N., Trilling, D., Helberger, N., de Vreese, C.: Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes? International Journal of Press/Politics **26**(1), 69–91 (2021)
10. Dufour, N., Pathak, A., Samangouei, P., Hariri, N., Deshetti, S., Dudfield, A., Guess, C., Hernández Escayola, P., Tran, B., Babakar, M., Bregler, C.: AMMeBa: A Large-Scale Survey and Dataset of Media-Based Misinformation In-The-Wild (2024)
11. European Parliament and Council: Regulation (EU) 2024/1689 (Artificial Intelligence Act), Article 50: Transparency obligations (2024)
12. Gilardi, F., Alizadeh, M., Kubli, M.: ChatGPT outperforms crowd workers for text-annotation tasks. Proceedings of the National Academy of Sciences **120**(30) (2023)
13. Goldstein, J.A., Lohn, A.: Deepfakes, Elections, and Shrinking the Liar's Dividend. Brennan Center for Justice (2024)
14. Google: YouTube Data API Documentation (2024)
15. Groh, M., Sankaranarayanan, A., Singh, N., Kim, D.Y., Lippman, A., Picard, R.: Human detection of political speech deepfakes across transcripts, audio, and video. Nature Communications **15**(1) (2024)
16. Hameleers, M., van der Meer, T.G.L.A., Dobber, T.: They Would Never Say Anything Like This! Reasons To Doubt Political Deepfakes. European Journal of Communication **39**(1), 56–70 (2024)
17. Hasler, D., Suesstrunk, S.: Measuring colourfulness in natural images. In: Proc. SPIE Human Vision and Electronic Imaging VIII. vol. 5007, pp. 87–95 (2003)
18. Hosseinmardi, H., Ghasemian, A., Clauset, A., Mobius, M., Rothschild, D.M., Watts, D.J.: Examining the consumption of radical content on YouTube. Proceedings of the National Academy of Sciences **118**(32) (2021)
19. Hosseinmardi, H., Ghasemian, A., Rivera-Lanas, M., Horta Ribeiro, M., West, R., Watts, D.J.: Causally estimating the effect of YouTube's recommender system using counterfactual bots. Proceedings of the National Academy of Sciences **121**(8) (2024)
20. Joo, J., Li, W., Steen, F.F., Zhu, S.C.: Visual Persuasion: Inferring Communicative Intents of Images. In: Proc. IEEE CVPR 2014 (2014)
21. Kapoor, S., Narayanan, A.: We Looked at 78 Election Deepfakes. Political Misinformation Is Not an AI Problem. Knight First Amendment Institute (2024)
22. Korshunov, P., Marcel, S.: DeepFakes: a New Threat to Face Recognition? Assessment and Detection (2018)

23. Lai, A., Brown, M.A., Bisbee, J., Tucker, J.A., Nagler, J., Bonneau, R.: Estimating the Ideology of Political YouTube Videos. Political Analysis **32**(3), 345–360 (2024)
24. Liu, H., Li, C., Wu, Q., Lee, Y.J.: Visual Instruction Tuning. In: NeurIPS 2023 (2023), lLaVA
25. Munger, K., Phillips, J.: Right-Wing YouTube: A Supply and Demand Perspective. International Journal of Press/Politics **27**(1), 186–219 (2022)
26. Peng, Y.: Same Candidates, Different Faces: Uncovering Media Bias in Visual Portrayals of Presidential Candidates with Computer Vision. Journal of Communication **68**(5), 920–941 (2018)
27. Peng, Y.: Athec: A Python Library for Computational Aesthetic Analysis of Visual Media in Social Science Research. Computational Communication Research **4**(1) (2022)
28. Qwen Team: Qwen3-VL Technical Report (2025)
29. Ribeiro, M.H., Ottoni, R., West, R., Almeida, V.A.F., Meira, W.: Auditing radicalization pathways on YouTube. In: Proc. ACM FAT* 2020. pp. 131–141 (2020)
30. Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., Nießner, M.: FaceForensics++: Learning to Detect Manipulated Facial Images. In: Proc. IEEE/CVF ICCV 2019 (2019)
31. Salganik, M.J., Dodds, P.S., Watts, D.J.: Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. Science **311**(5762), 854–856 (2006)
32. Simon, F.M., Altay, S., Mercier, H.: Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. Harvard Kennedy School Misinformation Review **4**(5) (2023)
33. Törnberg, P.: ChatGPT-4 Outperforms Experts and Crowd Workers in Annotating Political Twitter Messages with Zero-Shot Learning (2023)
34. Vaccari, C., Chadwick, A.: Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. Social Media + Society **6**(1), 1–13 (2020)
35. Verdoliva, L.: Media Forensics and DeepFakes: An Overview. IEEE Journal of Selected Topics in Signal Processing **14**(5), 910–932 (2020)
36. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018)
37. YouTube: How we're helping creators disclose altered or synthetic content. YouTube Official Blog (2024)
38. Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., Yang, D.: Can Large Language Models Transform Computational Social Science? Computational Linguistics **50**(1), 237–291 (2024)