Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube

ESLAM HUSSEIN[*] and PRERNA JUNEJA[*], Department of Computer Science, Virginia Tech
TANUSHREE MITRA, Department of Computer Science, Virginia Tech

Search engines are the primary gateways of information. Yet, they do not take into account the credibility of search results. There is a growing concern that YouTube, the second largest search engine and the most popular video-sharing platform, has been promoting and recommending misinformative content for certain search topics. In this study, we audit YouTube to verify those claims. Our audit experiments investigate whether personalization (based on age, gender, geolocation, or watch history) contributes to amplifying misinformation. After shortlisting five popular topics known to contain misinformative content and compiling associated search queries representing them, we conduct two sets of audits—*Search-* and *Watch-misinformative audits*. Our audits resulted in a dataset of more than 56K videos compiled to link stance (whether promoting misinformation or not) with the personalization attribute audited. Our videos correspond to three major YouTube components: *search results*, *Up-Next*, and *Top 5* recommendations. We find that demographics, such as, gender, age, and geolocation do not have a significant effect on amplifying misinformation in returned search results for users with brand new accounts. On the other hand, once a user develops a watch history, these attributes do affect the extent of misinformation recommended to them. Further analyses reveal a filter bubble effect, both in the *Top 5* and *Up-Next* recommendations for all topics, except *vaccine controversies*; for these topics, watching videos that promote misinformation leads to more misinformative video recommendations. In conclusion, YouTube still has a long way to go to mitigate misinformation on its platform.

CCS Concepts: • **Information systems → Personalization**; *Page and site ranking*; *Content ranking*; • **Human-centered computing → Human computer interaction (HCI)**.

Additional Key Words and Phrases: search engines, misinformation, algorithmic audit, misinformation audit, information retrieval, conspiracy theory, group fairness

ACM Reference Format:
Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 48 (May 2020), 26 pages. https://doi.org/10.1145/3392854

1 INTRODUCTION

Search engines are an indispensable part of our lives. Despite their importance in selecting, ranking, and recommending what information is considered most relevant for us—a key aspect governing our ability to meaningfully participate in public life [24]—there is no guarantee that the information is credible. Numerous scholars have emphasized the need for systematic statistical investigations, or audits of search systems so as to uncover societally problematic behavior [69]. For example,

[*] **Both authors contributed equally to this research.**

Authors' addresses: Eslam Hussein, ehussein@vt.edu; Prerna Juneja, prerna79@vt.edu, Department of Computer Science, Virginia Tech; Tanushree Mitra, Department of Computer Science, Virginia Tech, tmitra@vt.edu.

2573-0142/2020/5-ART48 $15.00
https://doi.org/10.1145/3392854

multiple studies have audited search engines for the presence of partisan bias [33, 63] and gender bias [10, 17]. Yet, none have empirically audited them for misinformation. Moreover, investigation of video search engines, like YouTube is rare (work by Jiang et al. is one exception [34]), despite popular prediction that by 2022, 82% of internet traffic will come from videos [14]. YouTube has also faced years of criticism for surfacing misinformative content [8, 18, 83]. Critics have gone as far as calling YouTube a *conspiracy ecosystem* [1]. Despite such vehement criticisms, there has been little effort towards quantifying the extent of misinformation in video search platforms, or investigating user attributes that might have an effect. What is the effect of attributes, such as user's demographics and geolocation on the amount of misinformation returned and recommended on YouTube? How does it change with user's watch history, where watch history is progressively built by watching videos rife with inaccuracies or videos presenting extensive debunks. This paper grapples with these questions and sheds light on the phenomenon of algorithmically surfaced misinformation on YouTube and how that is affected by penalization attributes (gender, age, geolocation, and watch history). We study the conspiracy facet of misinformation and perform our audits on trending and perennial misinformative topics that are widely known to be false (details in Section 3). In particular, we examine five misinformative topics namely, *9/11 conspiracy theories*, *chemtrail conspiracy theory*, *flat earth*, *moon landing conspiracy theories* and *vaccine controversies*.

We conduct two sets of audit experiments—*Search* and *Watch* audits to examine YouTube's search and recommendation algorithms, respectively. While *Search* audits are conducted using brand new user accounts, *Watch* audits examine user accounts that have built watch history by systematically watching either all promoting, neutral, or debunking videos of potentially misinformative topics. Both audits control for extraneous factors that can lead to potential errors in our audit data collection. We create more than 150 Google accounts to audit YouTube. Our experiments collect 56,475 YouTube videos, spread across five popular misinformative topics and correspond to three major components of YouTube: videos present in *search results*, *Up-Next*, and *Top 5* recommendations.

We find little evidence to support that users' age, gender and geolocation play any significant role in amplifying misinformation in search results or recommended videos for brand new accounts. On the other hand, watch history exerts a significant effect on the amount of misinformation present in the *search results* corresponding to the *vaccine controversy* topic. Watch history also significantly affects the extent of misinformation in recommended videos (both *Up-Next* and *Top 5*) for all five misinformative topics. Interestingly, we observe a filter bubble effect in recommendations, where watching promoting misinformative videos lead to more promoting videos in the *Up-Next* and *Top 5* video recommendations. This filter bubble effect for recommended content is observed for all topics, except *vaccines controversies*. For the vaccine topic, while filter bubble is not observed for the *recommended* videos, it exists for the *search results*. Specifically, people who watch anti-vaccination videos are presented with less misinformation in their recommendations but more misinformation in their search results, compared to those who watch neutral or debunking vaccine videos.

Overall, our work makes the following contributions:

- We develop a methodology to audit search engines for misinformation and open up a new avenue in the domain of algorithmic-audit research. By applying our methodology on YouTube, our study is the first to systematically investigate the effect of personalization attributes on the extent of misinformation returned via searches and recommendations on a video search platform.
- Overall, our audits result in a novel dataset of 56,475 videos compiled to link the stance of the video with the personalization attribute audited[1]. A byproduct of our audit is an extensive qualitative coding scheme built to measure the stance of video content. Using this scheme,

[1] https://social-comp.github.io/YouTubeAudit-data/

we manually coded each of the unique videos (a total of 2,943) with a discrete misinformation stance (promoting, neutral, or debunking).

- Our audit study revealed variability in YouTube's behaviour towards different components across different misinformative topics. We find evidence to suggest that a filter bubble effect exists in the *recommendations* of all misinformative topics except *vaccine controversies*. We find that YouTube recommends debunking videos in its *Top 5* and *Up-Next* components to the accounts that watch promoting anti-vaccination videos. However, a filter bubble effect still exists in the *search results* of the aforementioned topic—people that watch videos promoting anti-vaccination agenda are presented with more such videos in their *search results.*

2 RESEARCH QUESTIONS AND HYPOTHESES

Our work is guided by the following main research question: What is the effect of personalization (based on age, gender, geolocation, or watch history) on the amount of misinformation presented to users on YouTube? We formulate the following sub-questions and hypotheses to investigate the effects of each of these personalization attributes.

RQ1 [*Search & Watch Experiments*]: What is the effect of demographics (age, gender) and geolocation on the amount of misinformation returned in various YouTube components?

RQ1a [*Search Experiments*]: How are *search results* affected for brand new accounts?

RQ1b [*Watch Experiments*]: How are *search results*, *Up-Next*, and *Top 5* recommendations affected, given accounts have a watch history?

Users provide their demographic information, including age and gender while signing-up for a new Google account. They use the same Google account for accessing YouTube. Prior studies investigating associations between user demographics and engagement with misinformation have found that the likelihoods for sharing misinformation vary across user groups [25]. For example, adults aged 65 or older were seven times more likely to share articles from fake news domains compared to younger age group users. Another study indicated that women have a higher likelihood of sharing misinformation [13]. Different demographics having different likelihoods of sharing misinformation might imply that certain groups are exposed to more misinformative content than others. Thus, given the interplay between demographic differences and engagement with misinformation, we hypothesize that YouTube's algorithm could indeed be biased, exposing older people and females to more misinformation while presenting content related to our five misinformative topics.

H1a. Older people (50 years or older) will be presented with more misinformative content than younger age groups.

H1b. Females will be presented with more misinformative content than males.

Prior studies have also shown that search algorithms, specifically Google search, leverage user's geolocation information to present personalized search results [28]. Moreover, Google keeps track of the region-based popularity of search topics and search queries through Google Trends data [31]. Hence, we hypothesize that geolocation will exert an effect, which in turn will depend on how popular the misinformative search topic is in that region.

H1c. Regions where misinformative topics are popular (hot regions) will be presented with more misinformative content compared to regions where such topics are rarely searched (cold regions).

While RQ1 investigates the effect of attributes that are directly connected to a user's account, RQ2 delves into the second order effect of a user's accumulated watch history. Hence, in RQ2, we ask:

RQ2 [*Watch Experiments*]: What is the effect of watch history on the stance of misinformative content returned in various YouTube components?

Technology critics have raised concerns on search engines' tendency to create a filter bubble over time by presenting less diverse and more attitude confirming search results and recommendations [57, 73]. Some media reports have gone so far as to claim that YouTube recommendations drive users down the conspiracy rabbit-hole by recommending increasingly more pro-conspiracy theory videos [65]. Hence, we hypothesize:

H2. Watching more videos belonging to a particular misinformative stance (promoting, neutral or debunking) leads YouTube's search and recommendation algorithm to present more videos reflecting that particular stance to users.

RQ3 [*Search & Watch Experiments*]: How does the amount of misinformative content differ across misinformative topics?

> RQ3a [*Search Experiments*]: How does misinformative content present in *search results* of brand new accounts differ across topics?
>
> RQ3b [*Watch Experiments*]: How does misinformative content present in *search results*, *Up-Next*, and *Top 5* recommendations of accounts having a watch history differ across topics?

Some misinformative topics are more popular than others. For example, topics like *vaccine controversies* have been widely discussed in the popular media. In the last few years, several social media platforms received backlash for harboring anti-vaccination content [23, 44]. In the beginning of 2019, a handful of them, including YouTube, pledged to take measures against vaccine misinformation [70, 76]. Does that indicate that YouTube's algorithm will present less misinformative content for such topics, given we performed our audit experiments in the middle of 2019? We hypothesize that when attention received by misinformative topics vary, the amount of misinformative content presented by topics will also vary.

H3. The amount of misinformative content returned will differ across misinformative topics.

3 STUDY CONTEXT: MISINFORMATION

The research community has referenced online misinformation with different names and definitions. A few popular characterizations include "fake news" [26, 32], "hoaxes" [39], "rumors" [22, 61], "conspiracy theories" [6, 66], "information credibility" [9, 48] and "perceived accuracy" [7, 58]. In our study, we focus on the conspiratorial aspect of misinformation and use these terms interchangeably. Conspiracy theories are narratives that embody the belief that secret and influential organizations are behind the occurrence of a particular event [93]. Note that conspiracy theories are not always false. There have been several cases in the past where conspiracy theories turned out to be true (for example, Watergate Scandal [71] and Project MKUltra [87]). To differentiate true conspiracy theories from false, we depend on the theory of social constructionism where a fact is only considered "true" if its claim is widely cited, replicated, and accepted without contest [41]. For the purpose of this research, we focus on demonstrably false misinformative topics that have been persistently discussed (e.g. anti-vaccine conspiracies) and for which the mainstream view of reality is known—for e.g., "vaccines do not cause autism". The mainstream perspective of such theories is either backed by expert authorities or scientific research and is widely accepted by a large number of people. At present, "what majority of people believe in" is our best effort in determining the truthfulness of conspiracy theories. The truth value may change in the future if new information is available. In this research, we focus on five topics namely, *9/11 conspiracy theories*, *chemtrail conspiracy theory*, *flat earth*, *moon landing conspiracy theories* and *vaccine controversies*. In the rest of the paper, we refer to these as misinformative topics. All these topics are demonstrably false, perennial and denied by authoritative sources or backed by scientific research. We next describe each topic and demonstrate how these are demonstrably false and perennial.

3.1 Five Misinformative Topics: Demonstrably False and Perennial

3.1.1 9/11 misinformative topic. There are several conspiracy theories surrounding the *9/11* attacks [50]. Some of them claim that authorities had foreknowledge of the attacks and that they deliberately aided the attackers. Few attribute the collapse of the Twin Towers to a controlled demolition or explosives [50]. Possible motives for these theories involve justification of the Iraq and Afghanistan invasion by the U.S. Government. Other theories assert that attacks were financed by Saudi Arabia's Royal family or were orchestrated by the Israel Government or Pentagon was hit by a missile launched under the orders of the U.S. Government [37]. All these accounts have been denied by authoritative sources and expert analysts [74]; hence the theory is demonstrably false. Yet, a New York Times poll conducted on 1,042 individuals revealed that 16% US adults do not believe in government's account of 9/11 attacks and 56% believe that the government is hiding something from them [78]. These statistics reveal that the theory is still persistent, despite being false.

3.1.2 Chemtrails misinformative topic. Chemtrails conspiracy theories claim that long lasting condensation trails, also known as Contrails, left by air-crafts and rockets in the sky are composed of harmful chemicals. The theories blame United States Air Force (USAF) for spraying these harmful chemicals with the intention of altering the weather, controlling the population and causing diseases. National Oceanic and Atmospheric Administration (NOAA) has constantly denied such allegations, citing research that has debunked these false claims [52]. Despite the scientific evidence, a recent study done with 1000 subjects found that 10% and 30% of Americans believe chemtrails conspiracy to be "completely" and "somewhat true", respectively [79].

3.1.3 Flat earth misinformative topic. Our third topic relates to *flat earth* conspiracies. *Flat earth* conspiracy theorists claim that National Aeronautics and Space Administration (NASA) and government agencies are duping the public into believing that Earth is spherical in shape. Surprisingly, a 2018 survey revealed that only 66% of millennials believed that the Earth is spherical [91].

3.1.4 Moon landing misinformative topic. Moon landing conspiracies claim that NASA's Apollo Mission's moon landing was staged by the agency. The theory was denied by NASA [49]. A 500 person poll revealed that 1 in 10 Americans still believe that moon landing never happened [4], justifying our perennial criteria for topic selection.

3.1.5 Vaccine misinformative topic. Conspiracy theories related to vaccines are based on the mistaken belief that vaccines contain harmful ingredients that can cause diseases like autism and sudden infant death syndrome (SIDS). Some theories also claim that childhood diseases can be automatically cured by the human body's immune system and thus, vaccination is not required. Such claims are denied by the World Health Organization among other authoritative sources and several scientific research [54, 55]. Yet, a recent survey conducted with 2000 participants revealed that 45% of American adults doubt vaccines [75]. We discuss how we empirically selected these five misinformative topics in detail in Section 5.

4 RELATED WORK

4.1 Misinformation in Search Systems

The literature on the phenomenon of misinformation and credibility of online content is extensive (see [38] for a survey). While most studies focus on social media based misinformation [2, 3, 62, 66, 67, 72, 94], systematic research on scrutinizing search engines for inaccurate content is rare despite their growing significance in our lives. It has been reported that 92% of the adult population rely on search engines for information [60], including information related to serious medical conditions and disability [16]. Dependence on these search systems have made us susceptible to their impact in critical ways [21]. For example, studies have shown that manipulating the rankings of search engine results may influence the votes of undecided citizens [19]. In another instance, researchers found that a significant number of people ended up believing that the Earth is flat after watching recommended videos in Youtube—one of the most popular video search platforms [53]. Another report outlined that searching for "vitamin K shot" on Google and YouTube returned web pages and videos asking parents to skip the vitamin shot [18]. The top search results also promoted anti vaccine conspiracies [18]. YouTube has been repeatedly accused of featuring conspiracy videos in its trending and recommendation sections in response to searches related to the 9/11 event, the Las Vegas shooting, and vaccination [83]. In a separate study analyzing election bias in YouTube, video analyses showed that most videos favored Trump and a substantial amount of them contained fake news and conspiratorial stories about Clinton [42]. While this study examined only the *Up-Next* component of videos, we measure how personalization (age, gender, geolocation, watch history) affects the amount of misinformation present across multiple components on YouTube: search results, *Up-Next*, and the YouTube recommendations. Moreover, all the aforementioned studies provide anecdotal evidence of search misinformation without systematic audit investigation. Our study fills this gap by auditing YouTube, a popular video search platform, to determine if personalization affects the amount of misinformation returned in search results and recommendations.

4.2 Search Engine Audits

In recent times, search engines have been critiqued for promoting misinformative and biased results [82]. One of the key methodologies used to identify, study, and quantify such bias, discrimination and misinformation is the *audit methodology*. An audit comprises of systematic statistical probing of an online platform to uncover societally problematic behavior underlying its algorithms [69]. Using audit techniques, researchers have investigated several issues pertinent to algorithmically driven online platforms. For example, they have explored the presence of partisan bias in search engine components [33, 46, 63]; investigated representativeness issues, such as racial and gender bias in online freelance marketplaces [30] and resume search engines [10]; presence of price discrimination and algorithmic manipulation in e-commerce websites [12, 29]; opacity in price surging algorithms used by ride sharing services [11]; lack of news source diversity in information returned by search platforms, [80]; and the extent of personalization and localization used by search engines [28, 35]. Yet, auditing online platforms for algorithmic misinformation is practically non-existing. By focusing on auditing YouTube for misinformation, our study takes a first step in the direction of auditing algorithms for misinformation.

All the aforementioned audit studies have also exposed numerous methodological challenges faced during audit investigations. The first roadblock is determining a viable set of search queries that will result in meaningful measurements. Surely, we cannot feed all possible search queries to the system under audit. Researchers have adopted several techniques to compile and shortlist meaningful search queries. For example, to audit Google's Top stories box, researchers selected Trending topics from Google Trends at a fixed time every day and then manually shortlisted the trending queries related to those topics [80]. An audit conducted on Google, Yahoo and Bing search engines, during the 2016 United States Congressional elections, used the names of electoral candidates as queries [47]. To investigate gender bias in resume database, researchers used the most commonly searched job titles [10]. To audit for partisan bias in Google search, scholars compiled autocomplete suggestions for multiple root queries related to Donald Trump's presidential inauguration [63]. We leverage both, queries from Google Trends as well as YouTube's autocomplete suggestions to ensure that the query set is trending and relevant to the platform. We then narrowed down this query set by manually removing semantically similar queries.

The second challenge of audit methodologies relate to carefully controlling the experimental setup for meaningful audit investigations. These comprise decisions on setting the data collection framework, selecting the components to audit, and controlling for confounding factors or noise. In one study researchers launched surveys to short-list real-life audience who were then instructed to install browser extensions that were used for data collection [63]. We, on the other hand, manually crafted the Google accounts and used automated scripts to collect data so as to have more control over our experiments—a technique that has been widely used by multiple prior audit studies [28, 29]. What components should we select for our audit experiments? Some audit studies focus on one component of the search engine, such as Google's Top stories box [80] or Google's search results [28]). Others focus on multiple components combined, such as various Google search page components including people-ask, news-card, twitter, people-search etc. [64]). Our study focuses on three YouTube components namely, *search results*, *Up-next* video and *Top 5* components. We also leverage previous literature [28] to control for any confounding factors that could possibly affect the outcome of our experiments.

The third challenge for conducting search engine audits lies in identifying the attributes and actions that could possibly affect the feature one is auditing (we focus on personalization). Several audit studies have focused on geolocation based personalization. For example, to investigate the effects of geolocation on web-based personalization, researchers focused on nation-level (randomly selected states in USA), state-level (counties within Ohio) and county-level (voting districts in Cuya-hoga County) locations. They found that personalization in search results increases with physical distance [36]. Instead of focusing on a single state or randomly selecting a handful of states, we determine "hot" and "cold" regions—states where search queries related to a misinformative topic are the most and least searched respectively. This selection provides us a unique opportunity to determine whether web traffic from a region affects YouTube's algorithm. Moreover, prior audit studies have also investigated the effects of demographics, search-history, click history and browsing history on Google's web search results as well as prices of commodities on e-commerce platforms [28, 29]. Motivated by these studies, we investigate the effects of attributes like demographics, geolocation and watch history on the amount of misinformative content present in various YouTube components.

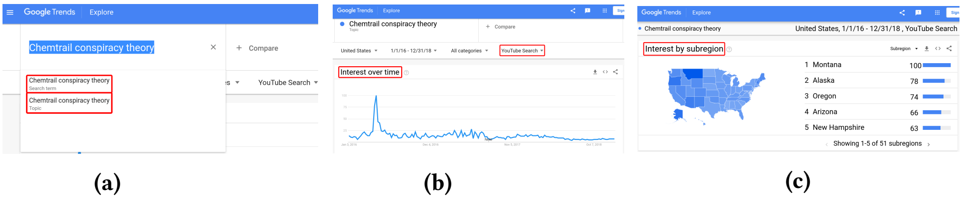
The last challenge for conducting online audits relates to properly defining how one is measuring the output label of the phenomenon that is being audited. For example, if a study investigates partisan bias, how do you define and label bias in a valid way? For our misinformative search audits, we label YouTube videos as promoting, debunking or neutral based on whether their narrative supports, debunks, or presents general information on any conspiratorial view respectively (more details in Section 5.3)

5 METHODOLOGY

Here, we first present our methodology for compiling high impact misinformative queries, the design and implementation of our audit experiments, steps for collecting audit data, including components of YouTube's Search Engine Results Page (SERP) and video pages, and our qualitative coding scheme for determining stance of the returned videos.

5.1 Compiling High Impact Topics and Queries

Our selection methodology to identify relevant and impactful *misinformation search topics* and *queries* comprises of three key steps.



Chemtrail conspiracy theory

Chemtrail conspiracy theory — Search term

Chemtrail conspiracy theory — Topic

United States · 1/1/16 - 12/31/18 · All categories · YouTube Search

Interest over time

United States, 1/1/16 - 12/31/18 , YouTube Search

Interest by subregion

1	Montana	100
2	Alaska	78
3	Oregon	74
4	Arizona	66
5	New Hampshire	63

Showing 1-5 of 51 subregions

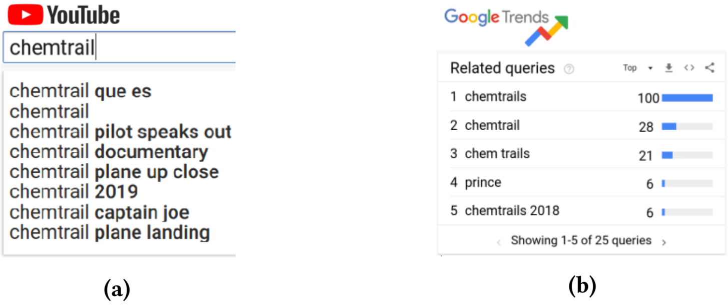
Fig. 1. (a) Google Trends allows users to specify search query as either a topic search or a term search. (b) Interest over time graph. (c) Popularity of *chemtrail conspiracy theory* topic in YouTube searches in the United States between January 1st, 2016 and December 31st, 2018. Color intensity in the heatmap is proportional to the topic's popularity in that region.

Search Topic	Seed Query	Hot	Cold	Sample Search Query
9/11 conspiracy theories	9/11 and 9/11 conspiracy	Maryland	Ohio	9/11 inside job 9/11 tribute 9/11 conspiracy
Chemtrail conspiracy theory	chemtrail	Montana	New Jersey	chemtrail chemtrail flu chemtrail pilot
Flat Earth	flat earth	Montana	New Jersey	flat earth proof is the earth flat
Moon landing conspiracy theories	moon landing	Ohio	Georgia	moon moon hoax moon landing china
Vaccine controversies	vaccines	Montana	South Carolina	anti vaccine vaccines vaccines revealed

Table 1. Seed query, hot & cold regions, and sample search queries for the five misinformation search topics.

5.1.1 Selecting misinformative topics via Wikipedia and related research: We curate a list of relevant misinformative topics (see Table 1) by referring to Wikipedia pages on conspiracy theories [84, 85] (e.g., 9/11, chemtrails, sandy hook, pizzagate conspiracy, etc.). We also refer to past studies that examine misinformation and conspiratorial phenomena in online communities [67, 88]. From this list, we exclude topics whose "truth" value is uncertain, that is, topics for which we were either unable to determine the mainstream perspective or the mainstream perspective is not backed by authoritative voices or scientific research. We manually identify and eliminate such topics. For example, we removed "Malaysian Airlines Flight MH370" topic since official investigations about the flight's disappearance have presented inconclusive reports [5, 40, 86]. Next, we leverage Google Trends to identify the most popular topics—continuously trending, high interest topics—that are searched on YouTube by a large number of people.

5.1.2 Selecting high impact misinformation search topics via Google Trends: Google Trends (Trends for short) is a good indicator for real-world activities impacting a large number of people [20]. Trends also provides interest data across different Google search services including YouTube. Figure 1a demonstrates how Trends could be used to search either as a *Term* or as a *Topic*. For example, searching as a *Topic*, *chemtrail conspiracy theory* will give results for several queries related to the topic (*chemtrails, contrails*—a common word used to refer to chemtrails*)*, whereas searching as a *Term* will return results that contain text strings "chemtrail," "conspiracy," and "theory." We opted to search as a *Topic* and selected "YouTube search" as our preferred service (refer to Figure 1b). This step discarded a few topics for which no trends data was returned. Next, we compare the *interest over time* plots for all remaining search topics from January 1, 2016 to December 31, 2018



chemtrail

chemtrail que es
chemtrail
chemtrail pilot speaks out
chemtrail documentary
chemtrail plane up close
chemtrail 2019
chemtrail captain joe
chemtrail plane landing

(a)

Related queries · Top

1	chemtrails	100
2	chemtrail	28
3	chem trails	21
4	prince	6
5	chemtrails 2018	6

Showing 1-5 of 25 queries

(b)

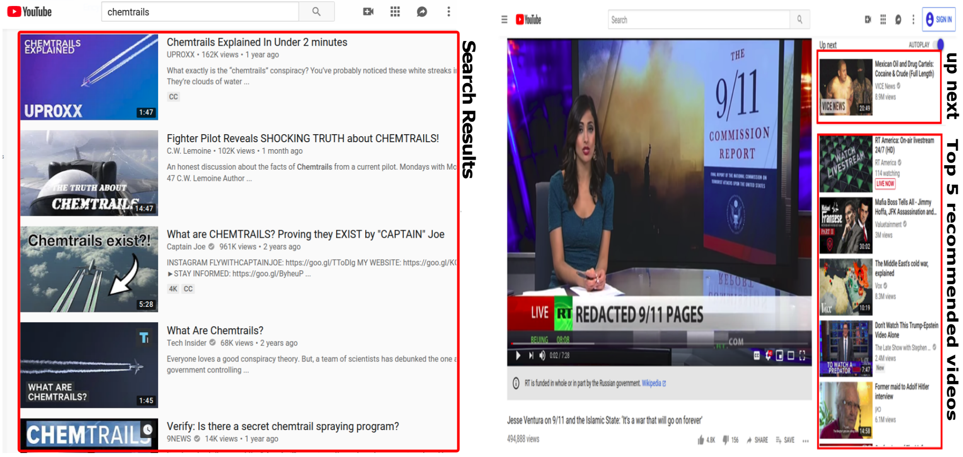
Fig. 2. (a) YouTube search's auto-complete suggests 10 trending queries. (b) Google Trends displays the top search queries related to the term or topic entered in the search box.

to ensure that the topics have been persistently discussed in the last two years. Then, we select the top 5 topics which represent the most searched topics, resulting in our list of highly impactful misinformative topics. Table 1 provides a list.

5.1.3 Selecting Search Queries. Our next step is to generate a set of queries for each of the misinformation search topics which we can use in our subsequent audit experiments and SERP data collection. We need to ensure that our query set comprises of both relevant and high impact or popular queries. We feed seed queries per search topic in both YouTube and Trends. Since our study audits YouTube, query suggestions on YouTube represent the most trending queries searched on the platform, whereas Trends helps identify the most prevalent and impactful queries. YouTube's search box's auto-complete feature suggests 10 popular queries once a seed query is fed into the search box (refer to Figure 2a). We add those to expand our query set. Searching on Google Trends as a *Topic* displays top related queries; number can vary by topic. We also include those in our query set (refer to Figure 2b). Thus, our query set comprises queries suggested by both YouTube and Trends platforms. Next, we manually removed duplicates and replaced semantically similar queries with a single relevant query. We retain the most impactful (trending and most searched) queries by keeping the seed query as well as queries that appear both in the top 5 YouTube suggestions and top 5 related queries list in Trends. We find that shorter queries (length ≤ 4) were better representative of the misinformative topic. Queries comprising more than 4 keywords (for e.g., "the flat earther's $100,000 challenge" and "moon landing press conference analysis") were overly specific. Hence, we only retain more representative generic queries that had a maximum of 4 keywords. Our final query set for the *9/11 conspiracy theories* and *vaccine controversies* topics had 11 queries each. Query sets for *chemtrails*, *flat earth* and *moon landing conspiracy theories* topics had 10, 8 and 9 queries, respectively. In total, we had 49 queries. Table 1 presents a sample.

5.2 Overview of Audit Experiments

YouTube utilizes age, gender, geolocation, and watch history as features in its recommendation system [15]. To determine if these features amplify the amount of conspiratorial content returned to users, we conduct a series of four audit experiments. Our audits collect three primary YouTube components. We annotate the collected videos with stance values: promoting, debunking, or neutral stance towards the topic. Finally, we conduct statistical comparison tests on the annotated data. Our audit experiments also control for multiple sources of noise. Unfortunately, in search engine audit studies, difference in search results and recommendations cannot be solely attributed to personalization. Confounding factors (or noise), if not controlled, can also influence the results. For example, users' choice of web browser could impact Google's search results and recommendations,



Chemtrails Explained In Under 2 minutes
UPROXX • 162K views • 1 year ago
What exactly is the "chemtrails" conspiracy? You've probably noticed these white streaks i
They're clouds of water ...
CC

Fighter Pilot Reveals SHOCKING TRUTH about CHEMTRAILS!
C.W. Lemoine • 102K views • 1 month ago
An honest discussion about the facts of **Chemtrails** from a current pilot. Mondays with Mo
47 C.W. Lemoine Author ...

What are CHEMTRAILS? Proving they EXIST by "CAPTAIN" Joe
Captain Joe 961K views • 2 years ago
INSTAGRAM FLYWITHCAPTAINJOE: https://goo.gl/TToDlg MY WEBSITE: https://goo.gl/K
►STAY INFORMED: https://goo.gl/ByheuP ...
4K CC

What Are Chemtrails?
Tech Insider 68K views • 2 years ago
Everyone loves a good conspiracy theory. But, a team of scientists has debunked the one a
government controlling ...

Verify: Is there a secret chemtrail spraying program?
9NEWS 14K views • 1 year ago

RT is funded in whole or in part by the Russian government. Wikipedia

Jesse Ventura on 9/11 and the Islamic State: 'It's a war that will go on forever'
494,888 views

Mexican Oil and Drug Cartels: Cocaine & Crude (Full Length)
VICE News
8.9M views

RT America: On-air livestream 24/7 (HD)
RT America
114 watching
LIVE NOW

Mafia Boss Tells All - Jimmy Hoffa, JFK Assassination and...
Valuetainment
3M views

The Middle East's cold war, explained
Vox
8.3M views

Don't Watch This Trump-Epstein Video Alone
The Late Show with Stephen ...
2.4M views
New

Former maid to Adolf Hitler interview
JAO
6.1M views

(a) SERP

(b) Video page

Fig. 3. Three components collected from YouTube: (a) *search results* from a SERP and (b) *Up-Next* and *Top 5* recommended videos from a video page

and hence could lead to noisy inferences. Thus, following prior search engine audit work [28], we control for browser noise by selecting one single version of Firefox browser for all experiments. Firefox was selected over Google Chrome to avoid the possibility of Chrome browser tracking Google accounts used in our experiments. All interactions with YouTube happened in incognito mode to remove any noise resulting from tracked cookies or browsing history. We also control for temporal effects by performing simultaneous searches. Additionally, all machines used in our experiments had the same architecture, configuration, and version of the operating system (64bits, Ubuntu 14.04, 3.75GB Ram). This step ensures that there are no temporal effects due to the differences in machines' speeds. In the remaining section, we describe the collected YouTube components and layout our experimental setup.

5.2.1 YouTube Components. We collect the following components: (a) *search results*. These consist of top 20 videos in YouTube's SERP (Search Engine Results Page) returned in response to a search query. (b) *Up-Next* corresponds to the next recommended video that will be played immediately after the current video finishes, (c) *Top 5* relates to the top five recommended videos on the right of the video page. Figure 3 demonstrates the three components.

5.2.2 Search Experiments: Auditing with brand new accounts. For our *Search* experiments, we conduct two experiments to test whether demographics (age and gender) and geolocation for a new user (with no prior history on YouTube) have a significant effect on the proportion of misinformative content returned by the platform.

Experiment #	Category	Feature	Tested Values
Search (Exp 1)	Demographics	Age	<18, 18-34, 35-50, >50
		Gender	Male, Female
Search (Exp 2)	Geolocation	IP Address	Georgia, Montana, New Jersey, Ohio, South Carolina
Watch (Exp 3)	Demographics	Age	<18, 18-34, 35-50, >50
		Gender	Male, Female
	Watch history	Watch history	Promoting, Neutral, Debunking
Watch (Exp 4)	Geolocation	IP Address	Georgia, Montana, New Jersey, Ohio, South Carolina
	Watch history	Watch history	Promoting, Neutral, Debunking

Table 2. List of user features for our audit experiments.

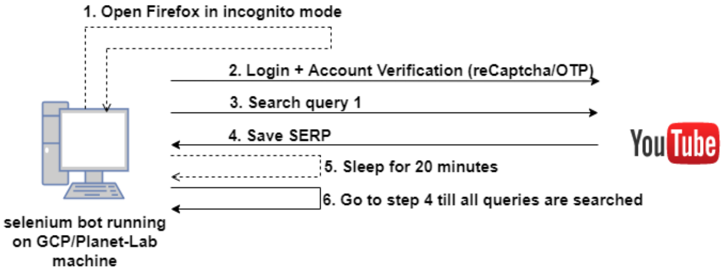

1. Open Firefox in incognito mode
2. Login + Account Verification (reCaptcha/OTP)
3. Search query 1
4. Save SERP
5. Sleep for 20 minutes
6. Go to step 4 till all queries are searched
selenium bot running on GCP/Planet-Lab machine
YouTube

Fig. 4. Steps performed in *Search* experiments 1 and 2.

Experiment 1: Search & Demographics (age and gender). We consider four age groups (less than 18 years old, 18 – 34, 35-50, and greater than 50) and two gender values (male and female) (see Table 2). We create eight different Google accounts—2 (gender values) X 4 (age group values)—each having a unique combination of gender and age. We manually crafted these accounts by following Google's account setup process of adding profile details (age and gender), and including a recovery email and phone verification.

Implementation: Each account is managed by a selenium bot. The bot runs on a virtual machine created on Google Cloud Platform (GCP). When testing for demographics, searches across all accounts are performed from the same location (Mountain View, California) to control for the effect of geolocation. Figure 4 shows the experimental setup. Each bot controlling an account opens Firefox browser in incognito mode and logs in to YouTube using that account's credentials. Each bot conducts searches on YouTube's homepage by drawing queries from the query sets of all misinformative topics. The searches are done in sequence similar to Vincent et al's approach in [81]. The bot sleeps for 20 minutes after every search to neutralize the carry-over effect—noise introduced in search results from dependency present in consecutive searches. Prior audit experiments on Google Web Search showed that carry-over effect is observed if the interval between two sequential query execution is less than 11 minutes [28]. We use this value as the benchmark and decide to keep a time interval of 20 minutes between two YouTube searches to control for carry-over effects. We collect SERP data for each of the 49 search queries, scrape these html-based SERPs to extract URLs of the top 20 videos present in the search results.

Experiment 2: Search & Geolocation To study the effect of geolocation, we need to identify physical locations corresponding to each search topic from where automated YouTube searches will be performed. We make use of Google Trend's *interest by sub-region* feature to shortlist locations that have the highest (or lowest) interest corresponding to each topic under audit investigation. We searched Trends 50 times for each of the misinformative search topics with the same parameters (region="US," time="1/1/2016 to 12/31/2018," service="YouTube search"). We calculate the average *interest-by-region* value for each sub-region (i.e. state), shortlist 15 sub-regions with the highest interest scores (referred to as hot regions in the paper) and bottom 15 regions with lowest scores (cold regions). Intuitively hot and cold regions are states in the U.S. where the search topic is the most and least popular, respectively. We select one hot and one cold sub-region for each search topic based on its availability on the list of active working nodes in geographically dispersed machines, called Planet-Lab [59]. For example, for *flat earth* topic, among the 15 hottest sub-regions (e.g. North Dakota, Montana, Oregon, etc.) we selected Montana because of its availability among Planet-Lab active working nodes. Table 1 shows the selected hot and cold sub-regions across all topics.

Implementation: For each search topic, we run two selenium bots, each corresponding to either a hot or cold geolocation. The bots run on the virtual machines created on the GCP. These bots

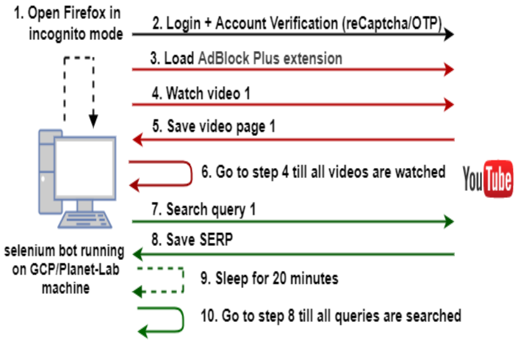

1. Open Firefox in incognito mode
2. Login + Account Verification (reCaptcha/OTP)
3. Load AdBlock Plus extension
4. Watch video 1
5. Save video page 1
6. Go to step 4 till all videos are watched
YouTube
7. Search query 1
8. Save SERP
9. Sleep for 20 minutes
10. Go to step 8 till all queries are searched
selenium bot running on GCP/Planet-Lab machine

Fig. 5. Steps performed in *Watch* experiments 3 & 4. These experiments have two phases: (1) watch phase (denoted by →), (2) search phase (denoted by →).

***Watch* experiments, for each misinformative topic:**

Stance	No. of accounts (Demographics)	No. of accounts (Geolocation)	
		Hot	Cold
Debunking (-1)	8	1	1
Neutral (0)	8	1	1
Promoting (1)	8	1	1
Total accounts	24	6	

Table 3. Accounts created to execute *Watch* experiments for each misinformative topic. In total, we created 120 (24X5) accounts to run experiment 3 and 30 (6X5) accounts for experiment 4. Here 5 denotes the number of topics.

connect to the Planet-Lab machines deployed in the hot and cold regions (refer to Table 1) for that misinformative topic through ssh tunneling. Figure 4 presents the steps performed in this experiment. After searching every query, every bot saves the SERP. Later, we scrape all the saved SERPs and extract the URLs of the top 20 videos present in them (i.e. *search results*). After completion of both search experiments (demographics and geolocation), we collected a set of 848 unique videos.

5.2.3 Watch Experiments. The goal of our *Watch* experiments is to examine the effect that a user's watch history exerts on the amount of misinformation presented to the user in both YouTube's search and video pages. We also determine how that effect varies with user demographics and geolocation. The experimental setup comprises of two phases, 1) *watch* and 2) *search*. The watch phase builds the watch history of every Google account followed by the search phase that conducts searches on YouTube. During the watch phase, after watching every video, we extract the *Up-Next* video and the *Top 5* recommendation components.

Experiment 3: Watch & Demographics. The aim of this experiment is to test the effects in the presence of a user's watch history. Hence, we first need to build the history of new user accounts by automatically making them watch videos that are either all debunking, neutral or promoting the particular misinformative topic under audit investigation. We create three sets of 2 (gender values) X 4 (age group values) Google accounts to audit each misinformative topic where each set watches 20 videos from each of the three stances. We obtain the videos from our previous set of *Search* experiments. We select 20 most popular videos for each of the misinformative topics. Popularity is calculated as the engagement accumulated by the video at the time of our experimental runs;

$$\begin{aligned} Popularity\ metric\ (pm) = \ & view\ count + like\ count + dislike\ count + \\ & favorite\ count + comment\ count \end{aligned}$$

We have released all videos corresponding to each stance (promoting, neutral, debunking) that were used to create watch histories of Google accounts along with their popularity values as the part of the online dataset[2].

Two authors annotated the video collection with stance values: -1 (debunking), 0 (neutral) and 1 (promoting). We describe our qualitative coding scheme and process in Section 5.3. Table 3 shows the count of accounts created for each misinformative topic for this experiment.

[2] https://social-comp.github.io/YouTubeAudit-data/

Implementation: Our *Watch* experiment for studying the effects of demographics is similar to our *Search* experiment runs. The only difference being that accounts build their watch history by watching, in its entirety, 20 popular videos from a particular stance set (all having the same stance in a set, either -1, 0, or 1) before conducting any search operation on YouTube. Figure 5 presents the steps for the *Watch* experiment.

Experiment 4: Watch & Geolocation. The aim of this experiment is to test the effect of the hot and cold geolocations on the amount of misinformation presented to the users in YouTube, given that each user has a watch history. Similar to the previous *Watch* experiment, the history is created by making each account watch YouTube videos of a particular stance. We create three sets of two Google accounts (see Table 3), each corresponding to a hot or cold region (refer to Table 1). The three sets build their watch histories following the same steps as in experiment 3.

Implementation: For each search topic, we run six selenium bots, three for hot and three for cold geolocations. After building their watch histories, the bot runs in a similar fashion as experiment 2—*Search & Geolocation*. After completion of the experimental runs, we collected 2,479 unique videos from both *Watch* experiments—demographics and geolocation. One author annotated one half of these videos, while the other half was annotated by the second author using the process described in Section 5.3.

5.3 Annotating our Data Collection

Through our audit experiments, we collected a total of 56,475 videos with 2,943 unique videos. We used an iteratively developed qualitative coding scheme to label our video collection. Qualitative coding is a process of interpreting data and labelling it into meaningful categories. First, the authors randomly selected 25 videos from the *Search* experiments' data collection, 5 from each topic. Next, six human annotators independently annotated all videos using a basic 3-scale annotation scheme: -1 (debunking), 0 (neutral), and 1 (promoting). All six annotators, including the authors, then discussed their individual annotations and the heuristics followed for the task. After discussions and multiple rounds of iterations, all raters reached a consensus on the annotation heuristics. The process resulted in a scale comprising 9 different annotation values: −1 to 7. This 9-point scale gives a microscopic view of the kinds of videos a user is exposed to when she searches for a misinformative topic (details in the next section). For example, the videos could either promote, discuss or debunk the misinformative topic being searched, or it could discuss a different misinformative topic—a topic that the user never searched for. Table 4 enlists our annotation values with description and examples. Please note that to curate misinformative topics for our study, we only considered demonstrably false conspiracy theories. But our annotation scheme does not classify videos for veracity, we rather check whether they promote, debunk or discuss a conspiratorial view related/unrelated to the search topic under audit.

5.3.1 *Annotation heuristics.*

We annotated videos as "debunking" (-1) when their narrative disputed, derided, or provided scientific evidence against any of the conspiratorial theories related to the particular misinformative topic being audited. For example, the video titled *Bill Maher Throws Out 9/11 Conspiracy Theorists On Live TV* was present in the *Top 5* recommendations while auditing the 9/11 misinformative topic. It mocks people supporting the 9/11 conspiracy theory and hence is annotated as "debunking". Conversely, we annotated videos as "promoting" (1) if they proposed, championed, or substantiated any theory or perspective that promotes inaccurate views related to the topic under audit. For example, the video titled *9/11 truthers attend Treason in America* shows interviews with 9/11 truthers—people who believe 9/11 was an inside job—and hence is annotated as "promoting". We annotated videos as "neutral" (0) when the content of the video presented a general discussion on the topic, without taking stance on conspiracy theories. For example, the

Annotation Value	Stance Description	Annotation Heuristics	No.of videos	Normalized Score	Sample Videos
					Video Title (Video URL, youtu.be/)
-1	debunking, mocking, disproving related misinformation	narrative of video disputes, mocks or provides authoritative evidence against conspiracy theories related to the topic under audit	430	-1 (D)	Bill Maher Throws Out 9/11 Conspiracy Theorists On Live TV (p80hXaM4QgU)
0	neutral & related to misinformation	narrative of the video does not take any stance on conspiracy theories related to the topic under audit	238	0 (N)	The Howard Stern Show and WCBS-2 On Sept. 11 (O3LT6FMF2f8)
1	promoting, supporting, justifying, explaining related misinformation	narrative of video promotes, supports or substantiates any conspiratorial views related to the topic under audit	374	1 (P)	9/11 truthers attend Treason in America (2-7GCs-2NUg)
2	debunking, mocking, disproving unrelated misinformation	narrative of video debunks, mocks or provides evidence against a conspiratorial view related to a topic different than the one under audit	64	-1 (D)	Did the Titanic Really Sink? The Olympic Switch Theory Debunked (_mpLRCqQ620)
3	neutral & related to another misinformation	narrative of the video does not take any stance on conspiracy theories unrelated to the topic under audit	25	0 (N)	JFK coverage 12:30pm-1:40pm 11/22/63 (pDOojsg62O0)
4	promoting, supporting, justifying, explaining unrelated misinformation	narrative of the video promotes, supports, justifies or explains any conspiratorial view unrelated to the topic under audit	66	-1 (P)	Mafia Boss Tells All - Jimmy Hoffa, JFK Assassination and Much More (__LxwaAEaL8)
5	not about misinformation	video content does not contain any conspiratorial views	1667	0 (N)	Former Abortionist Dr. Levatino At Virginia Tech (dIRcw45n9RU)
6	foreign language	video content in non-English language	35	translated & re-annotated	Las voces del 11S, documental en Español del Canal National Geographic (7rMQu2B_3vU)
7	undefined/unknown	annotators were unable to assign any of the above annotation values to the video	9	ignored	Ahmed Mohamed's Dad Pushes 9/11 Conspiracy Theories Online (CTkE0Etkszc)
8	removed	video removed from the platform at the time of annotation	35	ignored	n/a (tpSO7i70LHw)

Table 4. Description of the annotation scale and heuristics along with sample YouTube videos corresponding to each annotation value. We map our 9-point annotation scale to 3-point normalized scores with values -1 (Promoting, (P)) , 0 (Neutral, (N)) and 1 (Debunking, (D)). We have shared the list of 2,943 unique videos along with their annotation values in our online dataset.[3]

video titled *The Howard Stern Show and WCBS-2 On Sept. 11* shows clips depicting damage done to the World Trade Centre after the 9/11 attacks. We marked it as neutral since there is no discussion for and against 9/11 conspiracies.

Annotation values "2", "3", and "4" are similar to values "-1", "0", and "1", respectively, with the difference that they correspond to videos promoting, containing neutral content, or debunking conspiratorial information related to a topic different from the one being audited. For example, consider the scenario where audit experiments of 9/11 misinformative topic returned videos discussing conspiratorial information corresponding to John F. Kennedy's assassination or those pertaining to the Titanic's demise. To illustrate, we list two concrete examples here. Video titled *Did the Titanic Really Sink? The Olympic Switch Theory Debunked* was returned in the *Top 5* recommendations during the *Watch* audits of the 9/11 misinformative topic. The video content refutes the conspiracy theory that claims that the Titanic ship never sank. We annotated it as "debunking misinformation not related to the misinformative topic under audit" (annotation value = 2). In another example, a video titled *JFK coverage 12:30pm-1:40pm 11/22/63* showed news coverage about JFK's assassination without promoting or debunking any false conspiracies. We annotated that video as "neutral video not related to the misinformative topic under audit" (annotation value

[3] https://social-comp.github.io/YouTubeAudit-data/

= 3). On the other hand, a video *Mafia Boss Tells All - Jimmy Hoffa, JFK Assassination and Much More* discusses conspiracy theories surrounding JFK's assassination. We annotated that video as "promoting misinformation not related to the misinformative topic under audit" and assigned an annotation value of 4.

Additionally, we annotated videos as "not related to misinformation" (5) if the content of the video is not related to any misinformative topic. For example, one of the videos in our audit experiment, titled *SHOCKINGLY OFFENSIVE AUDITIONS Have Simon Cowell In A Rage! | ANGRY JUDGES | X Factor Global* is about a reality TV show audition. Since the content does not contain any information related to any misinformative topic, we annotated the video as unrelated to misinformation. Moreover, we annotated non-English videos as "foreign language" (annotation value = 6). We later translated the title, description, and the top few comments of these videos using Google Translate[4]. We then re-annotated them with the appropriate stance value lying between -1 to 5. For example, we re-annotated the Spanish video titled *Las voces del 11S, documental en Español del Canal National Geographic* as "debunking", since the comments within the video indicated that it debunks 9/11 conspiracy theory—the misinformative topic being audited. Finally, videos for which we were unable to assign any annotation value between -1 to 6, we annotated them as "undefined or unknown" (annotation value = 7). For example, the video titled *Ahmed Mohamed's Dad Pushes 9/11 Conspiracy Theories Online* mentions a 9/11 conspiracy tweet. Since the video neither discusses 9/11 events nor takes a stance for or against any conspiracy theory, the coder was unable to decide the annotation value. Because of the confusion it was marked as "unknown". During our annotation phase, we also find that YouTube had taken down 35 unique videos that were captured by our audit experiment. We make an ethical decision to not collect the data or annotate content that was removed by the platform.

After converging on our annotation scale and heuristic, two authors independently coded 158 videos to test for their inter-rater reliability. A high reliability score (Cohen's Kappa score of 0.80), suggested substantial agreement and offered credence to our annotation heuristic. The authors then split the annotation task of the remaining videos evenly between them. We next develop two scoring metrics to score the amount of misinformation in videos.

5.3.2 Normalized scores. The key goal of our audit investigation is to determine whether user activities—search and watch activities corresponding to a particular misinformative topic—leads to more misinformative content, either in the returned search result videos or through the recommended videos. Hence, for downstream analysis, we map our 9-point granular scale (-1 to 7) to a 3-point normalized score with values of -1, 0, and 1. The normalization process puts videos that contain any type of misinformation, whether related or unrelated to the searched topic, under the same bucket. For instance, if queries for the *9/11* topic result in a video enumerating conspiracies corresponding to missing Malaysian flight 370 (an example from our dataset), then we annotate the video is as promoting unrelated misinformation (annotation value = 4) with normalized score = 1. Annotation values of 2, 3, and 4 are mapped to -1, 0, and 1, respectively, while 5 and 6 are treated as neutral (see Table 4). We discard videos coded as 7 and 8, since annotators were either unable to identify their stance (value = 7) or the video was removed from the platform (value = 8). In total, we annotated 2,943 unique videos with 501, 1,980, and 462 videos marked as -1, 0, and 1.

5.3.3 SERP-MS Score. We develop a scoring metric **SERP-MS** (SERP Misinformation Score) that captures the amount of misinformation while taking into account the ranking of search results. SERP-MS = $\frac{\sum_{r=1}^{n}(x_i*(n-r+1))}{\frac{n*(n+1)}{2}}$; where r is the rank of the search result and n is the number of search results present in the SERP. We only consider the top 10 search results for computing SERP-MS.

[4] https://translate.google.com/

Feature	Topic	Stance	Comp.	Statistical Tests	Mean Diff.
Age	Flat Earth	N	Top5	KW H(3, 800)=18.28, p=0.0004	50 or older < all other age groups (post-hoc)
	Vaccine controversies	N	Top5	KW H(3,799)=24.65, p=1.8e-05	age 18-34 < all other age groups (post-hoc)
Gender	Flat Earth	N	Top5	MW U=74659, p=0.004	M > F
				MW U=3612, p=6.6e-07	M (50 or older) > F (50 or older)
	Moon landing conspiracy theories	N	Up-Next	MW U=2720, p=0.03	F > M
	Vaccine controversies	N	Top5	MW U=4068, p=0.002	M (age 35-50) > F (age 35-50)
				MW U=76206.5, p=0.02	M > F
		P	Top5	MW U=4443, p=0.01	M (age 18-34) > F (age 18-34)
			Up-Next	MW U=2880, p=0.04	M > F
				MW U=120, p=0.002	M (age 18-34) > F (age 18-34)
Geo-location	Moon landing conspiracy theories	P	Top5	MW U=4137.5, p=0.02	Hot > Cold

Table 5. RQ1b:*Watch* experiment results for demographics and geolocations, given accounts have built watch history after watching promoting (P), neutral (N) or debunking (D) videos. Mean corresponds to normalized scores for the annotated videos. Higher values indicate that accounts receive more promoting videos. For example, M (50 or older) >F (50 or older) indicates that males who are 50 or older and who watch neutral *flat earth* videos receive more promoting videos in their *Top 5* than females of the same age group.

Thus, SERP-MS is a continuous value ranging between -1 (all top 10 videos are debunking) to +1 (all top 10 are promoting).

6 RESULTS

In this section, we analyze our collected and annotated audit data to investigate our research questions and hypothesis (refer to Section 2). Our goal is to determine the effects of personalization attributes on the amount of misinformation returned in both *Search* and *Watch* experiments. Recall that, among the three YouTube components (*search results*, *Up-Next*, and *Top 5* recommendations), we can only collect *search results* for *Search* experiments. On the other hand, we collect all three components for *Watch* experiments. A test of normality reveals that our data is not normally distributed and our samples have unequal sizes. Hence, we opt for non-parametric tests. For all pairwise comparisons, we use Mann-Whitney U test. To perform multiple comparisons, we use Kruskal Wallis ANOVA followed by post-hoc Tukey HSD[5]. We report results using both normalized and SERP-MS scores. Note that the SERP-MS score is only calculated for the *search results* component.

6.1 RQ1: Effect of demographics and geolocation

In the first research question, we investigate the effect of demographics (age and gender) and geolocation on the amount of misinformation returned in various YouTube components for both brand new accounts and accounts that have build their watch history progressively by watching either promoting, neutral or debunking misinformative videos.

RQ1a [*Search* experiments]: How are *search results* affected for brand new accounts? We find no significant effect for gender (Mann-Whitney U = 7247667.0, p>0.48), age (Kruskal Wallis H(3,7616) = 0.00888, p>0.99), and geolocation (Mann-Whitney U=471803.0, p>0.496) when comparing using normalized scores. Use of SERP-MS score also shows non-significant results. Thus, *H1a*, *H1b* and *H1c* are not supported demonstrating that age, gender and geolocation do not have an impact on the amount of misinformation returned in search results for users who have newly created their YouTube accounts.

RQ1b [*Watch* experiments]: How are *search results, Up-Next,* and *Top 5* recommendations affected, given accounts have a watch history? We find that age has a significant effect for only

[5] Tukey HSD adjusts p-values automatically, thus controlling family-wise error rate for multiple comparisons.

Component	Topic	Test	Mean Diff (post-hoc)
Search Results	Vaccines controversies	KW H(2,6517)=6.2953, p=0.04	P >N & P >D
Top5	All	KW H(2,14740)=9.42, p=0.009	P >N & P >D
	9/11 conspiracy theories	KW H(2,2911)=186.68, p=2.9e-41	P >N & P >D
	Chemtrail conspiracy theory	KW H(2,2845)=73.20, p=1.31e-16	P >N & N >D
	Flat Earth	KW H(2,2980)=49.18, p=2.18e-11	N >P & D >P
	Moon Landing conspiracy theories	KW H(2,3005)=17.18, p=0.0002	P >N & D >N
	Vaccines controversies	KW H(2,2999)=48.54, p=2.9e-11	N >P & D >P
Up-Next	All	KW H(2,2963)=10.29, p=0.006	P >N
	9/11 conspiracy theories	KW H(2,487)=60.12, p=8.8e-14	P >N & P >D
	Chemtrail conspiracy theory	KW H(2,570)=16.12, p=0.0003	P >D
	Flat Earth	KW H(2,600)=26.29, p=1.96e-06	P >D & D >N
	Moon Landing conspiracy theories	KW (2,606)=5.99, p=0.049	D >N
	Vaccines controversies	KW H(2,600)=66.86, p=3.0e-15	D >N >P

Table 6. RQ2: Analyzing watch history effects on the three YouTube components. P, N, and D are means of the normalized scores of videos presented (via the YouTube components) to accounts that have built their watch histories by viewing promoting (P), neutral (N), and debunking (D) videos, respectively. For example, P > N indicates that accounts that watched promoting videos received more misinformation (or more promoting videos) compared to accounts that watched neutral videos.

two comparisons (refer Table 5). In both cases, older people do not receive more misinformation than the other younger age groups. Thus, *H1a* is rejected. Next, we find that gender has a significant effect across five comparisons involving certain combinations of search topics, watch stance, and YouTube components. Out of the five comparisons, *H1b* is supported for one case, where female accounts watching neutral moon landing videos receive more misinformation in their *Up-Next* component than corresponding male accounts watching the same videos. In all other significant comparisons, men receive more misinformation than females. For example, male accounts who watch neutral vaccination videos receive more promoting videos in their *Top 5* recommendations than female accounts that watch the same videos. Table 5 presents all the significant results.

We find that *H1c* holds only for the *Top 5* recommendations of *moon landing* topic. Accounts that watch promoting *moon landing* videos from Ohio (hot geolocation, region with the most interest) receive more promoting videos in their *Top 5* than those who watch the same videos from Georgia (cold geolocation or region exhibiting lowest interest in the topic). For other topics, geolocation did not have any significant effect on the amount of misinformation presented in *search results*, *Up-Next* and *Top 5* recommendations.

6.2 RQ2: Effect of watch history

Next, we explore the effect of watch history on the amount of misinformative content returned in our three YouTube components of interest. Note, that RQ2 only applies to our watch experiment, where an account has already built its watch history. Table 6 presents only the significant results. We discuss a handful. Statistical tests performed using SERP-MS did not give any significant results. Note that we apply this metric only on the *search results* component. Using the normalized score metric, we find that *H2* only holds for *search results* corresponding to the *vaccine controversies* topic (Kruskal Wallis H(2,6517)=6.2953, p=0.0429). This indicates that a user's previous watch history only affects the misinformative stance of videos presented in *search results* of the aforementioned topic. Post-hoc tests reveal that accounts that watch promoting anti-vaccination videos receive more promoting videos in their search results compared to those who watch neutral or debunking vaccination videos.

Next, we find that watch history has significant effects on the stance of misinformative videos presented in *Top 5* (Kruskal Wallis H(2,14740)=9.4235, p=0.0089) and *Up-Next* video recommendations (Kruskal Wallis H(2,2963)=10.2932, p=0.00581) when all topics are considered together. Post-hoc tests show that accounts that watch promoting videos receive more promoting results in both

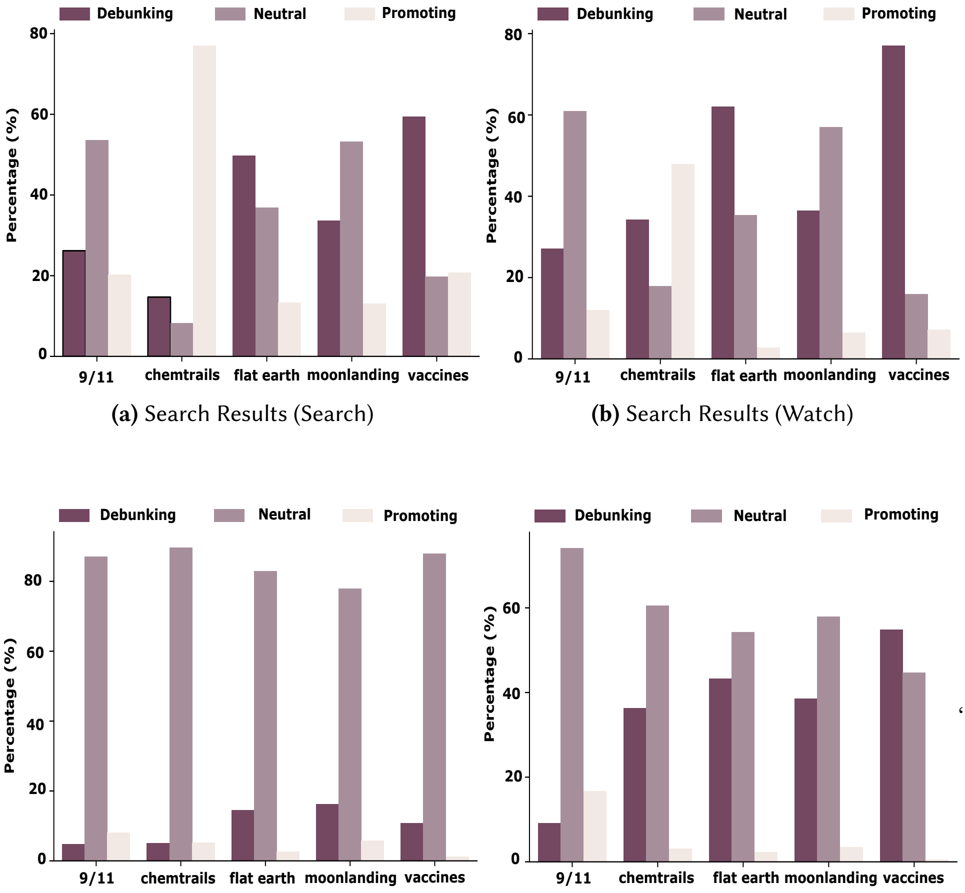
Up-Next and *Top 5* compared to those who watch either neutral or debunking videos. The effect of watch history for both these components is significant for all topics individually too. Thus, *H2* is supported for *Up-Next* and *Top 5* recommendations for all topics. We discuss the post-hoc test results for *vaccine controversies* and *chemtrail conspiracy theories* topics. Post-hoc tests for the *vaccine controversies* topic reveal that accounts that watch promoting anti-vaccination videos receive more debunking videos in their *Top 5* (Kruskal Wallis H(2,2999)=48.54, p=2.9e-11) and *Up-Next* (Kruskal Wallis H(2,600)=66.86, p=3.0e-15) components. This finding can be attributed to YouTube's initiative to reduce the recommendations of anti-vaccination videos. It is important to note that while recommendations of such videos have decreased, a filter bubble still exists with respect to the *search results*—people who watch promoting anti-vaccination videos were presented with more promoting content (Kruskal Wallis H(2,6517)=6.29, p=0.04). Post-hoc tests for *chemtrail conspiracy theories* topic demonstrate that accounts that watch videos promoting chemtrails conspiracies receive more promoting videos in their *Top 5* (Kruskal Wallis H(2,2845)=73.20, p=1.3e-16) and *Up-Next* (Kruskal Wallis H(2,5709)=16.12, p=0.0003) video recommendations than those who watch neutral and debunking videos. Whereas accounts that watch neutral chemtrails conspiracies receive more promoting videos in their *Top 5* compared to those who watch debunking videos of chemtrails. Table 6 lists the results for the remaining topic comparisons.

6.3 RQ3: Across topic differences

While in RQ1 and RQ2 we studied the effects of personalization attributes on the amount of misinformation presented to users in various YouTube components, in RQ3 we investigate whether misinformative content presented to users differ across the five misinformative topics.

RQ3a [*Search* experiments]: How does misinformative content present in *search results* of brand new accounts differ across topics? Figure 6a shows the proportion of promoting, neutral, and debunking videos across all topics in *Search* experiments. We find that *H3* is supported for *search results* of brand new accounts. Comparing both normalized scores (Kruskal Wallis H(4,1943)=467.29, p < 7.9e-100) and SERP-MS (Kruskal Wallis H(4,98)=51.1, p < 2.1e-10) across topics show that the amount of misinformation significantly differs among topics. Post-hoc comparisons using Tukey HSD (on both score metrics) reveal that the *chemtrail conspiracy theory* topic harbors significantly more misinformative *search results* compared to all other topics. Figure 6a also demonstrates the largest amount of promoting videos in the *chemtrails* topic. We discuss the possible reasons for this occurrence in Section 7.2

RQ3b [*Watch* experiments]: How does misinformative content present in *search results*, *Up-Next*, and *Top 5* recommendations of accounts having a watch history differ across topics? Figure 6b, Figure 6d and Figure 6c show the proportion of promoting, neutral, and debunking videos across all topics collected from *search results*, *Up-Next* and *Top 5* recommendations respectively in *Watch* experiments. *H3* is supported for all the three YouTube components for accounts having a watch history. Comparing both normalized scores and SERP-MS across topics, show that topics have a significant effect on the amount of misinformation present in *search results*, *Up-Next* (Kruskal Wallis H(4,2963)=375, p < 6.7e-80), and *Top 5* recommended videos (Kruskal Wallis H(4,14740)=390.6, p < 2.9e-83). Recall that SERP-MS is applicable only for the *search results* component. Post-hoc comparisons using Tukey HSD reveal that *chemtrail conspiracy theories* has significantly more misinformation in its *search results* compared to all other topics. Figure 6b exhibits the largest amount of promoting videos on that topic. On the other hand, the amount of misinformation present in *Up-Next* and *Top 5* recommendations for *9/11 conspiracy theory* topic is significantly more than other topics. This is also evident from Figures 6c and 6d.



(a) Search Results (Search)

(b) Search Results (Watch)

(c) Top 5 (Watch)

(d) Up-Next (Watch)

Fig. 6. RQ3: Percentages of video stances for each topic.

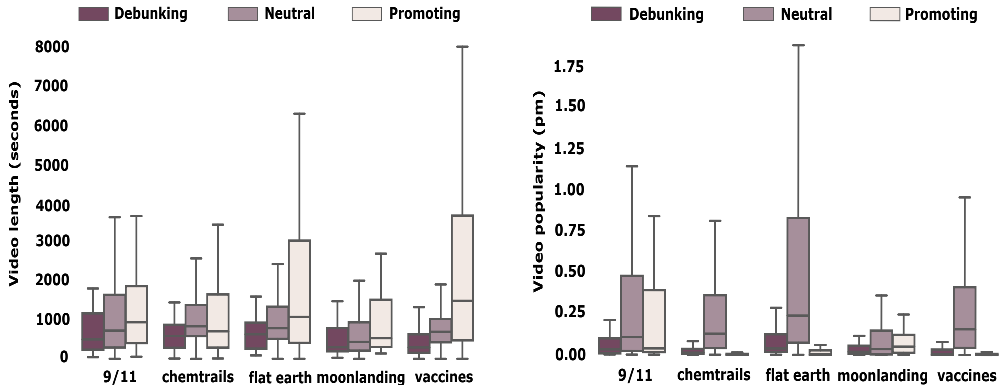
6.4 Analyzing Video Length and Popularity

Analyzing video length, we observe that promoting videos are longer than neutral and debunking videos across all misinformative topics, except *chemtrails conspiracies* where they are slightly shorter than neutral videos and longer than debunking ones (see Figure 7a). For all topics, debunking videos are the shortest compared to other stances. We also observe that neutral videos are the most popular (see Figure 7b), where popularity is calculated using the *popularity metric (pm)*. Topic *9/11* has more popular videos compared to other topics. On the other hand, for topic *moon landing*, popularity of videos under each stance is almost the same. Although the percentage of videos promoting *chemtrails conspiracies* is highest when compared to other misinformative topics, they are the least popular videos.

7 DISCUSSION

7.1 Effect of demographics and geolocation on misinformation

Modern search engines filter, rank, and personalize results before presenting to a user. These information retrieval systems make decisions about the relevance of results without considering accuracy and credibility—a fact most people are unaware of [43]. Motivated by several media



(a) Video length

(b) Video popularity

Fig. 7. Box plots of (a) video length in seconds and (b) video popularity (*pm*) for each stance under each topic.

reports pointing out the prevalence of misinformation on search spaces, we audited YouTube to empirically determine the extent of conspiratorial content present in its search and recommended video results. We investigate the role played by personalization attributes (age, gender, geolocation, and watch history). Our analyses show little evidence to support that user demographics and geolocation play any role in amplifying misinformation in search results for users who have newly started their search journey—those with brand new accounts. On the contrary, once they have a watch history, we find that demographics and geolocation attributes do exert an effect. However, this effect pertains to only certain combinations of personalization attributes and varies with the topic under audit investigation. We saw significant gender differences in 8 comparisons and in all but one case men (account gender set to "male") were recommended more misinformative videos. Perhaps more surprisingly, in 4 of these cases men were watching neutral videos and yet ended up with significantly higher misinformative videos recommendations. While we do not know why YouTube's algorithm showed this behavior, our observed gender-based differences have important societal implications, especially for certain misinformative topics, such as *vaccine controversies.* For example, a survey of 2,300 people in United States revealed that percentage of male anti-vaxxers are more than females [45]. Therefore, recommending videos that promote misinformative topics to men can inflict more harm by reconfirming their pro-conspiracy beliefs. Moreover, recommending promoting videos to men who are drawn to neutral information and have yet not developed a strong pro-conspiracy belief towards the topic is even more problematic because it might increase their chances of forming pro-conspiracy beliefs.

7.2 Effect of watch history on misinformation

One of the goals of our audit investigations was to verify several anecdotal claims criticizing YouTube for surfacing misinformative content in its recommendations [8, 42, 83]. These claims accused the platform of driving users into a misinformation rabbit hole—a phenomenon where people watching videos promoting misinformation are presented with more such videos in the search results and recommendations. Contrary to these blanket claims, we observe variability in YouTube's behavior in presenting recommendations to accounts having a watch history across different misinformative topics. Comparing the stances of our annotated data obtained from the search results of accounts with a watch history shows that YouTube's search algorithm fares well for the flat earth and vaccine topic. On the other hand, we witness a large proportion of videos promoting misinformation for the *chemtrails* topic (refer to Figure 6b). This observation can be attributed to YouTube's recent effort to censor misinformative content belonging to select search topics. In an announcement to the public, the platform pledged to reduce misinformative content belonging to topics like 9/11, flat earth and medical misinformation [76]. Thus, we believe the percentage of *search results* promoting these misinformative topics is less compared to other topics like *chemtrail conspiracies*.

Our audits reveal that people who watch promoting videos for certain misinformative topics (for example, 9/11 conspiracies) are recommended more of such videos in their *Up-Next* and *Top 5* recommendations compared to those who watch neutral or debunking videos. These findings indicate that the recommendation algorithm is biased towards the stance of videos watched by the user for certain misinformative topics (refer Table 6). In another observation we find that for users watching videos on the vaccine topic, both *Top 5* and *Up-Next* recommendations return negligible proportion of videos promoting "vaccine hesitancy", 1.2% and 0.5% respectively. Statistical tests reveal that people watching promoting anti-vaccination videos receive more debunking videos in their recommendations compared to people who watch neutral or debunking videos. However, a filter bubble effect still exists for the *search results* component, where people watching anti-vaccination videos are presented with more such results. This variability in YouTube's behavior across search topics suggests that YouTube is modifying its search ranking and recommendation algorithms selectively, handpicking topics that are highlighted by media reports and technology critics (for e.g. reports around anti-vaccine video recommendations). These observations are concerning, since all misinformative topics are high impact, popular and perennial and hence are likely to affect a large population of users' search experiences. Our findings serve as an important call-to-action for YouTube to develop more universal approach that offers a comprehensive solution to the problem of misinformation.

7.3 Tackling search engine enabled misinformation

Complete eradication of misinformation from YouTube requires time and significant resources. In the interim, YouTube can take several steps to tackle the problem of misinformation on its platform. It can begin by giving priority to monitoring certain misinformative topics that have a wider negative impact on society. Which misinformative topics are a threat to public well being? While "vaccine hesitancy" is now one of the top 10 global threats of 2019 [56] and has led four European nations to lose their "measles free" status [51], seemingly harmless pizzagate conspiracy led a man to fire shots in a pizza parlor [77]. We recommend that YouTube should identify high-impact and popular misinformative topics. Our work itself suggests a technique to curate such misinformative topics that are perennial, popular, and searched by a large number of people. Misinformative content belonging to the selected impactful topics can be filtered, fact-checked, and accordingly censored from the platform.

But is censoring the misinformative content enough? Our audit experiments reveal that YouTube recommendations are still biased towards the misinformative stance of videos watched by a user. Given that almost 500 hours of content is uploaded to YouTube every hour [27], censorship might not be a comprehensive solution to fix this algorithmic bias. There is a need to break the filter bubble effect by recommending debunking videos to people who watch videos promoting misinformative content. YouTube can start by identifying and modifying recommendations of vulnerable populations who could be targets for certain misinformative topics. Our audit experiments revealed one such demographics. For example, we found YouTube recommending promoting videos to men who watched neutral misinformative videos.

Our audits also revealed variability in YouTube's behavior towards certain misinformative topics— an indication of a reactive strategy of dealing with misinformation. We recommend the platform to also proactively reveal the workings of its algorithm. For example, users can be told "you are recommended video A because you viewed videos C and D". Given the complexity of algorithms used by search engines and the interplay between the data and algorithm, even an expert in the area might not be able to predict algorithmic output [68]. Thus, there is also an inherent need for platforms to conduct audit studies that can help reveal biases present in their algorithm.

While we discussed some nascent steps that YouTube can take towards eradication of misinformation from its platform, this feat cannot be achieved without having proper content policies and infrastructure in place. Currently YouTube's community guidelines do not disallow misinformative content [92]. There is a need to have appropriate policies in place that not only prohibit posting misinformative content on the platform but also ensure that posting advertisements on misinformative videos is not financially incentivized. The challenge of having appropriate infrastructure to implement these policies still remains.

8 LIMITATION AND FUTURE WORK

Our study is not without limitations. We do not perform repeated searches of our search queries over several days which is essential to study the longitudinal effect of personalization. We plan to conduct continuous audit runs with repeated searches in future. We also tested the effect of geolocation feature only for regions within Unites States, but conducting audits over a global scale is a fertile area for future endeavors. Our *Search* and *Watch* audit runs had a gap of three months. Thus, we do not perform any comparisons between the *search result* components of the two audits. We do not take into account the stance of a search query and how that affects the search results. We make this conscious choice because our methodology for compiling high impact search queries, by definition, focuses on realistic searches that were *most used* by real users on YouTube.

Identifying videos that promote conspiracies and inaccurate content or those that debunk them is a challenging task. To make such distinctions with high precision, we used qualitative coding to annotate videos. In addition to the video content, we referred to metadata attributes, such as video title, description, and user reactions present in the comments section. We found that videos relating to misinformative topics exhibit special characteristics. For example, pro-conspiracy videos are mostly longer while neutral videos are more popular. We believe that such distinctive features along with features used in our manual annotation process can be leveraged to build machine learning models that can identify the stance of videos.

While we audit three major components of YouTube, other components such as home-pages and trending section can also be examined. Auditing search queries presented by YouTube's auto-complete feature for their stance is also left for future investigation. Moreover, understanding how misinformative search results and recommendations affect users' search intent [89, 90] are other compelling avenues for future research.

9 CONCLUSION

In this study, we conducted two sets of audit experiments on YouTube platform to empirically determine the effect of personalization attributes (age, gender, geolocation and watch history) on the amount of misinformation prevalent in YouTube searches and recommendations. Our audits resulted in a dataset of 56,475 videos that we annotated for their stance (promoting, neutral and debunking) and relevance towards the misinformative search topics. We found that the personalization attributes affect the amount of misinformation in recommendations once the user develops a watch history. Our study also suggests that YouTube is modifying its search and recommendation algorithm for certain misinformative topics like *vaccine controversies*. Our audit methodologies can be used for investigating other search engines for misinformative search results and recommendations. We believe such audit studies will inform the need for building search engines that retrieve and present results ranked according to both relevance and credibility.

10 ACKNOWLEDGEMENTS

We would like to thank our reviewers and entire Social Computing lab group at Virginia Tech for providing valuable feedback that helped improve this work. This work is partially supported by the Social Science Research Council's Social Media and Democracy Research grant.

REFERENCES

[1] Jonathan Albright. 2018. *UnTrue Tube – YouTube's Conspiracy Ecosystem.* https://datajournalismawards.org/projects/untrue-tube-youtubes-conspiracy-ecosystem/

[2] Hunt Allcott and Matthew Gentzkow. 2017. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31, 2 (2017), 211–36. https://doi.org/10.1257/jep.31.2.211

[3] Gordon W. Allport and Leo Postman. 1946. An Analysis of Rumor. *Public Opinion Quarterly* 10, 4 (1946), 501–517. https://doi.org/10.1093/poq/10.4.501

[4] Rebecca Lee Armstrong. 2019. New Survey Suggests 10% of Americans Believe the Moon Landing Was Fake. (2019). https://www.satelliteinternet.com/resources/moon-landing-real-survey/

[5] Woodrow Bellamy. 2019. Malaysia Airlines Flight 370 Final Report Inconclusive. (2019). https://www.aviationtoday.com/2018/08/02/malaysia-airlines-flight-370-final-report-inconclusive/

[6] Alessandro Bessi, Mauro Coletto, George Alexandru Davidescu, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2015. Science vs conspiracy: Collective narratives in the age of misinformation. *PloS one* 10, 2 (2015), e0118093. https://doi.org/10.1371/journal.pone.0118093

[7] Thomas DG Burgess II and Stephen M Sales. 1971. Attitudinal effects of "mere exposure": A reevaluation. *Journal of Experimental Social Psychology* 7, 4 (1971), 461–472. https://doi.org/10.1016/0022-1031(71)90078-3

[8] Nick Carne. 2019. 'Conspiracies' dominate YouTube climate modification videos. (2019). https://cosmosmagazine.com/social-sciences/conspiracies-dominate-youtube-climate-modification-videos

[9] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web.* ACM, 675–684. https://doi.org/10.1145/1963405.1963500

[10] Le Chen, Ruijun Ma, Anikó Hannák, and Christo Wilson. 2018. Investigating the Impact of Gender on Rank in Resume Search Engines. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).* ACM, Article 651, 651:1–651:14 pages. https://doi.org/10.1145/3173574.3174225

[11] Le Chen, Alan Mislove, and Christo Wilson. 2015. Peeking beneath the hood of uber. In *Proceedings of the 2015 internet measurement conference.* 495–508. https://doi.org/10.1145/2815675.2815681

[12] Le Chen, Alan Mislove, and Christo Wilson. 2016. An empirical analysis of algorithmic pricing on amazon marketplace. In *Proceedings of the 25th International Conference on World Wide Web.* 1339–1349. https://doi.org/10.1145/2872427.2883089

[13] Xinran Chen, Sei-Ching Joanna Sin, Yin-Leng Theng, and Chei Sian Lee. 2015. Why students share misinformation on social media: Motivation, gender, and study-level differences. *The Journal of Academic Librarianship* 41, 5 (2015), 583–592. https://doi.org/10.1016/j.acalib.2015.07.003

[14] Cisco. 2019. Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper. (2019). https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html

[15] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems.* https://doi.org/10.1145/2959100.2959190

[16] Munmun De Choudhury, Meredith Ringel Morris, and Ryen W White. 2014. Seeking and sharing health information online: comparing search engines and social media. In *Proceedings of the SIGCHI conference on human factors in computing systems.* 1365–1376. https://doi.org/10.1145/2556288.2557214

[17] Nicholas Diakopoulos, Daniel Trielli, Jennifer Stark, and Sean Mussenden. 2018. I Vote For– How Search Informs Our Choice of Candidate. *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple, M. Moore and D. Tambini (Eds.)* 22 (2018). https://www.academia.edu/37432634/I_Vote_For_How_Search_Informs_Our_Choice_of_Candidate

[18] Renee Diresta. 2018. *The Complexity of Simply Searching for Medical Advice.* https://www.wired.com/story/the-complexity-of-simply-searching-for-medical-advice/

[19] James N Druckman and Michael Parkin. 2005. The impact of media bias: How editorial slant affects voters. *The Journal of Politics* 67, 4 (2005), 1030–1049. https://doi.org/10.1111/j.1468-2508.2005.00349.x

[20] Andrea Freyer Dugas, Yu-Hsiang Hsieh, Scott R Levin, Jesse M Pines, Darren P Mareiniss, Amir Mohareb, Charlotte A Gaydos, Trish M Perl, and Richard E Rothman. 2012. Google Flu Trends: correlation with emergency department

influenza rates and crowding metrics. *Clinical infectious diseases* 54, 4 (2012), 463–469. https://doi.org/10.1093/cid/cir883

[21] Robert Epstein, Ronald E Robertson, David Lazer, and Christo Wilson. 2017. Suppressing the search engine manipulation effect (SEME). *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–22. https://doi.org/10.1145/3134677

[22] Adrien Friggeri, Lada Adamic, Dean Eckles, and Justin Cheng. 2014. Rumor cascades. In *Eighth International AAAI Conference on Weblogs and Social Media*. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewFile/8122/8110

[23] Eric Gaillard. 2019. Facebook Under Fire for Permitting Anti-Vax Groups. (2019). https://www.thedailybeast.com/facebook-under-fire-for-permitting-anti-vaccination-groups

[24] Tarleton Gillespie. 2014. The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society* 167 (2014). https://www.microsoft.com/en-us/research/wp-content/uploads/2014/01/Gillespie_2014_The-Relevance-of-Algorithms.pdf

[25] Andrew Guess, Jonathan Nagler, and Joshua Tucker. 2019. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science advances* 5, 1 (2019), eaau4586. https://doi.org/10.1126/sciadv.aau4586

[26] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. 2013. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 729–736. https://doi.org/10.1145/2487788.2488033

[27] James Hale. 2019. More Than 500 Hours Of Content Are Now Being Uploaded To YouTube Every Minute. (2019). https://www.tubefilter.com/2019/05/07/number-hours-video-uploaded-to-youtube-per-minute/

[28] Aniko Hannak, Piotr Sapiezynski, Arash Molavi Kakhki, Balachander Krishnamurthy, David Lazer, Alan Mislove, and Christo Wilson. 2013. Measuring Personalization of Web Search. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13)*. ACM, 527–538. https://doi.org/10.1145/2488388.2488435

[29] Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. 2014. Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 conference on internet measurement conference*. 305–318. https://doi.org/10.1145/2663716.2663744

[30] Anikó Hannák, Claudia Wagner, David Garcia, Alan Mislove, Markus Strohmaier, and Christo Wilson. 2017. Bias in Online Freelance Marketplaces: Evidence from TaskRabbit and Fiverr. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, 1914–1933. https://doi.org/10.1145/2998181.2998327

[31] Google Trends Help. 2020. Explore results by region. (2020). https://support.google.com/trends/answer/4355212?hl=en

[32] Benjamin D Horne and Sibel Adali. 2017. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Eleventh International AAAI Conference on Web and Social Media*.

[33] Desheng Hu, Shan Jiang, Ronald E. Robertson, and Christo Wilson. 2019. Auditing the Partisanship of Google Search Snippets. In *The World Wide Web Conference (WWW '19)*. ACM, 693–704. https://doi.org/10.1145/3308558.3313654

[34] Shan Jiang, Ronald E Robertson, and Christo Wilson. 2019. Bias Misperceived: The Role of Partisanship and Misinformation in YouTube Comment Moderation. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 278–289.

[35] Chloe Kliman-Silver, Aniko Hannak, David Lazer, Christo Wilson, and Alan Mislove. 2015. Location, Location, Location: The Impact of Geolocation on Web Search Personalization. In *Proceedings of the 2015 Internet Measurement Conference (IMC '15)*. ACM, 121–127.

[36] Chloe Kliman-Silver, Aniko Hannak, David Lazer, Christo Wilson, and Alan Mislove. 2015. Location, location, location: The impact of geolocation on web search personalization. In *Proceedings of the 2015 Internet Measurement Conference*. ACM, 121–127. https://doi.org/10.1145/2815675.2815714

[37] Peter Knight. 2008. Outrageous conspiracy theories: Popular and official responses to 9/11 in Germany and the United States. *New German Critique* 103 (2008), 165–193. https://doi.org/10.1215/0094033X-2007-024

[38] Srijan Kumar and Neil Shah. 2018. False information on web and social media: A survey. *arXiv preprint arXiv:1804.08559* (2018).

[39] Srijan Kumar, Robert West, and Jure Leskovec. 2016. Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In *Proceedings of the 25th international conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 591–602. https://doi.org/10.1145/2872427.2883085

[40] William Langewiesche. 2019. What Really Happened to Malaysia's Missing Airplane. (2019). https://www.theatlantic.com/magazine/archive/2019/07/mh370-malaysia-airlines/590653/

[41] Bruno Latour and Steve Woolgar. 2013. *Laboratory life: The construction of scientific facts*. Princeton University Press.

[42] Paul Lewis and Erin McCormick. 2018. How an ex-YouTube insider investigated its secret algorithm. (2018). https://www.theguardian.com/technology/2018/feb/02/youtube-algorithm-election-clinton-trump-guillaume-chaslot

[43] Ramona Ludolph, Ahmed Allam, and Peter Schulz. 2016. Manipulating Google's Knowledge Graph Box to Counter Biased Information Processing During an Online Search on Vaccination: Application of a Technological Debiasing

Strategy. *Journal of Medical Internet Research* 18 (2016), e137. https://doi.org/10.2196/jmir.5430

[44] Logan McDonald and Caroline O'Donovan. 2019. YouTube Continues To Promote Anti-Vax Videos As Facebook Prepares To Fight Medical Misinformation. (2019). https://www.buzzfeednews.com/article/carolineodonovan/youtube-anti-vaccination-video-recommendations

[45] Annalisa Merelli. 2015. The average anti-vaxxer is probably not who you think she is. (2015). https://qz.com/355398/the-average-anti-vaxxer-is-probably-not-who-you-think-she-is/

[46] Danaë Metaxa, Joon Sung Park, James A Landay, and Jeff Hancock. 2019. Search Media and Elections: A Longitudinal Investigation of Political Search Results. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–17. https://doi.org/10.1145/3359231

[47] P Takis Metaxas and Yada Pruksachatkun. 2017. Manipulation of search engine results during the 2016 US congressional elections. (2017).

[48] Tanushree Mitra and Eric Gilbert. 2015. Credbank: A large-scale social media corpus with associated credibility annotations. In *Ninth International AAAI Conference on Web and Social Media*. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/download/10582/10509

[49] NASA. 2001. NASA Facts. (2001). https://web.archive.org/web/20151213100852/ftp://ftp.hq.nasa.gov/pub/pao/media/2001/lunar_landing.pdf

[50] BBC News. 2011. 9/11 conspiracy theories: How they've evolved. (2011). https://www.bbc.com/news/magazine-14665953

[51] BBC News. 2019. Measles: Four European nations lose eradication status. (2019). https://www.bbc.com/news/health-49507253

[52] National Oceanic and Atmospheric Administration. 2016. Do Contrails Affect Conditions on the Surface? (2016). https://www.nesdis.noaa.gov/content/do-contrails-affect-conditions-surface

[53] Alex Olshansky. 2018. Conspiracy Theorizing and Religious Motivated Reasoning: Why the Earth 'Must' Be Flat. (2018).

[54] World Health Organization. 2019. MMR and autism. (2019). https://www.who.int/vaccine_safety/committee/topics/mmr/mmr_autism/en/

[55] World Health Organization. 2019. Six common misconceptions about immunization. (2019). https://www.who.int/vaccine_safety/initiative/detection/immunization_misconceptions/en/index3.html

[56] World Health Organization. 2019. *Ten threats to global health in 2019.* https://www.who.int/emergencies/ten-threats-to-global-health-in-2019

[57] Eli Pariser. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think.* Penguin.

[58] Gordon Pennycook, Tyrone D Cannon, and David G Rand. 2018. Prior exposure increases perceived accuracy of fake news. *Journal of experimental psychology: general* (2018). https://doi.org/10.1037/xge0000465

[59] Larry Peterson, Tom Anderson, David Culler, and Timothy Roscoe. 2003. A Blueprint for Introducing Disruptive Technology into the Internet. *SIGCOMM Computer Communication Review* 33, 1 (2003), 59–64. https://doi.org/10.1145/774763.774772

[60] K Purcell. 2011. Findings: Search and email remain the top online activities| Pew Internet & American Life Project. *Pew Research Center's Internet & American Life Project* (2011).

[61] Vahed Qazvinian, Emily Rosengren, Dragomir R Radev, and Qiaozhu Mei. 2011. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics, 1589–1599.

[62] M. Rajdev and K. Lee. 2015. Fake and Spam Messages: Detecting Misinformation During Natural Disasters on Social Media. In *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Vol. 1. 17–20. https://doi.org/10.1109/WI-IAT.2015.102

[63] Ronald E. Robertson, Shan Jiang, Kenneth Joseph, Lisa Friedland, David Lazer, and Christo Wilson. 2018. Auditing Partisan Audience Bias Within Google Search. *Proceedings of ACM on Human Computer Interaction* 2, CSCW (2018), 148:1–148:22. https://doi.org/10.1145/3274417

[64] Ronald E. Robertson, David Lazer, and Christo Wilson. 2018. Auditing the Personalization and Composition of Politically-Related Search Engine Results Pages. In *Proceedings of the 2018 World Wide Web Conference (WWW '18)*. International World Wide Web Conferences Steering Committee, 955–965. https://doi.org/10.1145/3178876.3186143

[65] Ashley Rodriguez. 2018. YouTube's algorithms can drag you down a rabbit hole of conspiracies, researcher finds. (2018). https://qz.com/1215937/research-youtubes-algorithms-lead-down-a-rabbit-hole-of-conspiracies/

[66] Mattia Samory and Tanushree Mitra. 2018. Conspiracies Online: User Discussions in a Conspiracy Community Following Dramatic Events. In *ICWSM*. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/viewFile/17907/17025

[67] Mattia Samory and Tanushree Mitra. 2018. 'The Government Spies Using Our Webcams': The Language of Conspiracy Theories in Online Discussions. *Proceedings of the ACM on Human-Computer Interaction* 2 (2018), 1–24. https:

//doi.org/10.1145/3274421

[68] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. An algorithm audit. *Data and Discrimination: Collected Essays. Washington, DC: New America Foundation* (2014), 6–10. http://www-personal.umich.edu/~csandvig/research/An%20Algorithm%20Audit.pdf

[69] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry* 22 (2014). https://pdfs.semanticscholar.org/b722/7cbd34766655dea10d0437ab10df3a127396.pdf

[70] Susan Scutti. 2019. Facebook to target vaccine misinformation with focus on pages, groups, ads. (2019). https://www.cnn.com/2019/03/07/health/facebook-anti-vax-messages-bn/index.html

[71] Geoff Shepard. 2015. *The Real Watergate Scandal: Collusion, Conspiracy, and the Plot That Brought Nixon Down.* Simon and Schuster.

[72] Tamotsu Shibutani. 1966. *Improvised news: A sociological study of rumor.* Ardent Media. https://doi.org/10.2307/2574636

[73] Natalie Jomini Stroud. 2010. Polarization and partisan selective exposure. *Journal of communication* 60, 3 (2010), 556–576. https://doi.org/10.1111/j.1460-2466.2010.01497.x

[74] Cass R Sunstein. 2014. *Conspiracy theories and other dangerous ideas.* Simon and Schuster.

[75] American Osteopathic Association Media Team. 2019. 45% of American adults doubt vaccine safety, according to survey. (2019). https://osteopathic.org/2019/06/24/45-of-american-adults-doubt-vaccine-safety-according-to-survey/

[76] The YouTube Team. 2019. Continuing our work to improve recommendations on YouTube. (2019). https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html

[77] Los Angeles Times. 2017. Man inspired by false 'pizzagate' rumor on Internet pleads guilty to shooting at D.C. restaurant. (2017). https://www.latimes.com/nation/nationnow/la-na-pizzagate-shooting-20170324-story.html

[78] The New York Times. 2004. The New York Times/CBS News Poll. (2004). http://www.nytimes.com/packages/html/politics/20040429_poll/20040429_poll_results.pdf

[79] Dustin Tingley and Gernot Wagner. 2017. Solar geoengineering and the chemtrails conspiracy on social media. *Palgrave Communications* 3, 1 (2017), 12. https://doi.org/10.1057/s41599-017-0014-3

[80] Daniel Trielli and Nicholas Diakopoulos. 2019. Search As News Curator: The Role of Google in Shaping Attention to News Information. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).* ACM, Article 453, 453:1–453:15 pages. https://doi.org/10.1145/3290605.3300683

[81] Nicholas Vincent, Isaac Johnson, Patrick Sheehan, and Brent Hecht. 2019. Measuring the Importance of User-Generated Content to Search Engines. *Proceedings of the International AAAI Conference on Web and Social Media* 13, 01 (2019), 505–516. https://www.aaai.org/ojs/index.php/ICWSM/article/download/3248/3116/

[82] Xuanhui Wang, Nadav Golbandi, Michael Bendersky, Donald Metzler, and Marc Najork. 2018. Position bias estimation for unbiased learning to rank in personal search. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining.* ACM, 610–618. https://doi.org/10.1145/3159652.3159732

[83] Cale Guthrie Weissman. 2019. Despite recent crackdown, YouTube still promotes plenty of conspiracies. (2019). https://www.fastcompany.com/90307451/despite-recent-crackdown-youtube-still-promotes-plenty-of-conspiracies

[84] Wikipedia. 2002. Conspiracy theory. (2002). https://en.wikipedia.org/wiki/Conspiracy_theory

[85] Wikipedia. 2003. List of conspiracy theories. (2003). https://en.wikipedia.org/wiki/List_of_conspiracy_theories

[86] Wikipedia. 2019. Malaysia Airlines Flight 370. (2019). https://en.wikipedia.org/wiki/Malaysia_Airlines_Flight_370

[87] Wikipedia. 2019. Project MKUltra. (2019). https://en.wikipedia.org/wiki/Project_MKUltra

[88] Michael Wood. 2013. Has the internet been good for conspiracy theorising. *PsyPAG Quarterly* 88 (2013), 31–34. http://www.psypag.co.uk/wp-content/uploads/2013/09/Issue-88.pdf#page=33

[89] Zhijing Wu, Yiqun Liu, Qianfan Zhang, Kailu Wu, Min Zhang, and Shaoping Ma. 2019. The influence of image search intents on user behavior and satisfaction. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining.* ACM, 645–653. https://doi.org/10.1145/3289600.3291013

[90] Xiaohui Xie, Yiqun Liu, Maarten De Rijke, Jiyin He, Min Zhang, and Shaoping Ma. 2018. Why people search for images using web search engines. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining.* ACM, 655–663. https://doi.org/10.1145/3159652.3159686

[91] YouGov. 2018. Most flat earthers consider themselves very religious. (2018). https://today.yougov.com/topics/philosophy/articles-reports/2018/04/02/most-flat-earthers-consider-themselves-religious

[92] YouTube. 2020. YouTube Community Guidelines. (2020). https://www.youtube.com/about/policies/#community-guidelines

[93] Marvin Zonis and Craig M Joseph. 1994. Conspiracy thinking in the Middle East. *Political Psychology* (1994), 443–459. https://doi.org/10.2307/3791566

[94] Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. 2018. Detection and Resolution of Rumours in Social Media: A Survey. *ACM Comput. Surv.* 51, 2, Article 32 (2018), 32:1–32:36 pages. https://doi.org/10.1145/3161603