Analysis of Faces in a Decade of US Cable TV News

James Hong¹, Will Crichton¹, Haotian Zhang¹, Daniel Y. Fu¹, Jacob Ritchie¹ Jeremy Barenholtz¹, Ben Hannel¹, Xinwei Yao¹, Michaela Murray¹ Geraldine Moriba², Maneesh Agrawala^{1,2}, Kayvon Fatahalian¹ Stanford University¹ Brown Institute for Media Innovation²



Figure 1: (a) Our dataset contains over 244,000 hours of video aired on CNN, FOX, and MSNBC from January 1, 2010 to July 23, 2019. The screen time of news content (and commercials) in our dataset is stable from 2012 onwards, representing near 24/7 coverage. (b) The ratio of time of when male-presenting faces are on-screen to when female-presenting faces are on-screen is 2.1× on average, but has narrowed from 2.4× to 1.9× over the decade. (c) The top 100 people by face screen time (top 10 labeled). Of the 100 people, 18 are US politicians and 85 are news presenters (3 are both). Source images © CNN, Fox News, and MSNBC.

ABSTRACT

Cable (TV) news reaches millions of US households each day. News stakeholders such as communications researchers, journalists, and media monitoring organizations are interested in the visual content of cable news, especially who is on-screen. Manual analysis, however, is labor intensive and limits the size of prior studies.

We conduct a large-scale, quantitative analysis of the faces in a decade of cable news video from the top three US cable news networks (CNN, FOX, and MSNBC), totaling 244,038 hours between January 2010 and July 2019. Our work uses technologies such as automatic face and gender recognition to measure the "screen time" of faces and to enable visual analysis and exploration at scale. Our analysis method gives insight into a broad set of socially relevant topics. For instance, male-presenting faces receive much more screen time than female-presenting faces (2.4x in 2010, 1.9x in 2019).

To make our dataset and annotations accessible, we release a public interface at https://tvnews.stanford.edu that allows the general public to write queries and to perform their own analyses.

KDD '21, August 14-18, 2021, Virtual Event, Singapore

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-8332-5/21/08...\$15.00 https://doi.org/10.1145/3447548.3467134

CCS CONCEPTS

- Applied computing \rightarrow Law, social and behavioral sciences.

KEYWORDS

Cable news; screen time; visual analysis at scale

ACM Reference Format:

James Hong, Will Crichton, Haotian Zhang, Daniel Y. Fu, Jacob Ritchie, Jeremy Barenholtz, Ben Hannel, Xinwei Yao, Michaela Murray, Geraldine Moriba, Maneesh Agrawala, Kayvon Fatahalian. 2021. Analysis of Faces in a Decade of US Cable TV News. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21), August* 14–18, 2021, Virtual Event, Singapore. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3447548.3467134

1 INTRODUCTION

Cable (TV) news reaches millions of US households each day [19], and national news networks such as CNN, FOX, and MSNBC exert great influence over public opinion and discourse on current events. While cable news has been on air for over 40 years, there has been little longitudinal analysis of its visual content, such as the presence of people on-screen. Despite this, cable news contains answers to many questions that are of great interest to academic researchers (studying gender, age, or political bias); journalists and opinion writers (commenting on current and past events); media monitors and watchdogs (which already collect such information manually); and the general public. For example, *How often are faces on-screen in cable news? What is the screen time of men vs. women? Which political*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

candidates and news presenters receive the most screen time? How are victims and perpetrators of violence portrayed? Who is on-screen when different topics are discussed?

In this paper, we investigate how AI, in the form of automatic annotation by off-the-shelf models, at scale, can provide answers to such socially relevant questions by analyzing a decade of US cable news video. Our dataset, obtained in collaboration with the Internet Archive [2], contains over 244,038 hours of cable news programming. It represents a near-uninterrupted 24/7 video stream from January 2010 to July 2019 from each of the three largest cable news networks in the US – CNN, FOX, and MSNBC, which serve both liberal (CNN, MSNBC) and conservative (FOX) leaning audiences [23]. Automating annotation and analysis of the people (i.e., faces, public figures, gender, etc.) in this video dataset would allow large-scale studies by media researchers; data exploration; and fact-checking and corroboration of trends found in prior work on other datasets and news mediums (e.g., print, social media).

Several challenges make automated analysis of cable news interesting and difficult: (1) the mismatch between what off-the-shelf models measure/detect (e.g., faces) and the questions of interest to stakeholders (e.g., "are women under-represented?"); (2) the large amount of video data; and (3) the potential for unreliable model predictions. To address (1), we focus on counting the "screen time" of faces and show that by presenting results in the form of screen time and time-series plots, we can reveal a variety of insights, patterns, and trends. Concretely, we define "screen time" as the temporal interval of a phenomenon in the video. It can refer to faces (i.e., the time a face is on-screen), to faces with other attributes such as identity (i.e., the time Donald Trump is on-screen), and to words (i.e., the time during which a word is said). To enable analysis at scale (2), we preprocess the dataset by detecting faces, predicting presented binary gender, identifying prominent public figures, and aligning text captions to audio (at sub-second granularity). To quantify uncertainty (3), we validate each stage of our processing pipeline and the labels that it produces. Our approach differs from prior news analyses that have largely relied on manual annotation of video, focused on text, or reported findings from small data samples.

In addition to our analysis, we release a public analysis tool built on top of our annotation database, akin to the Google Ngram Viewer [25], which demonstrated the broad applications of usable, interactive word frequency analysis of 5.2 million books and print media from 1800 to 2000. Our tool is updated daily and allows interactive queries on the full dataset from 2010 to the present day (290,000+ hours as of May 2021). We hope that releasing our annotation database and a query interface will encourage other researchers (including non-technical), journalists, and the general public to perform their own analyses on the cable news dataset.

In summary, we make three main contributions:

(1) We conduct a large-scale, longitudinal study of the last decade (January 1, 2010 to July 23, 2019) of US cable news. Our experiments demonstrate the value of applying face recognition technologies to analyze screen time at scale, and we evaluate this claim by answering a diverse set of questions over relevant social issues such as gender representation, visual bias in news content, and cable news presentation that are of interest to media researchers, journalists, activists, and other news stakeholders (section 4).

- (2) We describe and validate our data processing pipeline, which produces the annotations used in our large-scale analysis, and we document challenges that arise in applying off-theshelf and our own models to sensitive social topics and quantify the uncertainty for each set of data labels (section 3).
- (3) We present a web-based data analysis tool that enables other researchers, journalists, and the general public to leverage our annotations and write their own queries on the full cable news dataset, now containing over 290,000+ hours of news video from 2010 to the present. Our design prioritizes interactivity, accessibility, and the ability to inspect underlying data to validate results (section 5).

Our analysis tool is available at https://tvnews.stanford.edu along with an extended technical report and analysis. The code for data processing and visualization is open-source and can be used to visualize other large, longitudinal video datasets.

Statement on social responsibility

Analysis of news media can touch upon sensitive social topics. Likewise, AI, when applied without due care, can aggravate existing biases or introduce new ones [8, 11]. We adopt a policy of minimizing the potential for harm. Two examples of such harms include (1) inaccurate conclusions due to systematic labeling error and (2) inaccurate labeling of individuals (e.g., misgendering), which can be hurtful and a possible vector for abuse. We address (1) by statistically validating every stage of our pipeline and qualitatively vetting our results using our (public) analysis tool (section 5), and we report results in this paper that are robust under this examination. (2) is addressed in section 3. We believe that the benefits enabled by our application, in enabling news stakeholders to scrutinize the last decade of US cable news, outweigh the potential for harm.

2 RELATED WORK

Face recognition in media analysis Applying face detection to video analysis is not a new idea. [30] compares the accuracy of face detection and text counting metrics for measuring politicians' screen time, showing that deep neural network based face detectors can produce accurate estimates of face screen time in a dataset of 2,102 hours of NHK News 7 video. [31] uses face recognition on the same dataset to visualize connections between politicians as a graph. The Geena Davis Institute [16] uses gender classification to measure gender inclusivity (screen time and speaking time) in 200 Hollywood films from 2014 and 2015. Like [16, 31], we analyze annotations such as faces, but with orders of magnitude more video, an emphasis on time-series, and a set of analysis questions around US cable news. More broadly, we view our work as applying relevant ideas from visual search [22] to social questions about people in our video dataset.

Manual analysis of TV news and media Audio-visual analysis of TV news is well-established in communications research, and the breadth of topics in the literature indicates that there is academic and public interest in answering questions of representation, coverage bias, and presentation (in cable news and media in general). [20] studied how sound-bite coverage of US presidential elections changed from 1968-1988, noting that the average duration of TV soundbites fell over 4x in that time period. [3, 6] extend [20] to



Figure 2: Video processing pipeline.

"image-bites", where candidates are shown but not heard. [24] compared broadcast and cable news during the 1996 presidential elections in Taiwan, finding that state-owned broadcast media (unfairly) spent more time covering the ruling party. Several long running media monitoring projects measure gender [5, 17, 32, 35] and race [32] representation in news. A common approach is to systematically sample a collection of sound-bites, image-bites, newscasts, etc. and to manually code (annotate) them. Large scale studies are rare; most studies are limited to a few hundred hours/bites [6, 20] or a few shows [32]. GMMP [17], the largest study of gender in news, employs a network of volunteers across 100+ countries and publishes reports every five years. Similar analyses also appear in non-peerreviewed articles by journalists and columnists [10, 14, 27].

Our work is complementary to the above. "Screen time" is a finegrained, visual metric of salience in video, similar to non-visual metrics such as word counts and hand-curated lists of discrete names, bites, and airings. Annotating a near-complete decade of news video enables rapid exploration of dataset slices (and time resolutions) that would be cost prohibitive to sample by hand, while also controlling for bias in what historical data is preserved and analyzed. This can lead to surprising observations that motivate new traditional research; demonstrate existing concepts in novel data contexts; or challenge prior assumptions and interpretations. Public analysis tools Our analysis tool is heavily inspired by the Google Ngram viewer [25] and Google Trends [18], which demonstrate that automated computational analysis of word frequency, when performed at scale on centuries of digitized books or the world's Internet search queries, can serve as a valuable tool for studying trends in culture. These projects allow the general public to conduct analyses by creating simple time-series visualizations of word frequencies. We view our work as bringing these ideas to cable news video. The GDELT AI Television Explorer [15], provides a web-based query interface for caption text, on-screen chyron text, and black-box results from the Google Cloud Video and Natural Language APIs. We analyze nearly the same corpus of video, but, unlike GDELT, we label the video with information about the faces on-screen and estimate the accuracy of these labels in our dataset. Prior studies [8] and our experience with AWS Rekognition Celebrity Recognition [1] suggest that validation on application domain data, through statistical or qualitative means, is crucial to quantifying bias and avoiding incorrect/misleading conclusions.

3 METHODOLOGY

Our methodology is designed to support a wide variety of exploratory analyses on faces, motivated by questions from news stakeholders. We improve and validate our models as necessary to trust our results. To this end, we prefer simpler annotations and methods (e.g., counting faces, n-grams) that can be applied uniformly across the entire dataset, and validated by random sampling, over more complex methods (e.g., topic models, action recognition) that are applicable to and require validation on specific slices of data. Our annotation pipeline is shown in Fig. 2 and statistics are given in this section and in section A.1.

3.1 Data processing

Our core dataset consists of 244,038 hours of video, audio, and captions recorded by the Internet Archive's TV News Archive [2]. It is segmented into 215,771 standard definition videos, organized by the date and time of airing; channel (i.e., CNN, FOX, or MSNBC); and the name of the show (e.g., "Fox and Friends"). The amount of video available over the decade, per month, is shown in Fig. 1a.

Commercial detection Commercials make up approximately 27.9% of the video aired. We detect commercials using a number of heuristics: commercial segments in the dataset are often bracketed by black frames, have captions in mixed/lower case (as opposed to all uppercase for news content), or are missing caption text entirely. Commercials also do not contain » delimiters (for speaker changes). To validate our detection algorithm, we hand annotated 225 hours of videos with 61.8 hours of commercials. The precision and recall on this sample are 93.0% and 96.8%, respectively. We focus our analysis on the news content portion of the dataset.

Face detection We use MTCNN [36] to detect faces in a uniform sample of frames at every three seconds. Subsampling reduces computational cost and is possible because cable news consists of slow changing content (three seconds is approximately half the average shot length of 6.2 seconds between camera cuts). At this sample rate, we detect 306 million faces in total, of which 263 million lie in non-commercial video frames. For each detected face, we compute a 128-dimensional FaceNet [33] descriptor using the pixels contained within the face's bounding box. These descriptors are used to compute additional annotations such as binary gender presentation and person identification.

From a stratified random sample across the 10 years (see section A.1), we estimate the precision and recall of the face detector to be 98.5% and 74.5%, respectively. The majority of missed detections are on small (e.g., in the chyrons) or background (out-of-focus) faces. Note that we do not differentiate between faces in different contexts; a face can belong to a news anchor or guest in the studio; or be a still image, B-roll footage, or even part of an infographic.

Gender classification We annotate each face's presented (or apparent) binary gender using a k-NN classifier, with FaceNet descriptors as input features. The k-NN model indexes 12,669 manually annotated faces, sampled representatively from our dataset. On an independent sample of 6,000 faces, the k-NN classifier has 97.2% agreement with human annotators. In our validation, we found that our k-NN classifier outperforms standard, pre-trained models applied to our data such as [12] (\approx 91%). While more accurate models certainly exist, we find that k-NN, with representative data, is adequate, interpretable, and trivially updatable (if errors for recurring individuals are discovered).

We acknowledge that treating gender as a binary quantity fails to represent many transgender individuals [21]. An individual's appearance or gender presentation may not reflect their actual gender. Despite these simplifications, we believe that automatically estimating presented binary gender (distinguished as male- and female-presenting in this paper) is useful for understanding gender representation in media, given the importance of initiatives such as BBC 50:50 [5] and GMMP [17]. To minimize harm and distress to individuals, who may be mislabeled, we focus on aggregate statistics. We do not expose gender queries in our public tool, except to vetted parties, to reduce the potential for online abuse.

Face identification using AWS Rekognition We use the AWS Rekognition Celebrity Recognition API [1] to identify public figures. The API labels 46.2% of the faces in the dataset and we propagate these detections to an additional 10.7% of faces using k-NN on FaceNet [33] descriptors (within each video), reducing flickering.

We manually validate the precision and recall for the key public figures in our analyses and include statistics in section A.1. Among the issues that we discovered in validation are (1) low recall on some individuals, especially the most prominent ones such as Donald Trump, and (2) the "doppelgänger" problem, when a recurring person in the news is mislabeled as a visually similar celebrity. We discuss the doppelgänger problem further in section A.1, Table 3. The politicians for which we use AWS generated labels have a mean estimated precision and recall of 97.4% and 88.0%, respectively.

Face identification using our own models We employ our own models to identify significant individuals (such as news presenters) and politicians missed (or considered too inaccurate to use) by AWS Rekognition. These individuals include Donald Trump, Hillary Clinton, and other important politicians (for parity), and 27 additional news presenters. Our models are trained on FaceNet [33] descriptors, and statistics are also included in section A.1, Fig. 14.

News presenters While annotations are computed at the granularity of frames and faces, we also join manually curated information about news presenters. We use the term "news presenter" to refer broadly to anchors, hosts, and on-air staff (e.g., contributors, meteorologists, etc.) of a news network, and we manually enumerate 325 news presenters from the three networks. Their names are compiled from the public staff listings of each network as well as Wikipedia entries for past shows since 2010 (including the top 150 shows by duration, accounting for 96% of news content). In addition to name, we collected available information such as date-of-birth (or year if the former is unknown) and attributes such as hair color.

Caption-time alignment Closed captions are available for 94.9% of the videos. We use the Gentle word aligner [28] to perform sub-second alignment of words in a video's captions to the video's audio track, assigning each word a starting and ending time. (The source captions are only coarsely aligned to the video.) Alignment is considered successful if alignments are found for at least 80% of the words in the captions, interpolating any that are missed. By this metric, we are able to align captions for 92.4% of the videos. The average time it takes to speak a word in our dataset is 219ms.

3.2 Screen time as a unit of measurement

We use "screen time" as the quantity of measurement in our analyses. Each face detection has a temporal extent, defined as 3 seconds (the sample rate). Attributes of faces, such as presented gender and identity, inherit their temporal extents and screen time from faces. When aggregating the screen time of faces (under a predicate), the union of the temporal extents is taken. For example, the "screen time of female-presenting faces" is the amount of time when "at least one female-presenting face is on-screen." Likewise, a "percentage of screen time" can be interpreted as the probability that a predicate holds when viewing the cable stream at a random instant.

Time-aligned words also have temporal extents and intersecting them with face screen time allows us to compute probabilities such as when a female-presenting face is on-screen when a word is said.

4 RESULTS

We report our findings in three sections: analysis of *face screen time*; a case study on *visual portrayal of individuals*; and analysis of *faces and caption text*. In addition to asking novel questions, we guide our analysis topics by evaluating prior claims (both scientific and anecdotal) by media researchers and journalists. Our contribution in these replication examples is to demonstrate that automatic labels can efficiently verify and expand upon existing work with much greater scale, a fraction of the human labor, and novel slices of data.

4.1 Who is in the news?

How much time is there at least one face on-screen? People are an important part of cable news as both subjects and storytellers (news presenters). At least one face appears on-screen 75.3% of the time. Interestingly, over the decade, the percentage of time with at least one face on-screen has risen steadily from 72.9% in 2010 to 81.5% in 2019 and is similar across all three channels (Fig. 3).

The average number of faces on-screen has also increased (from 1.2 to 1.6). On CNN and FOX, this is marked by a decline in the amount of time when only one face is on-screen, while that value has remained constant on MSNBC. On all three channels, the amount of time when multiple faces (2 or more) are on-screen simultaneously has risen, accounting for the increase in time when at least one face is on-screen. In contrast, the average number of faces on-screen in commercials has increased less (from 0.42 to 0.52) and has remained flat on all three channels since 2013, suggesting that the increase is not due to changes in face detection (on higher quality video). While we refrain from speculating on the causes of increased face screen time, the data do suggest longitudinal shift in how cable news is presented.

How does screen time of male-presenting individuals compare to that of female-presenting individuals? Male-presenting faces are on-screen 60.2% of the time, compared to 28.7% of the time for female-presenting faces, a 2.1 to 1 ratio. These percentages are similar across channels and have slowly increased for both groups. Correspondingly, the ratio of male- to female-presenting screen time has narrowed from $2.4 \times$ to $1.9 \times$ over the decade (Fig. 1b), as both gender groups have gained screen time at a similar pace. While the trend indicates movement towards gender parity in screen time, the rate of change is slow, and these results also reinforce prior observations on long-running under-representation of women in media, in news [17] and film [16].

Which public figures receive the most screen time? We identify 1,260 unique individuals who receive at least 10 hours of screen time in our dataset. These individuals account for 47% of the 263 million faces that we detect in the news content and are on-screen for 45%



Figure 3: The percentage of time when at least one face is on-screen has increased on all three channels (thick lines), mostly between 2015 and 2018. The amount of time when multiple faces are on-screen has also increased, while the amount of time with only one face on-screen has declined on CNN and FOX, and stagnated on MSNBC.



Figure 4: The percentages of time when male- and femalepresenting faces are on-screen are similar on all three channels and have also increased with the rise in all faces noted in Fig. 3. Male- and female-presenting faces can be on-screen simultaneously so the lines can add to over 100%.

of total screen time. The top individual is Donald Trump, who rose to prominence in the 2016 presidential campaigning season and his presidency from 2017 to 2021 (Fig. 1c). Barack Obama is second, with 0.63× Trump's screen time, and is prevalent between 2010 (the start of the dataset) and 2017 (the end of his second term as president). Besides US presidents, the other top individuals consist almost exclusively of US politicians and news presenters (Fig. 5).

How much screen time do political candidates get before an *election?* In US elections, aspiring candidates first compete in (party) primary elections to become their party's nominee in the general election. Recognition and the amount of media publicity can influence perceptions of candidates in these critical months.

During the 2016 Republican presidential primaries, Donald Trump consistently received more screen time than other candidates (Fig. 6a). In the competitive primary season, from January to May 2016, Trump received 342 hours of screen time, while his closest Republican rival, Ted Cruz, received only 130 hours. In the same timespan, the leading Democratic candidates, Hillary Clinton and Bernie Sanders received more equal screen time (164 hours and 139 hours, respectively); both received far more screen time than the other Democratic primary candidates (Fig. 6b). Comparing the two presidential nominees, during the period from January 1, 2016 to the November 8, 2016, general election, Trump received 1.9× more screen time than Clinton. This disparity largely holds across all three channels and directly supports [10]'s claim that the media disproportionately focused on Trump in 2016.

Our dataset also allows a retrospective comparison with the 2012 election, between then President Barack Obama and Republican challenger Mitt Romney. Fig. 6c shows that Romney did not hold the edge in coverage until he became the presumptive Republican nominee. A more recent comparison between (Joe) Biden and Trump in



Figure 5: Distribution of individuals' screen time (stacked), separated by news presenters on each channel and other. 65% of the individuals with 100+ hours of screen time are news presenters. Note: the three leftmost bars are truncated.

2020, plotted in our public tool, shows over-representation toward Trump by $1.7 \times$ in the three months before the general election.

Who presents the news? A news presenter is on-screen 28.1% of the time – 27.4% on CNN, 33.5% on FOX, and 23.0% on MSNBC. On CNN, the percentage of time that a news presenter is on-screen increases by 13% between 2015 and 2018, while it remains mostly flat over the decade on FOX and MSNBC (Fig. 7a). Time is not shared equally, the top 5 presenters on each channel account for 31%, 22%, and 34% of this screen time on CNN, FOX, and MSNBC.

We can also shed light on gender representation in the newsroom. Across all three channels, there is shift towards gender parity in the screen time of news presenters early in the decade followed by divergence (Fig. 7b-d). Looking closely, CNN reaches genderscreen-time parity for news presenters in January-June 2012 and May-August 2015 (Fig. 7b). However, CNN diverges from September 2015 onward, as a 10% increase in male presenters (14% to 24%) outpaces a 3% increase for their female peers (13% to 16%). This gap is explained by Anderson Cooper, Don Lemon, and Chris Cuomo, who see $2.5 \times$, $4.5 \times$, and $5.5 \times$ growths in screen time since 2015. The disparity on FOX narrows from 2010 to 2016, but widens in 2017 due to an increase in the screen time of male presenters (Fig. 7c) near the departure of several female anchors (around the time of the Roger Ailes scandal). On MSNBC, the disparity as a percentage of screen time increases from May 2017 to July 2019 (Fig. 7d). This is due to a simultaneous drop in the screen time of both male and female presenters: from 17% to 13% and from 14% to 7%, respectively.

What is the average age of news presenters? Age disparity is a sensitive issue in media [17]; prior research on women in TV reporting found greater rates of turnover compared to male peers [13]. To explore changes in the age composition of news presenters across the decade, we measure the age-weighted screen time of faces. Age, of a face, is calculated as a difference between the broadcast date and presenter's date of birth, and its weighted sum can be interpreted as the expected age of a news presenter, as visible on-screen.

We discover that, overall, female presenters are younger on average than their male counterparts by 6.3 years. However, the gap has narrowed in recent years. News presenters are also becoming older on average; rising from 48.2 to 51.0 years (Fig. 8). This is visible on all three channels, though there are localized reversals that often correspond to the retirements of prominent hosts: for example, on CNN after Larry King's retirement in 2010 at age 76.

Are female news presenters on FOX disproportionately blonde? We assessed the prominence of hair colors among female news presenters, weighted by face screen time. This was done by manually attributing hair color (blonde, brown, black, other) to 145 female news presenters and aggregating their screen time. Blondes



Figure 6: (a) In 2016, Donald Trump received significantly more screen time than the other Republican candidates. (b) Hillary Clinton and Bernie Sanders received nearly equal screen time during the competitive primary season (January-May 2016). (c) In 2012, Mitt Romney did not overtake the other Republican candidates until after he became the presumptive nominee.



Figure 7: (a) The percentage of time when a news presenter is on-screen has remained mostly flat on FOX and MSNBC, but has risen by 13% on CNN since 2016. (b-d) The screen time of news presenters by presented gender (as a percentage of total news presenter screen time) varies across the decade. CNN reaches parity in 2012 and 2015, but has since diverged. Because male- and female-presenting news presenters can be on-screen simultaneously, the lines can add to over 100%.



Figure 8: The average age of news presenters, weighted by screen time, has increased (bold lines). FOX has the highest average age for both male and female news presenters.



Figure 9: Blonde female news presenters consistently receive more screen time on FOX than non-blonde female news presenters. CNN catches up to FOX from 2014 onward. The screen time of blonde female news presenters has risen on MSNBC since 2015, but does not exceed that of non-blonde female news presenters.



Figure 10: The 25 news presenters who receive the largest fraction of screen time on their own shows ("screenhogs") and the total amount of video content for their shows in the dataset. The top two shows by this metric, *Cuomo Primetime* and *Tucker Carlson Tonight*, are relatively recent shows, starting in June 2018 and November 2016, respectively.

account for 64.7% of female news presenter screen time on FOX, compared to 28.8% for non-blondes. This supports the stereotype that female news presenters on FOX fit a particular aesthetic [14]. Despite the stereotype being most widely attributed to FOX, FOX is not alone (Fig. 9); the proportion of blondes on CNN has risen, and the chance of seeing a blonde female news presenter on CNN is approximately equal to that on FOX (56.6% compared to 38.6% for non-blondes). The screen time of blonde female news presenters is lower on MSNBC (36.6%), where non-blonde female news presenters is lower on SNBC (36.6%), where non-blonde female news presenters account for 55.7%; brown is the dominant hair color at 40.8%, but 21.4% is due to a single brown-haired host (Rachel Maddow). On all three channels, the percentage of blonde female news presenters far exceeds the natural rate of blondeness in the US (\approx 11% [9]).

Which news presenters "hog" the screen time on their shows? We measure the percentage of time a news presenter is on-screen on their own show and plot the top 25 "screenhogs" (Fig. 10). Chris Cuomo (CNN) has the highest fraction of screen time on his own show (visible 70.6% of the time on *Cuomo Primetime*). Tucker Carlson (FOX) is second at 55.3% on *Tucker Carlson Tonight*. Both individuals are hosts of primetime shows, often featuring interviews, political opinion, and debate. Carlson also monologues frequently.

4.2 How are individuals portrayed?

News organizations control the images and graphics used to tell a story. We perform a case study around the Trayvon Martin shooting on February 26, 2012 in Sanford, Florida. Martin, an unarmed 17 year-old African-American high-school student, was fatally shot by neighborhood watch coordinator George Zimmerman. The circumstances of the incident were highly publicized; raised allegations of racial profiling and excessive use of force; and elicited polarized reactions from different political and social groups in the US.

Did each channel show different images of Trayvon Martin and George Zimmerman? Media depictions of both Martin and Zimmerman were scrutinized heavily as the story captured the national interest [27, 34]. The object of this scrutiny is captured in our dataset. We identified unique images (and video appearances)



Figure 11: In early coverage of the shooting of Trayvon Martin (by George Zimmerman) in 2012, all three channels used the same photos of Martin and Zimmerman. However, as the story progressed, depictions of Trayvon (left) differed significantly across channels. Depictions of Zimmerman (right) also evolved over time but largely reflect efforts by channels to use the most up-to-date imagery of Zimmerman, especially in the 2013 trial. Grey lines are the total screen time.

of Martin and Zimmerman in our dataset and tabulated the screen time of these images (see section A.1 for details).

Fig. 11 shows the four images of Martin (left) and Zimmerman (right) that received the most screen time in (1) the aftermath of the shooting and (2) during Zimmerman's 2013 trial. While the ranking by total screen time differed by channel (Fig. 11-top), the time-series tell a more complete story about how depictions evolved. In the initial week of coverage, all three channels used the same image of Martin (purple). This image generated significant discussion about the "baby-faced" depiction of Martin, although it was dated to a few months before the shooting. In the ensuing weeks (and later during Zimmerman's trial), differences in how the three channels depict Martin emerge. CNN most commonly used a photograph of Martin smiling in a blue hat (blue box). In contrast, the most commonly shown photo on FOX depicts an unsmiling Martin (orange). MSNBC most frequently used the black-and-white image of Martin in a hoodie (pink) that was the symbol for protests in support of Martin and his family. The three different images reflect significant differences in editorial decisions made by the three channels.

Depictions of Zimmerman also evolved with coverage of the shooting and reflect both editorial choices as well as efforts by the channels to use the most up-to-date photos for the story at hand. All three channels initially aired the same image of Zimmerman (purple). The photo, depicting Zimmerman in an orange polo shirt, was both out of date and taken from a prior police incident unrelated to the Martin shooting. A more recent photograph of Zimmerman (pink) was made available to news outlets in late March 2012. While FOX transitioned to using this new photo, which depicts a smiling Zimmerman, CNN and MSNBC continued to use both photos until mid-April. After mid-April 2012, depictions of Zimmerman on all three channels show him, primarily, in courtroom appearances.



Figure 12: The distribution of words by the difference in conditional probability of a female- and male-presenting face being on-screen. Note the stark differences in topic representation in the top 35 male and female associated words: foreign policy, conflict, and fiscal terms (male); and female health, family, weather, and business news terms (female).

4.3 What is said when people are on-screen?

Which words are most likely to be said when female-presenting faces are on-screen? We compute the conditional probability of observing at least one female-presenting (or one male-presenting) face on-screen when each word is said in the caption text (details in section A.1). Because of the gender imbalance in screen time, the conditional probability of a female-presenting face being on-screen when *any* arbitrary word is said is 29.6%, compared to 61.4% for a male-presenting face. We are interested in words where the difference between the female and male conditional probabilities deviates from the baseline 31.9% difference.

Fig. 12 shows the words that are most associated with male- and female-presenting faces on-screen, providing insight into the top female and male topics. Womens' health (e.g., breast, pregnant) and family (boyfriend, husband, mom(s), parenthood) are common female terms, as are gendered job titles (actress, congresswoman). Other topics such as weather (temperatures, meteorologist, blizzard, tornadoes) and business (futures, Nasdaq, stocks, earnings) are also near parity due to female weatherpersons (Indra Petersons/CNN, Janice Dean/FOX, Maria Molina/FOX) and business correspondents (Christine Romans/CNN, Alison Kosik/CNN, JJ Ramberg/MSNBC, Stephanie Ruhle/MSNBC, Maria Bartiromo/FOX). By contrast, the top words associated with male-presenting faces are related to foreign affairs, terrorism, and conflict (ISIL, Israelis, Iranians, Saudis, Russians, destroy, treaty); and to fiscal policy (deficits, trillion, entitlement(s)). While many underlying factors such as news presenters' and news subjects' presented genders affect the degree to which words audio-visually associate by gender, the stark difference in the topics represented motivates further investigation.

5 PUBLIC VISUALIZATION TOOL

Our interactive, web-based analysis tool enables the general public to perform analyses of the cable news dataset (Fig. 13). The tool generates time-series line charts of the amount of cable news video (aggregate time) matching user-specified queries. Queries may consist of one or more filters that select intervals of time when an individual appears on-screen (name="..."); an on-screen face has a presented gender (tag="male"); a keyword or phrase appears in



Figure 13: Our public analysis tool supports interactive time-series visualization of the cable news dataset. (Left) Users define queries using a combination of face, caption text, and video metadata filters. The tool plots the total amount of video (aggregate screen time) matching these queries. (Right) To provide more context for the segments of video included in the chart, users can click on the chart to bring up the videos matching the query. We have found that providing direct access to the videos is often essential for debugging queries and understanding the relevant video clips. Source images © CNN, Fox News, and MSNBC.

the captions (text="..."); or the videos come from a particular channel (channel="CNN"), show, or time of day.

To construct more complex analyses, the tool supports queries containing conjunctions and disjunctions of filters, which serve to intersect or union the video time intervals matched by individual filters (name="Hillary Clinton" AND text="email" AND channel="FOX"). We implemented a custom in-memory query processing system to execute screen time aggregation queries over the entire cable news dataset while maintaining interactive response times for the user. In addition to generating time-series plots of video time, the tool enables users to directly view underlying clips (embedded from [2]) that match queries by clicking on the chart.

A major challenge when developing this tool was making an easy-to-use, broadly accessible data analysis interface while still exposing sufficient functionality to support a wide range of analyses on faces and text in cable news. We call out three design decisions made during the tool's development.

(1) Limit visualization to time-series plots Time-series analysis is a powerful way to discover and observe patterns over the decade spanned by the cable news dataset. While time-series do not encompass the full breadth of analyses presented in this paper, we chose to focus the tool's design on the creation of time-series plots to encourage and simplify this important form of analysis.

(2) Use screen time as a metric We constrain all queries, regardless of whether visual filters or caption text filters are used, to generate counts of a single metric: the amount of screen time matching the query. Although alternative metrics, such as using word counts (to analyze of caption text) or counts of distinct individuals (to understand who appears on a show), may be preferred for certain analyses, we chose screen time because it is well suited to many analyses focused on understanding representation in the news. For example, a count of a face's screen time directly reflects the chance a viewer will see the face when turning on cable news. Word counts can be converted into screen time intervals by attributing each instance of a word, regardless of its actual temporal extent, to a fixed interval of time (textwindow="...").

The decision to make all filters select temporal extents simplifies

the query interface. All filters result in a selection of time intervals, allowing all filters to be arbitrarily composed. A system where some filters yield word counts and others yield time intervals would complicate the user experience by introducing the notion of different data types into queries.

(3) Facilitate inspection of source video clips It is important for the visualization tool to support user inspection of the source video clips that match a query (Fig. 13-right). Video clip inspection allows a user to observe the context in which a face or word appears in a video. This context in turn is helpful for understanding why a clip was included in a query result, which facilitates a deeper understanding of trends being investigated; aids the process of debugging and refining queries; and helps a user assess the accuracy of the automatically generated video labels relied on by a query.

6 **DISCUSSION**

Our results show that measuring the screen time of faces in a decade of cable news reveals patterns that are of interest to many news stakeholders. Here, we discuss the impact of our tool, our metrics and validation, and future directions.

Impact of the tool We released the tool in August 2020 to the general public. The tool has been used by journalists in several articles [7, 26] and several media research organizations have expressed interest in using the tool to conduct and verify their studies. Our tool and processing pipeline is applicable to other datasets, and versions of our tool have been used at two national European news broadcasters on their internal video archives.

Screen time of faces vs. n-gram counts Screen time of faces and counts of name mentions in the captions can be highly correlated; on days when an individual's coverage peaks, their screen time often spikes too. The metrics diverge, however, when an individual is shown but not being discussed or mentioned by name. For example, less-known presidential candidates are often on-screen (e.g., being interviewed or as a static image with other candidates) but rarely discussed. News presenters too are often on-screen but rarely addressed by name. Interestingly, Trump is mentioned 31%

more than Biden in the three months leading up to the 2020 US presidential election but received 68% more screen time.

Validation at scale We validated each pipeline component independently over random samples of data. Modularity makes manual labeling tractable and is necessary when exploring new downstream models, labels, and questions. However, it does not capture correlated error across components (e.g., a face detector that has a higher error rate by gender or race [29]). As such, we view our findings as salient disparities that motivate further study rather than exact measurements. As future work, our system and public tool would benefit from improved statistical modeling and presentation of errors (such as automatic sanity checking against known biases).

New annotations of high-level audio-visual concepts such as the context in which face appears (e.g., live video, still images, vs. replays; foreground vs. background), who is speaking (and their gender), face expressions, and topic modeling would enable a richer set of analyses but also require additional validation. Providing these annotations to users in a flexible and transparent way, to support interactive and higher-level analysis, is an exciting topic of research. Incorporating outside data such as polling statistics and viewer demographics would also enable analysis of how cable news impacts politics, elections, and viewers more generally.

7 CONCLUSION

We have demonstrated that applying off-the-shelf face recognition technology to measure screen time in a decade of cable news video enables interesting analysis on a broad range of topics in cable news, motivated by real-world communications research and journalism. The analyses in this paper were conducted by a small team of researchers and show the amplifying effect that visual analysis at scale can have for social science. We validated our processing pipeline and, guided by our experiences, have released a public analysis tool to enable other researchers and the general public to perform their own analyses at scale. The tool and dataset are updated daily, and we hope it will encourage further research into the presentation of this important form of news media.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation (NSF) under IIS-1539069 and III-1908727. This work was also supported by financial and computing gifts from the Brown Institute for Media Innovation, Intel Corporation, Google Cloud, and Amazon Web Services. We thank the Internet Archive for providing their dataset for academic use. Any opinions, findings, and conclusions expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsors.

REFERENCES

- [1] Amazon. 2020. Amazon Rekognition. https://aws.amazon.com/rekognition.
- [2] Internet Archive. 2020. TV News Archive. https://archive.org/details/tv.
- [3] Kevin G. Barnhurst and Catherine A. Steele. 1997. Image-Bite News: The Visual Coverage of Elections on U.S. Television, 1968-1992. Harvard International Journal of Press/Politics 2, 1 (1997), 40–58.
- [4] Edward Loper Bird, Steven and Ewan Klein. 2009. Natural Language Processing with Python. O'Reilly Media Inc, Sebastopol, CA, USA.
- [5] British Broadcasting Corporation (BBC). 2020. 50:50 Project.
- [6] Erik P. Bucy and Maria Elizabeth Grabe. 2007. Taking Television Seriously: A Sound and Image Bite Analysis of Presidential Campaign Coverage, 1992–2004. *Journal of Communication* 57, 4 (12 2007), 652–675.

- [7] Philip Bump. 2020. Turns out that media coverage of the virus wasn't just a ploy to hurt Trump. *The Washington Post* (Nov 2020).
- [8] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency. 77–91.
- [9] Bureau of Labor Statistics, U.S. Department of Labor. 2019. National Longitudinal Survey of Youth 1979 cohort, 1979-2016 (rounds 1-27).
- [10] Nicholas Confessore and Karen Yourish. 2016. \$2 Billion Worth of Free Media for Donald Trump. *The New York Times* (Mar 2016).
- [11] Kate Crawford and Trevor Paglen. 2019. Excavating AI: The Politics of Images in Machine Learning Training Sets. https://www.excavating.ai/.
- [12] Daniel Pressel. 2016. Rude Carnie: Age and Gender Deep Learning with Tensor-Flow. https://github.com/dpressel/rude-carnie.
- [13] Joe S. Foote. 1998. "Women Correspondent Visibility 1983-1997. Proceedings of the Annual Meeting of the Association for Education in Journalism and Mass Communication (81st, Baltimore, Maryland, August 5-8, 1998). Commission on the Status of Women. (Aug 1998).
- [14] Hadley Freeman. 2017. Why do all the women on Fox News look and dress alike? Republicans prefer blondes. *The Guardian* (Feb 2017).
- [15] GDELT Project. 2020. The GDELT AI Television Explorer.
- [16] Geena Davis Institute. 2016. The Reel Truth: Women Aren't Seen or Heard: An Automated Analysis of Gender Representation in Popular Films.
- [17] Global Media Monitoring Project. 2015. Who Makes the News? https:// whomakesthenews.org/gmmp-2015-reports/.
- [18] Google. 2020. Google Trends. https://trends.google.com/trends/.
- [19] Jeffrey Gottfried and Elisa Shearer. 2017. Americans' online news use is closing in on TV news use. *Pew Research Center* (2017).
- [20] Daniel C. Hallin. 2006. Sound Bite News: Television Coverage of Elections, 1968–1988. Journal of Communication 42, 2 (02 2006), 5–24.
- [21] Foad Hamidi, Morgan Klaus Scheuerman, and Stacy M. Branham. 2018. Gender Recognition or Gender Reductionism? The Social Implications of Embedded Gender Recognition Systems. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, Article 8, 13 pages.
- [22] Yushi Jing, David Liu, Dmitry Kislyuk, Andrew Zhai, Jiajing Xu, Jeff Donahue, and Sarah Tavel. 2015. Visual Search at Pinterest. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Sydney, NSW, Australia) (KDD '15). Association for Computing Machinery, New York, NY, USA, 1889–1898.
- [23] Mark Jurkowitz, Amy Mitchell, Elisa Shearer, and Mason Walker. 2020. U.S. Media Polarization and the 2020 Election: A Nation Divided.
- [24] Ven-hwei Lo, Pu-tsung King, Ching-ho Chen, and Hwei-lin Huang. 1996. Political bias in the news coverage of Taiwan's first presidential election: A comparative analysis of broadcast TV and cable TV news. Asian Journal of Communication 6, 2 (1996), 43–64.
- [25] Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. 2011. Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* 331, 6014 (2011), 176–182.
- [26] Musa al-Gharbi. 2020. Cable news profits from its obsession with Trump. Viewers are the only victims. *Columbia Journalism Review* (Sep 2020).
- [27] FOX News. 2016. Trayvon Martin Case: Photos Can be Deceiving. (Mar 2016).
- [28] Robert Ochshorn. 2018. Gentle forced aligner.
- [29] Inioluwa Deborah Raji, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. 2020. Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (New York, NY, USA) (AIES '20). Association for Computing Machinery, New York, NY, USA, 145–151.
- [30] Haolin Ren, Fan Yang, Benjamin Renoust, Yusuke Matsui, Tetsuro Kobayashi, and Shin'ichi Satoh. 2019. Evaluating Face Tracking for Political Analysis in Japanese News Over a Long Period of Time. In IEEE/WIC/ACM International Conference on Web Intelligence - Companion Volume (Thessaloniki, Greece) (WI '19 Companion). Association for Computing Machinery, New York, NY, USA, 51–58.
- [31] B. Renoust, D. Le, and S. Satoh. 2016. Visual Analytics of Political Networks From Face-Tracking of News Video. *IEEE Transactions on Multimedia* 18, 11 (Nov 2016), 2184–2195.
- [32] Rob Savillo. 2016. Diversity On The Sunday Shows In 2015. MediaMatters for America (2016).
- [33] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition. 815–823.
- [34] Alicia Shepard. 2012. The iconic photos of Trayvon Martin & George Zimmerman & why you may not see the others. *Poynter Institute* (Mar 2012).
- [35] Women's Media Center. 2017. The Status of Women in the U.S. Media.
- [36] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. 2016. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters* 23, 10 (2016), 1499–1503.

A REPRODUCIBILITY

This section condenses the most relevant aspects of our validation; please refer to our extended supplemental materials (on our webpage) for additional pseudo-code, tables, and dataset observations.

A.1 Additional methodology and validation

Face detection We assessed the precision and recall of the face detector [36] in a random sample of frames, stratified across 10 years. The results are tabulated in Table 1. Precision is very high, while recall is lower. Our validation metric includes all faces, regardless of size, sharpness, and importance. The per-face recall dips to its lowest in 2016, when footage from Clinton's and Trump's political rallies was common and featured a large number of background faces. For consistency, we do not set thresholds on face size and sharpness, though such options can be useful, depending on the analysis question. Likewise, despite more recent face detectors now being available, we use a single model across all of the data.

Year	Precision	Recall	Error (frame level)
2010	0.973	0.813	10.8%
2011	0.986	0.792	11.2%
2012	0.982	0.759	14.0%
2013	0.992	0.721	15.2%
2014	0.979	0.757	12.8%
2015	0.974	0.803	11.6%
2016	0.981	0.673	14.8%
2017	0.984	0.720	15.2%
2018	0.986	0.712	17.6%
2019	0.985	0.715	15.6%
All	0.985	0.745	13.9%

Table 1: Face detector precision and recall for all faces in 250 randomly sampled frames per year. We also report the percentage of frames containing at least one error.

Gender classification Table 2 shows the confusion matrix from our k-NN model and the class imbalance in the 6,000 randomly sampled validation faces. We use k = 7. Note that just from this validation sample, we can estimate that 31.5% (±1.2% at 95% confidence) of the faces in the dataset are female-presenting.

Identifying public figures We use a combination of our own models (trained on FaceNet [33] embeddings) and labels from AWS Rekognition Celebrity Recognition [1]. Our models are used for individuals missing from or poorly detected by AWS. Fig. 14 plots the precision and recall of the public figures, Trayvon Martin, and George Zimmerman, who are used in our analysis.

As mentioned in section 3, AWS Rekognition Celebrity Recognition [1] suffers from a doppelgänger problem when applied to cable news video, especially for uncommon identities. The effect, shown in Table 3, underscores the importance of visual validation of black-box API results prior to analysis. We manually verified that the individuals (e.g., politicians, news presenters) referenced in the paper do not fall under the "doppelgänger" category. A possible avenue for future work is to leverage weak, multimodal supervision from the captions or on-screen text to identify likely errors and adapt person identification models to the TV news domain.

Human labels \ Predicted labels	Male	Female
Male	4,058	51
Female	118	1,773

Table 2: Confusion matrix of k-NN generated labels and human-annotated gender labels. The estimated precision and recall for the male- and female-presenting classes are 97.2% and 98.8%; and 97.2% and 93.8%, respectively.



Figure 14: Precision is estimated on 100 randomly sampled faces identified as each individual. Recall is estimated on true instances of each individual's face found in a random sample of 2,500 faces, from 25 videos known to contain each individual. See extended supplemental materials for table.

Screen time	# of names	Est. % of doppelgängers	
0-10 min	129,138	-	
10-15 min	8,559	80%	
15-30 min	10,664	76%	
30-60 min	6,352	72%	
1-2 hr	3,403	84%	
2-5 hr	2,136	68%	
5-10 hr	795	52%	
10-20 hr	445	4%	
20-50 hr	415	4%	
50-100 hr	203	0%	
100-200 hr	90	0%	
200 hrs or more	107	0%	

Table 3: The Amazon Rekognition Celebrity Recognition API [1] makes identity predictions for 162,307 distinct names in our dataset. We noticed that the majority of uncommon names (i.e., individuals with less than 10 hours of screen time) predicted by the API are "doppelgängers" of the people who are actually in the news content (false positives). These doppelgängers include a large number of international musicians, sports players, and actors/actresses. To evaluate the effect of these errors, we randomly sampled 25 individuals (by name) from each screen time range and visually validated whether the individual is present only as a doppelgänger to other individuals. A threshold of 10 hours is needed to eliminate nearly all of the doppelgängers.

Portrayal of Trayvon Martin and George Zimmerman We cluster faces by their "source" image (before any editing; see Fig. 15 for examples). Clustering is performed with a human in-the-loop by selecting faces which correspond to unique images and partitioning using FaceNet [33] embeddings. Viewing the clusters can reveal new source or misclassified images, and the human can create new labels, fix existing labels, and repeat the process. We repeat until the clusters are clean (e.g., over 90% precise). We find that using a 1-NN classifier for partitioning is sufficient and that only a small number of manual labels are needed (fewer than 200) to obtain good accuracy in the clusters (Table 4).



(a) Trayvon Martin

(b) George Zimmerman

Figure 15: Examples of the top four images of Trayvon Martin and George Zimmerman. Images can have different backgrounds, color tone, sharpness, and contrast as a result of editing while the source image remains the same. For Zimmerman, who survived, we make a best effort to group faces from the same source event or setting (e.g., court appearances, an interview).

Trayvon Martin			R	
Precision (500 samples) Recall (500 samples)	0.996 1.000	0.978 1.000	0.988 1.000	0.986 0.994
George Zimmerman		P		S
Contains video?	yes	no	yes	no
Precision (500 samples)	0.970	0.996	0.948	0.990
Recall (500 samples)	0.941	1.000	1.000	1.000

Table 4: Estimated precision and recall for the top four clusters for Trayvon Martin and George Zimmerman. For each cluster (X), we estimate precision by sampling randomly in X and counting false positives. To estimate the number of false negatives (for recall) we sample faces randomly from all other clusters and count the number of faces that belong in cluster X but were wrongly assigned. The precision estimate is used to estimate the number of true positives.

Screen time of news presenters With few exceptions, we track the screen time of news presenters at the channel granularity. News presenters can change roles and shows multiple times in the decade, and we use "news presenter" as an umbrella term for staff who are on-air at a network. We also do not explicitly track the dates when news presenters are employed at a network, but because news presenters usually are not newsworthy in themselves, their screen time almost always drops to 0 once they leave a network. The exceptions to the channel granularity are presenters who were already public figures (e.g., politicians) prior to becoming presenters; these individuals include Mike Huckabee, Newt Gingrich, and David Axelrod, who we track at the granularity of their hosted shows.

Hair color for female news presenters Two of the authors independently labeled the visible hair color for each news presenter in 25 frames sampled from the dataset. There were five possible labels (blond(e), brown, black, red, white/gray, and bald). For each news presenter, we calculated the majority label according to each rater. The inter-rater agreement for the majority label for female news presenters was 92.4%. In these cases, the majority label was used in the analysis as the hair color label. The two raters reviewed and agreed upon a hair color label for the 11 female news presenters



Figure 16: Random image of each female-presenting news presenter, grouped by their assigned hair color label.

where their majority labels did not match. Fig. 16 shows example faces from each hair color group for the female news presenters that we analyzed. The data for male presenters was not analyzed due to low inter-rater agreement (only 75%) as a result of ambiguity between grey/white, blond, and bald labels. Our method does not account for presenters' changing their hair color, but this is rare.

Measuring association between words and male- and femalepresenting screen time The majority of the caption text tokens in the dataset (including rare words, but also misspellings) occur very infrequently (95.6% of unique tokens appear fewer than 100 times in the dataset). Because there are few face detections simultaneous to these tokens being said, their association with the presented gender of on-screen faces has very high variance and they are excluded. To focus on words with significant usage, we filter out news presenter names and NLTK English stop words [4], and we restrict our analysis to common tokens (the top 10% of remaining tokens with at least 13,462 utterances each).

We rank the tokens according to the difference in conditional probability of male- and female-presenting faces being on-screen given the word appearing in the captions. The top and bottom tokens in this list have the greatest gender disparity, and we report the top tokens for each presented gender. We manually filter out entries that are human names (e.g., Oprah) or specific to news program names (which associate with their hosts' gender(s)). The top female-associated word, futures is similar to other highlyranked tokens in the list (e.g., NASDAQ and stocks), but is also part of the name of a female-hosted news program (i.e., the 3-gram *Sunday Morning Futures* accounts for 14.6% of mentions). Newsroom, ranked 14th in female screen time association, is also part of several show names.

A.2 Implementation

Our data processing pipeline runs on Linux VM instances and can be parallelized when ingesting large video datasets. We process at least 72 hours of video daily (24 per channel) to keep the public tool up to date; this processing occurs on a VM with 8 vCPUs and 64GB of RAM. The resulting annotations are stored in a relational database and cloud storage. Our public analysis tool makes heavy use of in-memory processing for real-time queries but can comfortably service queries over 290,000+ hours of data with 64GB of RAM. Code for both systems is open-source: https://github.com/scannerresearch/{ tv-news-ingest-pipeline, tv-news-viewer }.