# UP NEXT:
# YouTube's Recommendation System and the 2019 Canadian Federal Election

by

Daniel Cockcroft

A thesis submitted in partial fulfillment of the requirements for the degrees of

Master of Arts

and

Master of Library and Information Studies

Digital Humanities and School of Library and Information Studies

University of Alberta

# ABSTRACT

In the months leading up to the 2016 election in the United States, YouTube's recommendation algorithm decidedly favored pro-Trump videos, fake news and conspiracy theories. In this thesis, I question whether such bias is present in the context of the 2019 federal election in Canada. To do so, I make use of open-source software to gather recommendation data related to three of the candidates: Conservative Party of Canada leader Andrew Scheer, New Democratic Party leader Jagmeet Singh, and Liberal Party of Canada leader Justin Trudeau. Using the same data, I will also study the media bias and factual accuracy of the sources recommended. My results show that YouTube's recommender system is susceptible to influence by audiences and shows bias towards Andrew Scheer and against Justin Trudeau. Given my results and evidence provided by other researchers, this study stresses the need for ethical algorithm design, including proactive approaches for increased transparency, regulatory oversight, and increased public awareness.

## ACKNOWLEDGEMENTS

I would like to extend a big thank you to my supervisory committee, Tami, Astrid, and Harvey for keeping me inspired and asking the right questions. Your countless hours of work have not gone unnoticed, and through your efforts you've made me a better researcher and writer.

I would also like to thank my friends for their inspiration and confidence in me, especially Kaitlyn and Kenzie. I am also immensely grateful for the support and flexibility given to me by my previous employers Leah Vanderjagt and Sharon Farnel. Special thanks to my graduate advisor, Nicola DiNicola. I couldn't have done it without you all!

Most of all, I'd like to thank my partner, Jill. Thanks for bearing with me.

# TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER ONE: INTRODUCTION

## Overview

YouTube is the most popular video hosting website online. Launched in 2005 and shortly thereafter acquired by Google, the website quickly grew into one of the most visited websites on the internet. The high amount of user traffic on YouTube paired with the large volume of video uploads every day results in unique information retrieval challenges. In order to best deliver this vast amount of content to the users who would engage with it the most, YouTube needed to start recommending similar content alongside the videos that users were watching already. To make YouTube more profitable, Google sought to keep viewers watching videos on their platform beyond their initial visits: More video views resulted in more advertisement views, and more advertisement views resulted in more revenue. In 2010, YouTube launched 'leanback', a feature that added a sidebar of similar or related titles to keep users watching (McCracken, 2010). Over time, this feature evolved to make use of complex ranking and sorting algorithms to deliver highly personalized recommendations to individual users. Much in part due to these developments, YouTube sees over a billion hours of content watched on their platform--per day (YouTube for Press. n.d.), 70% of which is a result of recommendations (Zhou et al., 2010).

For most people, browsing YouTube is a relatively effortless activity. User attention is kept occupied by a never-ending flow of personalized video suggestions. That effort is reduced further as YouTube's autoplay feature begins playing a recommended video immediately after the current video has finished. As users view videos, YouTube's system keeps note of what they have watched, how long they watch, and other engagement patterns including likes and comments. In addition, YouTube's system also looks at which videos other users are most likely to watch next. Making an algorithmic approximation based on these signals, the system can make predictions of other videos for users to enjoy. Clicking on a video and watching a high percentage of it is interpreted as a signal of interest. As a result, YouTube's system continues to feed viewers with content that is met with positive signals. At its core, the problem with this method is that there are other ingredients in a good meal: YouTube must look beyond engagement to consider what is ultimately good for the viewer, and by extension, society. While it is possible to argue that YouTube should not interfere with user content preferences, this line of thinking is troubled by the algorithm's use of the activity of other users to make recommendations (Davidson et al., 2010). The proliferation of conspiracy videos, fake news, propaganda and far-right content on YouTube has implications for every user as a metric called ‘co-visitation’ ensures that the choices of individual viewers influence the recommendations served to other users. As YouTube’s primary goal is to hold the viewer’s attention in the pursuit of advertisement revenue, little moral consideration is put into what is being recommended, only that viewers keep watching (Bahara et al., 2019). According to former YouTube engineer Guillame Chaslot, the most provocative content is often the most popular, a problem also encountered on Facebook (Maack, 2019). This means that content which skirts YouTube’s rules has an algorithmic advantage. If this oversight is consciously exploited, it is possible to receive a video recommendation that represents the interests of another; including those intentionally attempting to spread problematic content. The underlying problem with YouTube has less to do with the videos hosted on their platform, and more with the process in which their recommender system chooses what to recommend; the average user is unlikely to search for problematic content directly, but may become radicalized if their tastes are algorithmically nudged over time. While the experience of browsing YouTube might require little cognitive effort, every line of code running behind the scenes is purposeful and driven from some higher objective, explicit or otherwise. That is to say, no choice is accidental. Just as every inaction is an action, a line of code does not only *include*, but also *excludes*. In YouTube’s case, the goal is to get and keep people watching videos (“YouTube Now”, 2012) --regardless of the content. To extrapolate capitalist imperatives from this objective, YouTube’s endgame, like most, is to generate profit for their shareholders.

Many scholars have raised warning flags regarding personalized algorithms, suggesting that such systems could even negatively impact democracy (Sunstein, 2004). More recently, the internet activist Eli Pariser described the cocoon of personalized content that surrounds internet users as a ‘filter bubble’, a space that is both difficult to study and hard to escape (Pariser, 2011). Others critique algorithms for their inevitable codification of social ills like sexism and racism (Buolamwini, 2016; Noble, 2018). Could YouTube’s video recommendation algorithm bear similar flaws? Shifting YouTube’s reputation from being a harmless entertainment source to a complicit radicalizer, two major studies have found similar conclusions regarding the influence the system has over human thought and behaviour, particularly concerning a new breed of internet-literate extremism: the alt-right (Lewis, 2018; O’Callaghan et al., 2014). After the unexpected outcome of the 2016 election in the United States, some believed YouTube’s influence over users deserved a closer look. The results provided by a 2016 study by the British news organization *The Guardian* raised concern about responsible platforming: YouTube was overwhelmingly promoting videos that supported one candidate, Donald Trump, alongside a chef's salad of clickbait, intentionally misleading content, and conspiracy theories (Lewis & McCormick, 2018). While some fingers have been pointed at Russian troll farms in the creation and promotion of some videos, many questions are rightfully directed at the systems YouTube uses to help shape seemingly innocent decisions about the content users consume.

To better understand the psychological biases that influence human information preference, theory can be used to explore how information systems are designed and for what purpose. Through a media theory lens, it is possible to think about how the algorithm as an information conduit or medium that informs audiences as much as audiences influence their design. The revolutionary effect of media as information structures on populations is central to my case for a closer examination of the algorithm as an information distributor. According to Marshall McLuhan, the effects of any single piece of media has significantly less influence than the medium it is carried on:

> *ABC News* host Robert Moore: When you say the medium is the message does that leave any room at all for criticism of individual, say, television programs?
>
> Professor Marshall McLuhan: Or content. You see, it doesn't much matter what you say on the telephone. The telephone as a service is a huge environment, and that is the medium. And the environment affects everybody. What you say on the telephone affects very few, and the same with radio or any other medium. What you print is nothing compared to the effect of the printed word. The printed word sets up a paradigm, a structure of awareness which affects everybody in very very drastic ways and it doesn't very much matter what you print as long as you go on with that form of activity. (ABC News (Australia), 1977).

As theory provides a way to apply a framework to any given situation, plugging in the relevant elements help theorize future outcomes. In this case, McLuhan's comments regarding media environments can be adjusted to begin thinking about recommender systems as information distributors and the effects they may have on media consumers.

Currently, evidence suggests that YouTube's recommendation system is complicit in radicalizing viewers (Bahara et al., 2019; Ribeiro et al., 2019; Roose, 2019), and had some level of influence over the 2016 American election (Lewis & McCormick, 2018). With over 1 billion users (around a third of internet users), YouTube's reach comes with a responsibility for their growing audience (YouTube for Press, n.d.). If users are presented with mis- or disinformation by the algorithm as a result of what similar viewers are watching, recommendations can quickly pull them into a dangerous information cocoon as each viewing compounds the effect, strengthening the co-visitation bond and causing the watching of a video to be interpreted as a signal asking for more of the same. Surrounded by content that presents an alternative reality, this environment has the potential to negatively influence how people make informed choices, including voting. While *The Guardian* was able to study YouTube's role in the 2016 U.S. election, the work done here will extend that investigation to a Canadian context. One of the only studies looking at YouTube's recommender system in a Canadian context was conducted by Fenwick McKelvey and Luciano Frizzera (2019). Frizzera is a graduate from the University of Alberta's Humanities Computing program. In their research, the authors investigated the political discoverability of four candidates in the 2018 Ontario provincial election as they appear in YouTube's recommendations. Like my own research, McKelvey and Frizzera made use of Guillaume Chaslot's open-source software to collect recommendation data on all four candidates. While the researchers found that extremist content was not typically found in the recommendations, they did find bias in the way that the candidates were represented. For example, the authors did find a trend towards negative videos about Liberal leader Kathleen Wynne and positive videos about Progressive Conservative leader Doug Ford (McKelvey & Frizzera, 2019). These findings suggest that YouTube's algorithm could have an impact on Canadian politics as well. To study whether the platform's influence had an impact on Canada's 2019 Federal election, research will be completed to investigate the portrayal of the election's three primary candidates in YouTube's recommendations, including the nature of the sources that are being promoted alongside them. This thesis will analyze approximately 3,000 thousand video recommendations as retrieved from searches for Andrew Scheer, Jagmeet Singh, and Justin Trudeau to discover if there are biases in how each candidate is portrayed by YouTube's algorithm.

**Rationale**

My initial draw to recommender systems was a previous interest in algorithmic influence in Google's search engine. Having been exposed to Marshall McLuhan's thought-provoking ideas regarding the effects of media on society through his short book *The Medium is the Message*, I quickly became fascinated by notions of systems and their influence on information consumption. As my studies in Library and Information Studies began, I became increasingly intrigued by the hidden complexities of information retrieval on the internet. As Google was pervasive to the point of becoming a verb, I wondered where the concern in my circles was over the corporate monopoly of the largest and greatest information portal in human history. Google is often the default search engine used in public libraries around the world and is globally one of the most popular search engines used overall. Given the popularity of online video, it becomes obvious to extend this line of inquiry to YouTube.

YouTube is a service that I regularly enjoy. On an almost daily basis, I consume content from a variety of creators, and as a visual learner, I use YouTube if I want to learn how to do something new. Like many others, I found this platform attractive for its ever-expanding catalogue and its function as a free entertainment source. Most of all, to a further extent than most media formats before it, YouTube has the combination of being both widely accessible and not curated to serve a general audience. Thus, it can serve every niche interest, free from the constricting influence of traditional gatekeepers of other media while still reaching large audiences. However, as discussed in a later chapter, YouTube's strengths can also be recontextualized as its weaknesses, and vice-versa.

Although there are many aspects of this study that have interested me for years, the true catalyst for this research project was born from the aforementioned investigative exposé published by the British news outlet *The Guardian* on February 2nd, 2018. Titled "'Fiction is outperforming reality': how YouTube's algorithm distorts truth," the article alleges that YouTube may have been complicit in getting Donald Trump elected as president of the United States of America in 2016. These claims are backed up with research, for authors Paul Lewis and Erin McCormick teamed up with an ex-Google employee named Guillaume Chaslot to complete the analysis themselves. Noticing a worrying trend in the types of videos that were promoted by the largest video hosting website in human history, Chaslot used his experience and knowledge of the industry to build a program called YouTube-explore, a piece of software that could be used to track and collate the videos that YouTube was recommending in their 'up-next' sidebar. Concerned with technological influence on the 2016 American election, Chaslot recorded data in the months beforehand, focusing on the two major candidates, Donald Trump and Hillary Clinton. Later, the database was handed over to Lewis and McCormick, who used the information to uncover startling truths regarding which videos were featured more than others. In their findings, the authors reveal that the overwhelming majority of partisan videos related to the candidates recommended by the recommendation algorithm at that time using the keywords 'Trump' and 'Clinton' bore anti-Clinton or pro-Trump sentiments. It was not only the content of the videos that they found surprising, but rather the quantity that was being pushed by the supposedly 'neutral' system. As with many other American headlines, the first thing that came to mind when I read that article was 'Could that happen in Canada?' And, with a federal election just a year away, I had a thesis topic.

## Objective

At its core, my research examines the extent to which YouTube's recommendation algorithm unequally promotes videos of the candidates in the 2019 Canadian Federal Election. Although my thesis question is direct and focuses on specific examples, the underlying probe is a larger question about systems: How do media structures inform user's experience of content? To the Canadian media theorist Marshall McLuhan, this is the heart of his famous phrase "the medium is the message" (Fiore & McLuhan, 1967), which is perhaps more clearly put with his other words: "You shape your tools, and they shape you. It's a loop. You start out a consumer, and you wind up consumed" (McLaughlin, 2003). Plainly put, this research seeks to extend research completed by *The Guardian* into a Canadian context, as well as to consider the ways the recommendation algorithm as a medium influences the process of experiencing content on YouTube. Using Guillame Chaslot's exploratory software to probe YouTube's recommendations, I will determine how and to what extent the system recommends videos related to one candidate over the other. The secondary objective will consider the sources of all videos captured in order to assess what is being recommended alongside the candidate videos.

To meet these goals, the database produced by Chaslot's software will undergo coding and subsequent quantitative analysis. Based only on personal speculation, it is hypothesized that there will be no significant findings regarding the promotion of one candidate over the other, at least in part due to the fact that Canada's relatively small population likely results in a much smaller pool of viewers interested in Canadian political content, in turn resulting in fewer videos related to Canadian political leaders. In addition, it is noted that the time of collection is much further away from the election than *The Guardian*'s research was from their own, and perhaps more significantly, YouTube's recent move towards removing misleading content from its recommendation services may have significant impact on my research (Wong & Levin, 2019). Finally, it will also be suggested that the overall number of videos (including those unassociated with Canadian political candidates) with obvious right-leaning bias will vastly outnumber the videos with obvious left-leaning bias, much in part due to the existence of right-wing bubbles as cited in other literature (O'Callaghan et al., 2014). After YouTube's recent crackdown, conspiracy theory videos should be much less likely to appear (Wong & Levin, 2019). Although my own human bias will influence my work, care will be taken to avoid any kind of motivated reasoning in my methodological approach, something YouTube accused *The Guardian* of doing in their research: "Our only conclusion is that *The Guardian* is attempting to shoehorn research, data, and their incorrect conclusions into a common narrative about the role of technology in last year's election. The reality of how our systems work, however, simply doesn't support that premise" (Lewis & McCormick, 2018). As I unpack how YouTube's systems work, I am committed to pursuing my research questions rigorously, regardless of the common narratives surrounding them.

## Justification

There are several reasons why this topic is worth studying. Recommender system algorithms, misinformation, YouTube, and Canadian politics are important in their own way, each having a unique significance on society. As a whole, the work done here has potential implications on both the trajectory of ethically minded algorithmic design and the cyber-dimension of politics in Canada.

### *Algorithms and Recommender Systems*

In the most basic terms, an algorithm is a complex set of instructions designed to meet a goal or goals often related to automation (Corman, 2001). The type of algorithm discussed in this thesis relates directly to heuristics designed to facilitate information retrieval. Widespread use of these systems has grown rapidly since the inception of the internet, and the surge in their use for information filtering purposes has understandably been met with some degree of concern from both academic and media spheres. Using some of the most powerful systems in the world as examples, it is a fact that algorithms have the power to influence human thought processes, and in the case of political influence networks, are even designed to do so (Murthy et al., 2016). Considering the ways in which personalization fragments audiences, algorithms could also impact democracy (Sunstein, 2004), a fact that becomes increasingly important worldwide as the percentage of people with internet access continues to rise every day. Not all algorithms necessarily operate on the internet, and not all algorithms are inherently negative. However, the automation of information retrieval must be critiqued because the ethical principles held by many human information gatekeepers are rarely coded into the systems we have used to outsource retrieval processes. In many ways taking on a readers' advisory role, recommender systems must be designed with a commitment to responsible information distribution: When people are fed a steady diet of misinformation, they may become accustomed to the taste (Tufekci, 2018).

*Misinformation*

Outside of algorithms, misinformation on its own has a tremendous impact on society, including affecting the outcome of elections (Howard et al., 2018). As such, library and information studies (LIS) professionals, researchers, and others have an ethical obligation to expose efforts to algorithmically amplify what is now known as 'fake news'. The importance of this responsibility is highlighted by the American Library Association's core value of social responsibility, which includes "ameliorating or solving the critical problems of society" ("Mission, priority areas", n.d.). While the ALA stresses the importance of access to all perspectives in pursuit of this goal, this is tempered by their resolution on disinformation, in which they declare that "inaccurate information, distortions of truth, excessive limitations on access to information and the removal or destruction of information in the public domain are anathema to the ethos of librarianship and to the functioning of a healthy democracy" ("Resolution on disinformation", 2005). Specifically speaking to governmental efforts to influence public opinion using 'mainstream' media, the ALA's resolutions appear outdated and do not apply perfectly to the propagation of misinformation by YouTube's recommendation algorithm. However, these ideas have been explored by library professionals elsewhere: Sullivan (2018) argues that current LIS approaches to the problem of fake news are outdated, as the focus has been put on the accuracy of information, rather than the effects that misinformation can have on the human mind. It is important to make two distinctions at the outset: Core values of free speech and access to information can still be upheld as I do not advocate for the removal of purely misinformative content, and private companies like YouTube remain free to host or remove legal content as they please. My qualms are with systems that blindly promote sources regardless of factual accuracy, posing a significant threat to ideals of an informed public.

*YouTube*

Understanding YouTube is equally important. It is the largest video-hosting website in human history: YouTube has over 1 billion users, "meaning almost one-third of the internet". On their press page, they boast that "over 1.9 Billion logged-in users visit YouTube each month" and daily, "people watch over a billion hours of video" (YouTube for Press. n.d.). Munger & Philips (2019) claim that YouTube is the most used social network in the United States, while Gruzd et al. demonstrate that YouTube is the second-most used social media platform used in Canada. As an online tool for the masses, YouTube has unique affordances for creators as well: Due to the lack of more traditional gatekeepers and the ever-dropping cost-barrier to video equipment, it is increasingly affordable and easy to produce content for YouTube. Regardless of subject matter, videos have the potential to reach large, global audiences on the platform. If content is promoted by the recommendation algorithm, YouTube is in some ways conducting marketing on the user's behalf. Needless to say, YouTube is a gargantuan force on the internet, and certainly one that deserves an appropriate level of scrutiny for the societal influence it holds.

Particularly interesting to me is the false notion that YouTube is only an entertainment platform. While it certainly can be used for entertainment, the ways people use the service seem to contradict that common assumption. For example, a 2018 PEW research study on the platform found that 86% of adults in the United States said that YouTube was at least somewhat important for 'figuring out how to do things they hadn't done before'. On top of that, 53% said that the website was at least somewhat important when it comes to 'understanding things that are happening in the world' (Smith et al., 2018). From this study alone, it is possible to see that YouTube isn't just used for entertainment, but also exists as a learning platform. One can imagine that younger generations more comfortable with technology will use YouTube as a non-traditional educational resource even more than they do today. In any case, YouTube's role is not only as an entertainment provider. If responsible education is a priority to researchers, then it is necessary to scrutinize how content is being recommended to users on YouTube.

While YouTube as a platform is important to study because of its use, the algorithm at the heart of the system deserves research in particular. Recommended videos make up over 70% of all of YouTube's daily traffic (Zhou et al., 2010). In other words, the recommended bar on the right-hand side of all videos with the automatically playing 'up-next' feature has significantly more impact on views than the original video that people begin to watch in the first place. This usage is reflected by YouTube's default landing page, which is the 'Home' tab featuring suggested videos interspersed with videos from channels users are subscribed to, as opposed to the 'Subscriptions' tab, which, as the name suggests, contains only videos from channels users are subscribed to. Semantically, one could even argue that since YouTube's Terms of Service should hypothetically eliminate any content that the company does not want on their platform, use of the word 'recommend' could signal endorsement of the featured videos (Schmitt, 2018). By creating a queue of content selected for individual users, YouTube has in some ways implicitly conveyed an approval of the content in those videos, beyond the assumed approval of continued existence on the platform.

The kinds of audiences that the algorithm is cultivating on YouTube is another reason this topic needs further investigation. As YouTube itself puts it, "The algorithm follows the audience”, meaning recommendations are largely based on suppositions of what people want to watch (Lewis, 2018). As innocent as this design philosophy sounds, it can help form dangerous bubbles of insular thought, in extreme cases, leading to radicalization. By automatically amplifying personal interest, the recommendation system can coax occasional viewers of ‘misinformational’ videos into consuming a steady diet of fake news or other harmful content. Although a separate conversation needs to be had on the alarming growth of the alt-right presence on YouTube, the system that perpetuates these information cocoons deserves closer scrutiny. After finding evidence of a YouTube-created right-wing filter bubble, researchers Jonas Kaiser and Adrian Rauchfleisch (2018) point towards the Thomas Theorem, a sociological theory that suggests that situations are real as far as we define them as real. Applying this thinking to the algorithm, the authors posit that if algorithms define situations as real, they are real in their consequences. That is to say, the information environments that YouTube’s algorithm fosters could incentivize the creation of videos in those communities; as such, there is a circular relationship between content and the system that distributes it.

*Canadian Politics*

The 2019 federal election in Canada also deserves close academic study as well, much in part to the current political climate: In the lead up to the 2019 election, Prime Minister Justin Trudeau characterized it as "perhaps what will be the most divisive and negative and nasty political campaign in Canada's history", citing the rise of nationalism and populism as part of the problem (Perkel, 2018). While it is obvious that the election is of great import to Canadians, what is particularly interesting to me are the factors that came into play in electing our next leader, including the possibility of algorithmic bias or the targeting of human bias by malicious foreign entities. The 2018 National Cyber Threat Assessment by the Canadian Centre for Cyber Security issues a strong warning regarding the legitimate threat facing Canadians, stating: “In the coming year, we anticipate state-sponsored cyber threat actors will attempt to advance their national strategic objectives by targeting Canadians’ opinions through malicious online influence activity” (National cyber threat, 2018, p.3). This caution is not unwarranted, for a recent report by Oxford University’s Computational Propaganda Project found evidence of external influence on the 2016 American election. The report reveals that Russia’s Internet Research Agency (IRA) used every major social media tool to help Trump get elected, including YouTube: “The IRA’s heavy use of links to YouTube videos leaves little doubt of the IRA’s interest in leveraging Google’s video platform to target and manipulate US audiences” (Howard et al., 2018, p. 8). External influence is being exerted on Canadian public opinion using social media and YouTube is a platform on which such attacks have occurred in the past. Putting the two together, it is reasonable to conclude that it is vital to Canadian democracy to study how YouTube's algorithm recommends Canadian-focused content.

Regardless, political influence by foreign states do not absolve YouTube itself of any fault, as the design of their systems may enhance the effectiveness of disinformation campaigns. Or, as the authors of "Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems" put it: "YouTube's status as the most popular video sharing platform means that it is especially useful to political extremists" (O'Callaghan et al., 2014, p. 460). Prioritizing revenue over any other metric, YouTube's reckless pursuit of views has resulted in endless plains of clickbait, a right-wing ideological bubble (Kaiser, 2018; O'Callaghan, 2014), and disproportionate promotion of divisive content and conspiracy videos (Lewis & McCormick, 2018). While cited evidence points towards a generous spread of research in this area, there is limited research on YouTube's recommendation system in a Canadian context. Given the existing literature on algorithms, YouTube, and Canadian politics, I believe that this subject is worthy of study.

**Conclusion**

This thesis is organized as follows. This first chapter serves as an introduction to the research as a whole, including my rationale and objectives. Chapter 2 will provide necessary context to the subject explored, including algorithms, recommender systems, YouTube, Canadian issues, and questionable or 'borderline' content on YouTube. In Chapter 3 I review the peer-reviewed literature related to information retrieval and algorithms, focusing on issues like recommender systems, YouTube, and ethics. Chapter 4 outlines the theoretical frameworks presented earlier, honing in on psychological and media theories associated with my research context. In Chapter 5, I begin outlining the original research done here by way of my methodology, followed by a chapter dedicated to revealing the results and discussing those findings in depth. Finally, in Chapter 7 I identify the major issues faced by the platform today and explore potential solutions to those problems. In addition, I target directions for future research, evaluate the software used to collect the data, and summarize the thesis.

The following chapter works to provide the necessary context to the complex question of YouTube, Canadian politics, and the conversations currently taking place around the internet's role in radicalization. YouTube is discussed within a Canadian context, specifically focusing on the 2019 federal election. Serving as an introduction to the website, the layout of the video platform is laid out, and their policies are also examined to highlight gaps between the company's publicly stated goals and the effects its recommendation algorithm is having on extremism.

# CHAPTER TWO: CONTEXT

## Introduction

In order to provide a complete perspective on this topic, a comprehensive contextual understanding is needed. Beginning with an overview of the Canadian political landscape, this chapter moves on to consider the part YouTube may have played in the 2019 Federal election. This potential influence is then weighed against YouTube's publicly stated goals as well as the policies they have put in place to help achieve those objectives. Finally, I explore the potential repercussions of the problematic content on the platform that comes close to breaking the community guidelines. Overall, this chapter provides the reader with the necessary information to better understand the issues surrounding my central thesis questions.

## Canadian Context

### *Politics*

On October 15th, 2015, Liberal Party leader Justin Trudeau won 184 seats across Canada, forming a majority government. The campaign that preceded that election was one of the longest in Canadian history, with only two election campaigns being longer: 1867 and 1872 ("Imminent federal election", 2015). It also saw the highest voter turnout (84.3%) since the 1993 federal election (Swartz, 2015). In 2019, the leaders of the three primary parties were Justin Trudeau for the Liberal Party of Canada, Jagmeet Singh for the New Democratic Party (NDP), and Andrew Scheer for the Conservative Party of Canada (CPC).

On August 19th, 2018, Justin Trudeau announced that he would be running for re-election in 2019 ("Trudeau formally announces", 2018) To briefly summarize some of his policy positions, Trudeau is pro-choice, identifies as a feminist, and has advocated for electoral reform in his initial election campaign, a goal that was abandoned in early 2017 following the decision of a parliamentary committee (Kilpatrick, 2014; Saul, 2015; Smith, 2017). Trudeau also led the way to a nation-wide decriminalization of cannabis in 2018. In May 2017, Andrew Scheer was elected leader of the Conservative Party of Canada, edging out opponent Maxime Bernier. In terms of political ideology, Scheer positions himself as being staunchly anti-Carbon tax, pro-life, but nevertheless "absolutely" a feminist (Boesveld, 2017; Hutchins, 2017). If he were elected Prime Minister, Scheer stated that he would not pursue a free trade deal with China and would balance the budget within two years of forming government (Gilmore, 2018; Hutchins, 2017). Finally, Jagmeet Singh was elected as leader of the NDP on October 1st, 2017 and refers to himself as a social democrat (Jones, 2017). He is in favor of decriminalising possession of all drugs, supports raising the minimum wage to $15 per hour, and advocates the removal of tax deductions for the highest-income earners (Britneff, 2017; theJagmeetSingh, 2018; Woo, 2017). Although only these three candidates will be included in this study, there are six major parties holding seats in the House of Commons, including the Bloc Québécois, the Green party, and Maxime Bernier's newly formed People's Party of Canada (PPC). As all three of these less-popular parties currently hold a combined 12 seats (out of 338), they do not inspire a significant enough amount of YouTube content to be considered in this study.

Although Trudeau ended 2018 with an advantage popularity-wise due to his relatively unmarred leadership, the tides began to shift in the following year in response to controversy. In February 2019, *The Globe and Mail* reported that the Prime Minister's office had attempted to influence an ongoing criminal case related to the Montreal-based construction company SNC-Lavalin (Fife, Chase, & Fine, 2019). On June 4th, 2019, average projections from *CBC*'s *Canada Votes* poll tracker had the CPC in the lead at 34.4%, the Liberal Party following with 30.8%, and the NDP trailing with 16.1% support (Grenier, 2019). As an addendum, the 2019 election saw the re-election of Justin Trudeau, giving the Liberal Party of Canada a minority government.

*Cyber Threats*

In a January 2019 speech, Canadian Public Safety Minister Ralph Goodale expressed the need to protect democratic institutions from hostile state activity in the face of the upcoming election (McKenna, 2019). Speaking in regard to Russian influence of the American election, Goodale said the following: "We would be foolish and naive not to conclude that something similar might be attempted with respect to Canada" (Guignard, 2019). This sentiment has been echoed in the 2018 National Cyber Threat Assessment by the Canadian Centre for Cyber Security, where the danger is explicitly forewarned: "We assess that in 2019, state-sponsored cyber threat actors will very likely attempt to advance their national strategic objectives by targeting Canadians' opinions through malicious online influence activity" (p. 14). The report goes on to outline the existence of 'troll farms' that are designed to make information 'more compelling and distracting', 'create fraudulent news', and 'promote extreme opinions'. Among the highlighted targets in the report are the opinions of Canadians, which 'cyber threat actors' seek to sway with the use of botnets, many of which can be described as large networks of automated social media accounts. According to the authors, the objective of these troll farms is to negatively influence democratic processes. A particularly relevant quote in the report is a statement regarding Russian influence on Canadian issues:

> A recent study revealed Twitter accounts connected to the Russian-based Internet Research Agency that promoted divisive and inflammatory content before the 2016 United States presidential election also tweeted about events in Canada. About 8,000 of over 3 million archived tweets from the now-deleted accounts focused on Canadian issues, including the May 2016 fire in Fort McMurray, the January 2017 Québec City mosque shooting, and the increase in asylum-seeker border crossings in summer 2017. The Russian trolls attempted to create confusion by inserting false information into online discussions and exacerbating existing differences of opinion. This case demonstrates that Canadian social media users can be exposed to foreign malicious influence activity. (Canadian Centre for Cyber Security, 2018, p. 15)

When politicians like Goodale warn of the potential for influence by foreign actors on Canadian democracy, studies like these reveal that they are far behind the stark reality of the situation: Russia has already attempted to digitally influence Canadian democracy and will almost certainly do so again. According to the *National Observer*, for example, the hashtag campaign #trudeaumustgo found its way onto Twitter's trending lists, giving many the impression that there was widespread discontent with the Prime Minister prior to the election. After analysis, the *National Observer* found that "a non-trivial number of tweets came from accounts that were either brand new, displayed signs of automation or both" (Orr, 2019).

*Potential Implications*

In Ralph Goodale's speech on foreign influence over Canadian elections, he also warned of a rise in right-wing extremism in the country, specifically from neo-Nazis and white nationalist groups (McKenna, 2019). According to an environmental scan conducted by Dr. Barbara Perry and Ryan Scrivens for Public Safety Canada, there are over 300 white supremacist and neo-Nazi groups across Canada (Perry, 2015). While this number does not account for lone wolf extremists, it does present a general overview of the situation. In the 2017 Public Report on the Terrorist Threat to Canada by Public Safety Canada, right-wing extremism was described as a 'growing concern'. Regarding dissemination of ideology, the report states:

> In Canada, individuals who hold extreme right-wing views are predominantly active online, leveraging chat forums and online networks. Rather than openly promoting outright violence, those holding extreme right-wing views often attempt to create an online culture of fear, hatred and mistrust by exploiting real

> or imagined concerns when addressing an online audience. (2017 Public Report, 2017)

Several prominent far-right figures online are Canadians by birth. This circle includes the white supremacist Stefan Molyneux, conservative comedian Steven Crowder, and the alt-right talk show hosts Jean-Francois Gariepy and Andy Warski, all of whom have a relatively large following on YouTube. As of August 2019, Molyneux has almost a million subscribers, while Crowder has over 4 million. Other notables are former members of the Canadian alternative media group *Rebel Media*, including Lauren Southern, former Toronto mayoral candidate Faith Goldy, and the founder of the far-right men's group the 'Proud Boys', Gavin McInnes. For many of these figures, a primary issue is immigration. Although less focused on typical right-wing talking points, Toronto professor and skeptic Jordan Peterson is also Canadian and often interacts with other right-wing personalities on YouTube. In 2016, Peterson released a series of lectures on YouTube vocalizing his position against Bill C-16, a piece of legislation designed to add transgender individuals under an umbrella of protected groups (Winsa, 2017).

In the offline world, the propagation of far right ideology[1] online can have deadly consequences. In the United States, right-wing extremism manifested in outright violence and the murder of Heather Heyer at the 'Unite the Right' rally in Charlottesville, Virginia in August of 2017. Earlier that same year, right-wing violence erupted in Canada: On January 29th, 2017, an individual[2] entered a mosque in Quebec City and opened fire on worshippers (Kassam and Lartey, 2017). The shooter killed six and injured 19 others. In the investigation following the attack, it was hypothesized that several prominent right-wing figures had a role in radicalizing the young man, including Ben Shapiro, Alex Jones, and Tim Gionet (aka 'Baked Alaska') and others as determined by the shooter's internet browsing history (Riga, 2018). In the months leading up to the shooting, he had conducted a number of searches related to Donald Trump: 417 on Twitter, 337 on Google Search, and 63 on YouTube. Also released was a list of the Twitter feeds the shooter was visiting that month, including Canada's own Gavin McInnes and Stefan Molyneux. The example of the Quebec City shooting serves to underline the reality that far-right extremism is a present threat in Canada. While it is not known whether YouTube had a role in radicalizing the shooter, many of the profiles he frequented on Twitter had or continue to have a strong presence on the video platform. As the hateful online culture that festers on YouTube is often aided by an ethically blind recommendation system, many are questioning the extent to which YouTube itself is responsible for promoting hate speech, an act which may later result in physical violence. The problem is complex: While it is unlikely that any engineer or designer ever purposely set out to create a system that damages society, should a corporation be held responsible if their systems do so regardless of intent? And on the other side of the coin, does YouTube as a private company have a responsibility towards upholding free speech?

## YouTube

### *Layout*

When referring to recommended videos on YouTube, I am primarily concerned with a specific type of video. To diffuse potential misunderstandings, the following discussion delineates the six areas where the recommendation algorithm impacts the videos shown. The six areas include search results, the ‘suggested videos’ stream, the home page, the trending stream, recommendations, and through notifications (Brown, 2018). Given the history of changes made to the various sections of the website, different versions of the same algorithm (or altered ranking models) are used for different purposes, namely between the search and sidebar recommendations (Chen et al., 2019).

While direct information retrieval based on keyword searching does not generate as much traffic as recommended content on YouTube, there is a search system that works to find relevant video content using the search bar (Figure 2.1). According to YouTube, “Videos are ranked based on a variety of factors including how well the title, description, and video content match the viewer’s query” (“Search and Discovery”, n.d.). They add two qualifiers to this statement: that user engagement is considered when ranking the results, and that the results list is not simply a list of the most-viewed videos related to that keyword. Given the proprietary nature of YouTube’s business secrets, researchers must rely upon user-oriented public documentation to investigate how it works. Additionally, papers published by Google employees provide a more technical understanding of how the algorithm has been designed in the past (Chen, 2019; Covington et al., 2016; Davidson et al., 2010). Otherwise, some information can be gleaned from Google’s API, which allows for the study of video and channel statistics.

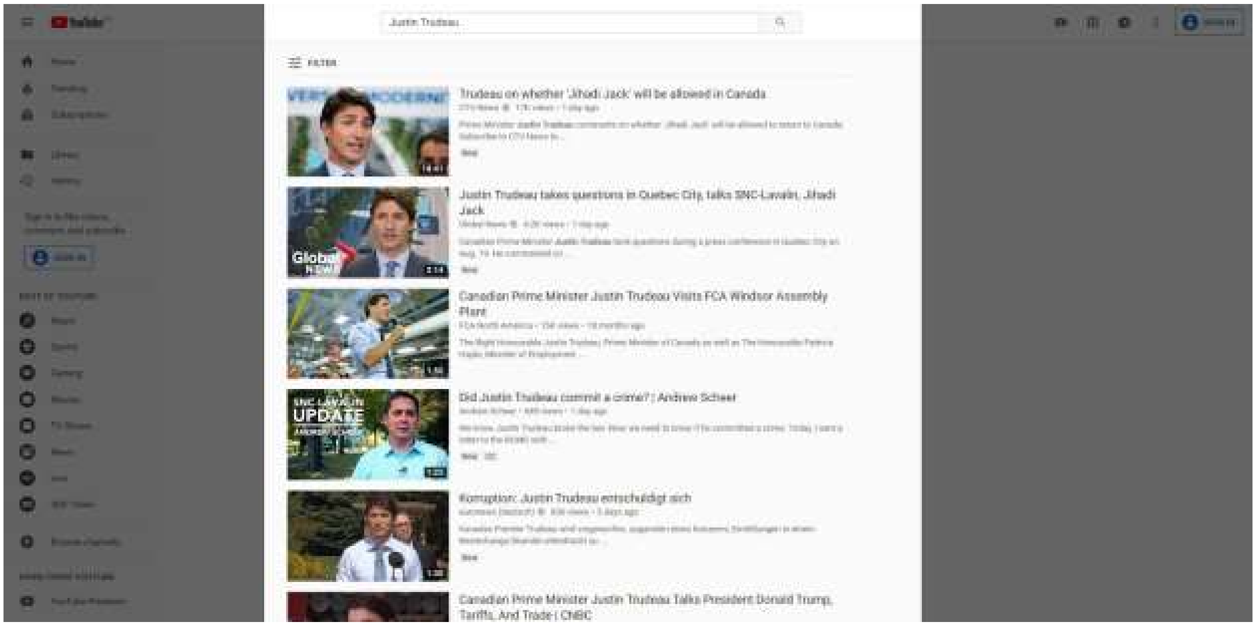



*Figure 2.1: Search Results*

Trending videos are displayed in an algorithmically curated section of the website where popular videos of the day are listed in four categories: Music, Gaming, News, and Movies (Figure 2.2). Right now, Trending videos can be found in the upper left-hand sidebar, between Home and Subscriptions. Trending videos prioritize a broad audience, aiming to "surface videos that a wide range of viewers will appreciate", and therefore are not personalized. Every fifteen minutes, the list is updated. Featured or 'surfaced' videos are selected based on meeting four objectives ("Search and Discovery", n.d.):

1. Are appealing to a wide range of viewers
2. Are not misleading, clickbaity or sensational
3. Capture the breadth of what's happening on YouTube and in the world
4. Ideally, are surprising or novel

To further narrow down which videos are selected, YouTube uses four signals to detect what might be the best fit for the list:

1. View count
2. The rate of growth in views
3. Where views are coming from (including outside of YouTube)
4. The age of the video

These signals are then combined to find videos to populate the list, which YouTube adamantly states cannot be influenced by underhanded means: "YouTube does not accept payment for placement on Trending. We do not include views from YouTube ads in selecting videos for Trending. YouTube does not favor specific creators" ("Trending", n.d.). While trending videos will not be included for analysis in my study, YouTube's transparency here when it comes to what they value is telling when investigating the other types of recommended videos. Given that the intent of YouTube's resources is to help creators succeed on their platform, the language used is vague and no effort is made to explain terminology in more detail. Vague terms like 'clickbaity', 'sensational', and 'surprising', are not expanded upon in these pages and some are only defined in a more concrete fashion elsewhere. For example, on the official YouTube blog, clickbait is loosely defined as content that makes use of misleading titles and descriptions ("Continuing our work", 2019). Additionally, there is no documentation suggesting how 'surprising' or 'novel' are measured; it seems likely that YouTube is intentionally vague about these terms as they are dependent on what are popular topics at any given time. In this thesis, 'clickbait' is regarded as consisting of provocative headlines or thumbnails that do not accurately represent the content of the item. Designed to attract viewership, clickbait overlooks accurate information for click-through rates.

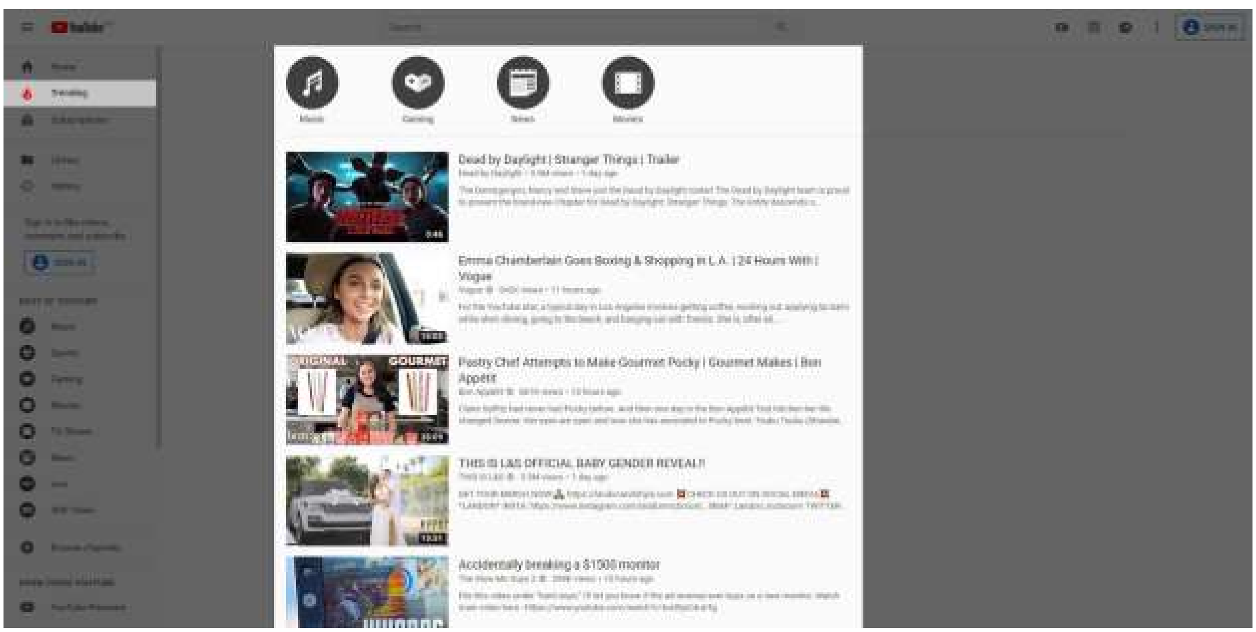



*Figure 2.2: Trending Videos Panel*

'Suggested for You' videos are selected based on a user's personal viewing history, related topics, and the video currently being watched (YouTube Creators, 2017 August 30). They can be found interspersed with the other recommended videos in the recommendation sidebar accompanying all videos (Figure 2.3). Even when a user is not signed in, they may still receive personalized suggestions thanks to demographic signals.

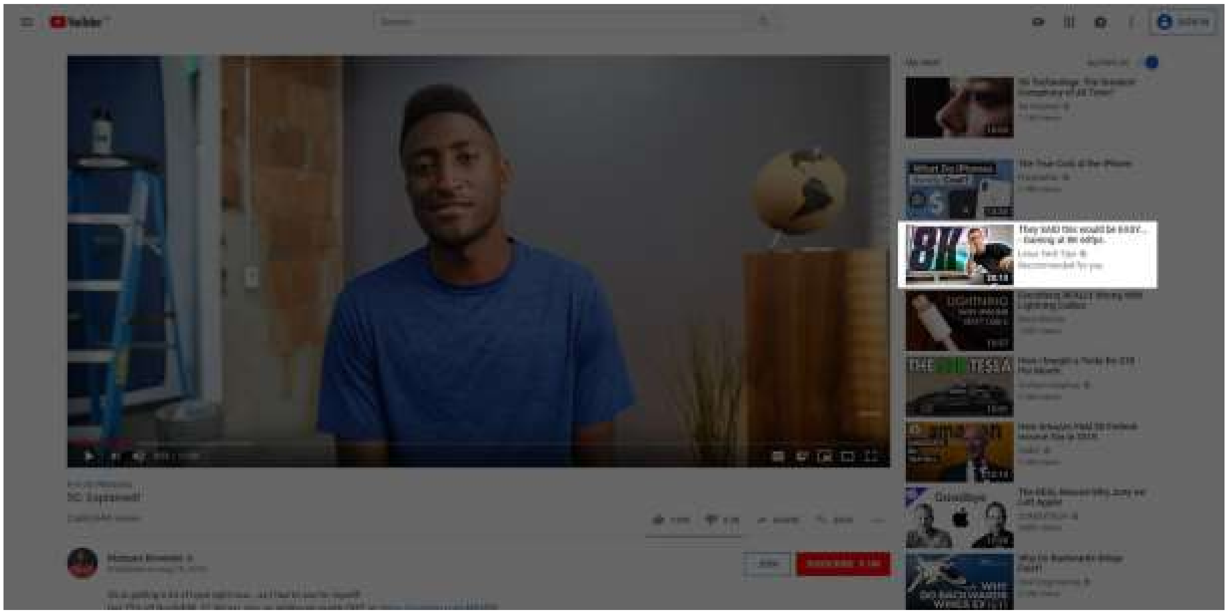


*Figure 2.3: 'Suggested for You' Video in the Recommendation Sidebar*

Recommended videos are items that can be found in the right-hand recommendation sidebar (Figure 2.4). Videos suggested on the Home page make use of the same algorithm. While Suggested for You videos can also be found in the recommendation sidebar, they will be excluded in this study due to their non-generalizable nature. The top-most video in this sidebar is set apart from the rest and labeled as 'Up-Next', a designation that indicates which video will automatically begin playing after the current video has finished. Accompanying the up-next video is an 'Autoplay' toggle that controls whether it will play automatically. By default, this feature is always set to the "on" position.

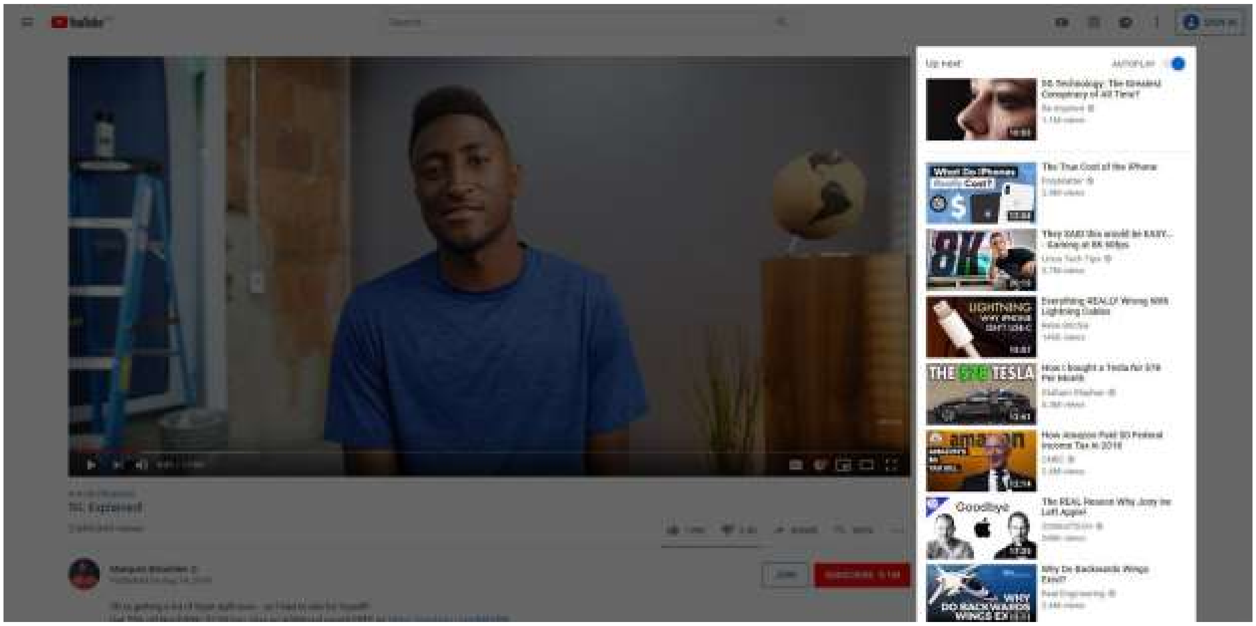



*Figure 2.4: Recommended Videos Sidebar*

*Values*

As a video-hosting platform, YouTube's reach is unrivalled. Given its popularity and its relevance to this study, a full understanding of the context surrounding YouTube is needed. First, it must be determined who and what YouTube is for. Is YouTube designed to benefit users, or are users themselves a product used to primarily benefit advertisers and shareholders? When it comes to problematic content that does not explicitly break their community guidelines, is YouTube a platform that upholds free speech, or is it a space for moderated discourse? According to their About page, YouTube is for everyone: "Our mission is to give everyone a voice and show them the world", (YouTube, n.d.). Going on, they list the four essential freedoms that define them:

1. Freedom of Expression
2. Freedom of Information
3. Freedom of Opportunity
4. Freedom to Belong

Understandably, these commitments to freedom are tempered by the company's community guidelines, which in some ways are designed to prevent one user's freedom from inhibiting the freedom of another. YouTube's corporate stance on responsible platforming is very clear: In response to an article questioning the company's delicate balancing act between profitability and moral obligations, YouTube CEO Susan Wojcicki tweeted "My #1 priority is responsibility, even if that comes at the expenses [sic] of growth" (SusanWojcicki, 2019).

*Policy*

To understand what YouTube considers to be responsible, a good starting point is the platform's policies on permitted and disallowed content. Although YouTube's Terms of Service agreement is often mistakenly cited to highlight questionable conduct on the platform, section 6E of that agreement directs readers to the Community Guidelines, a page that outlines the policies set out in order to make YouTube "fun and enjoyable for everyone" ("Policies and Safety", n.d.). Here, there are 11 major areas of policy:

- Nudity or sexual content
- Harmful or dangerous content
- Hateful content
- Violent or graphic content
- Harassment and cyberbullying
- Spam, misleading metadata, and scams
- Threats
- Copyright
- Privacy
- Impersonation
- Child safety

There are also some additional policies regarding the following:

- Vulgar language
- Inactive accounts
- Encouragement of ToS violations
- Age requirements

Of relevance is the section on hateful content. YouTube defines hate speech as content that promotes violence or hatred against individuals based on various attributes of protected groups, including:

- Age
- Disability
- Ethnicity
- Gender
- Nationality
- Race
- Immigration Status
- Religion
- Sex
- Sexual Orientation
- Veteran Status

The company goes on to provide further clarification on this policy by offering other scenarios where this policy may apply. These include dehumanization or implications of lesser status; use of racial, ethnic, religious, or other slurs; and the use of stereotypes that incite or promote hatred. Exceptions include documentaries focused on hate groups and depictions of hatred in a historical context. Determination of which content violates their policies is not typically done by employees, but rather by YouTube community members and automated services. On the page that details how content can be reported, YouTube states that they "rely on YouTube community members to report content that they find inappropriate." Beyond making use of their own users to police content, the platform also launched a Trusted Flagger Program to help regulate the daily influx of videos. This program consists of a set of tools that help selected individuals, government agencies, and non-governmental agencies identify policy-violating content on the platform. These tools include a bulk-flagging tool, prioritized flag reviews, a direct line of communication with YouTube, and the ability to view rationale on previously flagged videos. Also available for NGOs are occasional online training sessions on community guidelines and enforcement processes. To become a trusted flagger, the process differs based on the applicant. If the user belongs to a governmental agency or NGO, YouTube requires active outreach from the organization themselves and considers various factors before allowing them into the program. If an independent individual aspires to take part in the program, they must start by flagging “a large volume of videos with a high rate of accuracy”. Users meeting this criterion are invited to participate by YouTube.

Once a report is received, it is not automatically taken down but instead first reviewed by staff. If the content is deemed to have violated a community guideline, this is communicated with the channel owner. If the channel has not posted restricted content in the past, they will only receive a warning without punishment beyond removal of the rule-breaking content (“Making our strikes”, 2019). After three such strikes, the channel is terminated. Every three months, YouTube releases a transparency report that reveals the volume of channels that have been terminated since the last report (YouTube Community, n.d.). In the period between January and March 2019, over 2.8 million channels were removed. In that time period, 85.1% of the channels were removed for spam, misleading content, or scams. 9.6% was removed for nudity or sexual content, 2.8% for reasons surrounding child safety, and varying amounts of other violations. Only 0.1% of channels were terminated for hateful or abusive content, a strikingly low amount given the broad definitions of hate speech and harmful content outlined in the community guidelines. Another way YouTube cracks down on undesirable content is with the use of automated flagging systems. Of the 8,294,349 videos removed in the period between January and March 2019, the majority (6,372,936) were deleted by an automatic system that uses spam-detection technology to remove videos before they receive any human views (“why flagging matters”). Of the 8 million videos removed in this period, over 75% were removed before they received any views.

*Roles*

Of the values expressed, YouTube’s commitment to freedom of opportunity stands out as being particularly relevant to their algorithm’s design: “We believe everyone should have a chance to be discovered, build a business and succeed on their own terms, and that people—not gatekeepers—decide what’s popular.” By distancing themselves from the process of promotion, YouTube is keen to argue that it is a platform and not a publisher. However, there is a tension between their function as a video-hosting website while simultaneously wanting to police that content (Gillespie, 2010). In other words, YouTube wants control over the content on its website without having to be held responsible for it. While perhaps a somewhat uncharitable interpretation of their values, the argument that viewers are in full control over YouTube’s contents limits YouTube’s responsibilities when it comes to hosting and even promoting questionable content, including conspiracy videos, fake news and hate speech. Instead of taking on the role of a publisher and employing heavy curation, the platform instead shifts blame on the public when issues arise. Later discussion details how YouTube is forced to play gatekeeper when held up to the light of bad press, revealing a preference against moderation until public outcry. Despite this discrepancy, the platform typically has risen to the occasion and has taken responsibility for its reach by removing objectionable content, although arguably more so for advertisers than viewers. That same profit motive underpins the majority of businesses in a capitalist economy, and YouTube is no exception. While repeated examples of the company's efforts to highlight their commitment to responsible platforming can be found scattered throughout this chapter, so too are there repeated examples of how YouTube has failed to uniformly uphold their policies.

In addition, there are also inconsistencies between what YouTube outwardly proclaims as its values and what it prioritizes internally. For example, on February 5th, 2019, YouTube CEO Susan Wojcicki posted a year-end review on the YouTube Creator Blog (Wojcicki). In it, she emphasized the growth users on the platform had experienced over 2018: channels with over one million subscribers doubled, and the number of creators earning five or more figures grew more than 40%. But the trends weren't all positive, as Wojcicki also admitted that a small number of individuals had negatively impacted the entire creator ecosystem, and that changes had to be made to ensure "responsible growth". Despite the public positioning, reports indicate that the company's goals are not as idealistic internally. According to a journalist at *Bloomberg*, three separate employees questioning the recommendation system and its role in pushing videos from the alt-right were told "Don't rock the boat" (Bergen, 2019). One of the interviewed employees went on to state that CEO Wojcicki "never put her fingers on the scale", instead choosing to focus on engagement over responsibility. Additionally, staff were pushed away from proactive measures and told by legal teams that prior knowledge of problematic content undermined their ability to remain a neutral party, and therefore dodge responsibility for the content they hosted. This tension between platform and publisher remains at the core of a discussion on both hosted and recommended content.

*System-Level Remediation*

On a system level, YouTube's parent company Google has acknowledged and partially addressed flaws in the recommendation system of their synonymous search engine. Project Owl was launched in 2017 to combat instances of questionable search suggestions where holocaust denial queries were offered up to users when typing in the incomplete phrase "did the hol" (Sullivan, 2017). As an aside, this problem is not completely resolved as the system design forces Google to be reactive instead of proactive (Albright, 2018). After Project Owl, users could report suggestions for inappropriate content. Retrieval of fake news sources has also been cited as a problem; Google's various solutions have included boosting what they deem 'authoritative content', adding tags from fact-checking websites like Snopes and PolitiFact to applicable content, adding trust indicators to verified news outlets, and restricting news articles from websites that refuse to show their country of origin (England, 2017; Fingas, 2017; Summers, 2017). After 2017's deadly Las Vegas shooting, Google News Top Stories recommended a 4chan thread containing wildly incorrect information about the event for a few hours. In response, Google promised 'algorithmic improvements' (Alvarez, 2017). In 2018, the company launched the Google News Initiative (GNI), a three-year, $300 million-dollar effort to combat fake news. This project focuses on three main objectives:

1. To elevate and strengthen quality journalism
2. To evolve business models to drive sustainable growth
3. To empower news organizations through technological innovation

While GNI's goals are vague, they do represent a massive commitment to ethical platforming. In their own words, "Bad actors often target breaking news on Google platforms, increasing the likelihood that people are exposed to inaccurate content. So we've trained our systems to recognize these events and adjust our signals toward more authoritative content". This official blog post from Google's CBO Philip Schindler goes on to recognize that similar problems are occurring on YouTube, and that they are being addressed by "highlighting relevant content from verified news sources" (Schindler, 2018). The point is that Google has identified and acted on the flaws of its systems in the past--beyond only removing objectionable content. With these moves, Google has arguably assumed the role of publisher and taken responsibility for the content they curate as well as the systems that promote and distribute that information. Here, a differentiation must be made: Google is an indexer and YouTube is a content host. These are two different services with arguably different levels of accountability for the content that can be retrieved on them, as YouTube provides server space for creators while Google Search retrieves, indexes, and ranks links to websites hosted elsewhere. Nevertheless, the evidence provided here demonstrates that Google as a company has taken on a socially responsible role in the past, meaning that it is not inconceivable that the company could adopt a similar philosophy on other platforms they own, including YouTube.

Although YouTube claims to remove content from their service that violates their community guidelines, their content-focused approach to responsible publishing often overlooks system-based solutions, resulting in a reactive and often delayed response to problematic errors. For example, YouTube removed a video about the Parkland shooting in February 2018 that had made its way into the trending section of their website, citing their policies on harassment and bullying. The promotion of the video, which claimed survivor David Hogg was a crisis actor, was seen as a failure of YouTube's automated systems. In a statement, YouTube acknowledged their fault and took responsibility for the video's surfacing: "In 2017, we started rolling out changes to better surface authoritative news sources in search results, particularly around breaking news events. We've seen improvements, but in some circumstances these changes are not working quickly enough" (Herrman, 2018). Just a year later, similar problems persisted: The November 2018 Camp Fire in California was a serious event, leaving 86 dead and nearly 19,000 buildings destroyed ("Camp Fire", 2019). As per usual, the story was subject to wild conspiracy theories on YouTube, alleging the fire was started by the government with the use of lasers. But something else was going on: YouTube's search prediction algorithm was automatically completing searches for "california fire" with suggestions like "california fire conspiracy 2018", and "california fire laser beam" (Haskins, 2018). In other words, this automation blindly pointed viewers towards fake news by using prior engagement signals to determine what users were most likely searching for. While the search algorithm exists separately from the video recommendation algorithm, it is again a reminder of the consequences of over-automation, a reliance on metrics over ethics that allow misinformation to thrive.

To combat misleading content, YouTube decided to add fact-checking links to certain videos in 2018 (Glaser, 2018). On videos discussing topics that tend to attract conspiracy theories, YouTube automatically added Wikipedia and Encyclopedia Britannica articles to provide more authoritative information. 2019 saw more ethical approaches to content hosting, as YouTube added similar links to Wikipedia articles on anti-vaccination videos (O'Donovan, 2019). However, these automated systems sometimes lack the care required to make accurate designations. For example, on April 15th, 2019, the iconic French cathedral Notre-Dame suffered a major fire. On YouTube, live streams of the fire were accidentally accompanied by 'knowledge panels' featuring articles from Encyclopedia Britannica about 9/11 (Paul, 2019).

Similarly, the design of the recommendation algorithm itself repeatedly exposes its own flaws. A June 2019 article by the *New York Times* highlighted how the recommendation system was enabling pedophiles (Fisher & Taub, 2019). After a user uploaded an innocent video of her children playing in a backyard pool, she soon found that the video was receiving undue attention: 400,000 views attested to something strange working behind the scenes. Thanks to the recommendation algorithm's co-visitation metric, the video of her children was being exposed to other users who had sought out videos of prepubescent, partially clothed children. Without intention, the recommendation system had tied innocuous videos of children together by identifying similarities between viewers.

Regardless of effectiveness, YouTube is taking steps towards moderating the videos it hosts, albeit reactively. But where are the proactive algorithmic adjustments? Can automated systems ever replace the ethical sensitivity of human moderation? What will YouTube do to address the continuous flow of harmful content onto their platform? After *The Guardian* published their story outlining how the recommendation algorithm at YouTube was overwhelmingly pushing pro-Trump videos and fake news, the response from the company pointed at flaws in the study design and at their users' viewing habits rather than the algorithms that decided what viewers wanted to watch: "Our search and recommendation systems reflect what people search for, the number of videos available, and the videos people choose to watch on YouTube, that's not a bias towards any particular candidate; that is a reflection of viewer interest (Lewis & McCormick, 2018)." Although YouTube has taken a more content-focused approach to questionable content in the past, recent statements display a renewed commitment to systemic adjustments. In 2019, Google took a stand against conspiracy theory videos and misinformation (Wong & Levin, 2019). In a blog post, YouTube vowed to "begin reducing recommendations of borderline content and content that could misinform users in harmful ways—such as videos promoting a phony miracle cure for a serious illness, claiming the earth is flat, or making blatantly false claims about historic events like 9/11" ("Continuing our work", 2019). As the evidence suggests, repeated incidents in the last few years have forced YouTube to consider how they might redesign their systems to proactively and responsibly moderate content on their platform.

## YouTube & Content Issues

### *Borderline Content*

Before considering potential system-level solutions, the problems facing YouTube require further description. Several high-profile cases serve as a useful gauge as to how YouTube deals with videos that appear to break community guidelines. In July of 2016, staff running the YouTube conservative channel *Prager University* noticed that a number of their videos (according to *PragerU*, this number is now over 100) were placed in the platform's 'restricted mode', a parental-control-like version of the website that filters out non-family friendly content (Wilson, 2018). In addition, some videos were deemed ineligible for advertisements. To PragerU, this was an act of censorship and warranted a lawsuit against YouTube's parent company Google, a claim which was subsequently thrown out of court by a judge in 2018. According to the Judge, "Defendants are private entities who created their own video-sharing social media website and make decisions about whether and how to regulate content that has been uploaded on that website" (Neidig, 2018). On their website, PragerU maintains its position, adding the message "Wondering why our ideas are being suppressed? Us, too" to a list of the banned videos, which include titles like "If You Live in Freedom, Thank the British Empire", "Where Are the Moderate Muslims?", "The World's Most Persecuted Minority: Christians" and "Dangerous People Are Teaching Your Kids" ("Restricted by YouTube", 2018). Another prominent example of punitive measures occurred in August 2018, when YouTube deplatformed the prominent conspiracy theorist Alex Jones. After issuing a third and final strike to his channel *Infowars*, it was permanently removed from the service for hate speech, harassment, and circumvention of enforcement measures (Tobias, 2018).

Sometimes, however, there appears to be some inconsistencies with how YouTube upholds its policies. On April 30th, 2019, *Vox* Journalist Carlos Maza tweeted a compilation of homophobic slurs directed at him by the host of the YouTube channel *StevenCrowder* (gaywonk, 2019). Crowder's YouTube show, *Louder with Crowder*, was used to target Maza's sexual orientation repeatedly. In many of the videos, host Steven Crowder himself wore and promoted a purchasable t-shirt with the text "Socialism is for F*gs" on the front. In response to Maza's tweet, the official Team YouTube Twitter account responded 5 days later, stating that they had conducted an "in-depth review" of the videos in question. YouTube went on to reveal that no action was to be taken: "While we found language that was clearly hurtful, the videos as posted don't violate our policies" (TeamYouTube, 2019). This statement appears to directly contradict their own community guidelines, as YouTube's harassment and cyberbullying policy requires creators to refrain from posting content that makes "hurtful and negative personal comments/videos about another person" ("Harassment and cyberbullying policy", n.d.). Elaborating, they added further comments:

> As an open platform, it's crucial for us to allow everyone–from creators to journalists to late-night TV hosts–to express their opinions w/in the scope of our policies. Opinions can be deeply offensive, but if they don't violate our policies, they'll remain on our site. Even if a video remains on our site, it doesn't mean we endorse/support that viewpoint. (TeamYouTube, 2019)

A day later this rhetoric was amended; now Crowder's channel was demonetized until links to his offensive t-shirts were removed (TeamYouTube, 2019). YouTube also took the opportunity to introduce new methods to restrict videos that cause "widespread harm" to their "communities, viewers, and advertisers". These new steps extend beyond the strike system to include demonetization options, the removal of partnerships and other benefits, and finally the ability to restrict a video's ability to be recommended on the home page, trending tab or watch next area. Following criticism of YouTube's handling of the Maza/Crowder dispute, YouTube CEO Susan Wojcicki publicly apologized to the LGBTQ community, citing policy as a reason that Crowder's videos weren't removed: "It's just from a policy standpoint we need to be consistent — if we took down that content, there would be so much other content that we need to take down" (Recode, 2019 June 10).

As evidenced by the continued existence of Steven Crowder's channel, there is questionable material that remains on the platform in a restricted mode regardless whether it is flagged by users for removal. YouTube refers to this as 'borderline' content ("Continuing our work", 2019). Outlined in a June 2019 blog post, YouTube's examples primarily point toward fake news and their counterpart, conspiracy videos. Three more examples of this kind of content are found on the YouTube Help page for "Limited features for certain videos". While restricted videos face the removal of comments, ads, suggested videos, and likes, they are permitted to contain the following:

- Inflammatory religious or supremacist content without a direct call to violence or a primary purpose of promoting hatred
- Conspiracy theories ascribing evil, corrupt, or malicious intent to individuals or groups based on certain attributes
- Videos denying that a well-documented, violent event took place

The contradiction is glaring: Supremacist content by definition promotes hatred, conspiracy videos targeting specific groups is discriminatory, and historical revisionism is almost always done with hateful intentions. Given the company's seemingly hard stance against hateful content (even the opening line of this page reads "Hate speech is not allowed on YouTube."), there is no middle ground here; this policy does not fall in line with the rest of their guidelines. No explanation is given as to why this type of borderline content should exist on the platform at all. Curiously, these examples have now been removed from the default North American English language page but remain on the English (Great Britain) language version.

According to the help page and YouTube's recent blog post on improving recommendations, there are now actions taken by the platform against what they deem as borderline content. While the content is permitted to exist, the watch page is not accompanied by a comment section, suggested videos, and likes. Additionally, these types of videos are not eligible for advertisements, and are sequestered behind a warning message. The recent blog post concerning recommendations has also revealed that the ability of these videos to be recommended to other users has been curtailed. In June 2019, the company announced redoubled efforts against borderline content, stating that since January of that year, videos containing this kind of content were already being recommended 50% less ("Ongoing work to tackle hate"). Using the new category, the recommendation system is now tuned to be more likely to recommend authoritative sources from videos deemed to contain borderline content. According to Guillame Chaslot, borderline content is more engaging than other videos (Maack, 2019). This observation would suggest that borderline content is more profitable to YouTube, putting a tension between revenue pursuit and ethical responsibility.

### *Automating the Automated*

One of the ways YouTube's skewed value metrics have been exploited is with the use of content-generating algorithms that quickly discover the content that is most likely to keep viewers watching and pieces together disparate attributes that are deemed successful. One of the most well-known examples of this phenomenon has been the rise of strange and often disturbing videos targeted at preschoolers (Papadamou, 2019). In 2017, James Bridle, author of *New Dark Age: Technology and the End of the Future*, published a detailed inquiry into the world of kids videos on YouTube. What he discovered was more than a little odd: Channels unconnected to major intellectual properties were imitating videos on popular children's topics, mashing them together, and adding in seemingly unexplainable sexual and violent elements, as well as an overdose of toilet-topics. According to Bridle, one of the methods these bogus channels use to increase views is keyword and hashtag association, which is essentially a simple attempt to cash in on trends in children's entertainment. Using the word 'Elsa', for example, would heighten the chance that the video would be recommended in searches related to Disney's film *Frozen*. What the creators of these channels discovered was that they could optimize their chances of being recommended by simply adding *all* of the trending keywords. To do so, they used algorithms to vomit out titles like "Surprise Play Doh Eggs Peppa Pig Stamper Cars Pocoyo Minecraft Smurfs Kinder Play Doh Sparkle Brilho", and "Disney Baby Pop Up Pals Easter Eggs SURPRISE"

(Alexander, 2016). These titles are designed to provide search optimization, as the target audience usually lacks reading comprehension.

One of the major trends in this industry is ‘Finger Family’ videos, which, according to Bridle, has at least 17 million variants on YouTube with view totals in the billions partly no doubt thanks to a largely passive audience that is further drawn in by YouTube’s autoplay function. This success however is not entirely natural, for Bridle points out that “a huge number of these videos are essentially created by bots and viewed by bots, and even commented on by bots”. In a YouTube video further exploring this topic, *Folding Ideas* producer Dan Olson dives deeper only to find even more curiosities. Putting the magnifying glass to South-Asian media companies, Olson found massive networks of channels that originated from the same few source animation studios. The videos released on these channels are for the most part procedurally generated from a library of stock animations and likely do not see a great deal of human intervention. In his video, Olson hypothesizes that the strategy is to game the recommendation algorithm “by casting a very wide net with the illusion of diversity”, for while the system is designed to keep viewers watching, there are checks in place to keep a single channel from monopolizing the recommendations. The content of the videos themselves ranges from harmless concoctions of an algorithm mining some word cloud of popular topics to much stranger things, including themes of cartoon characters being buried alive and worse. For Bridle, the content isn’t so much the problem as the content delivery system and the audience, for “algorithms don’t discriminate—and neither do the kids”. He goes on to clarify that he does believe that exposing this content to kids is abuse, but the topics are not the issue, rather the “violence inherent in the combination of digital systems and capitalist incentives”, meaning the level of automation involved in the creation of these videos is doing little more than following numbers. While it can be argued that while shoveling nonsensical videos into the laps and iPads of children may be harmless, it still escapes the nuances of human curators. Without venturing into moral territory, it is fair to say that the recommender system is programmed to feed viewers what they want, and that toddlers lack the critical thinking skills to know what might affect them negatively. Bridle comes to a similar conclusion at the end of his article (emphasis mine):

> This, I think, is my point: **The system is complicit in the abuse**. And right now, right here, YouTube and Google are complicit in that system. The architecture they have built to extract the maximum revenue from online video is being hacked by persons unknown to abuse children, perhaps not even deliberately, but at a massive scale. (Bridle, 2017)

Gaming YouTube’s recommender system is to embrace its core philosophy: the longer the viewers are kept watching, the better. When that mantra is followed to its logical extension, ethics go out the window often alongside traditional media structures. As it turns out, a computer's vision of what humans like to watch is fragmentary and constantly changing, lacking human values like objectivity and truth.

Perhaps more concerningly, an automated approach has also been used to create 'news' videos that perform well under the algorithm's current objectives. Jonathan Albright, the research director for Columbia University's Tow Center for Digital Journalism, has found evidence of a network of videos created by T, an artificial intelligence platform. In his 2017 investigation, Albright found 19 channels with over 78,000 videos, all created without human intervention. After studying the upload times, he was able to conclude that the channels were able to generate new videos every three to four minutes. Drawing text from various news sources online, the videos mash sources together to create text-to-speech audio narration of current news topics, which are subsequently shared all over the internet. Everything about them, says Albright, "suggests SEO [search engine optimization], social politics amplification, and YouTube AI-playlist placement" (Albright, 2017). Although an unmonitored news-generation algorithm is cause enough for concern, their design is again an example of a way that YouTube's RS (recommender system) can be taken advantage of. This automated form of manipulation takes advantage of a demand for current news by spitting out thousands of videos without ethical oversight. Only focusing on what is being watched, YouTube's algorithm does not differentiate these videos from professional news sources and will surface them regardless.

*Malicious Interference*

If YouTube's algorithms prioritize views and engagement over any kind of commitment to truth and have the potential to negatively affect democracy, then it should come as no surprise that foreign agents might want to take advantage of the system for their own political or financial ends. The idea is simple: A bad actor can purposefully boost the engagement metrics of divisive topics to ensure that content is promoted to a larger audience. Similarly, bots could be used to create co-visitation counts *en masse* to fool the system into connecting previously unrelated audiences. While peer-reviewed literature on the subject is lacking, there is doubtless an incoming wave of academic research as details start to emerge regarding Russian and other external influence on North American information distribution systems, including recommender systems and other algorithms.

Using the word 'manipulation' in this sense does not indicate that the code itself of these systems was disrupted or even altered, instead it refers to the ways revenue-seeking parties and bad actors have been able to take advantage of already-existing systemic flaws. This research will take the position that digital information systems designed in capitalistic economies necessarily demand endless growth, often bypassing or compromising ethical considerations. In YouTube's case, this means a system designed with a relentless pursuit of views will always surface what people want to watch, or, in the case of the external actor, what the system is fooled into thinking people want to watch.

YouTube's vulnerability is recognized by members of the U.S. government: Mark Warner, a Senator from Virginia who served on the Senate Intelligence Committee, stated that "YouTube is a target-rich environment for any disinformation campaign — Russian or otherwise — that represents a long-term, next-generation challenge" (Wakabayashi & Confessore, 2017). As Warner highlights, the primary suspect of intentional subversion of North American social media platforms, including YouTube, has been the Russian government. In a statement to a Senate Judiciary subcommittee in May 2017 regarding potential Russian election interference, the former director of national intelligence James Clapper made it clear how significant this threat is:

> Russia's influence activities in the run-up to the 2016 election constituted the high water mark of their long running efforts since the 1960s to disrupt and influence our elections. They must be congratulating themselves for having exceeded their wildest expectations with a minimal expenditure of resource[s]. And I believe they are now emboldened to continue such activities in the future both here and around the world, and to do so even more intensely. If there has ever been a clarion call for vigilance and action against a threat to the very foundation of our democratic political system, this episode is it. I hope the American people recognize the severity of this threat and that we collectively counter it before it further erodes the fabric of our democracy. ("Full transcript", 2017)

Russian attempts to influence elections around the world are not conspiracy theories, nor are they vestiges of soviet-era cold-war tactics; this is happening, and it is a real threat to both YouTube and Canadian governance.

After *The Guardian's* expose of YouTube's skewed 'Up-Next' videos in relation to the 2016 American election, members of the parliament in the U.K. questioned whether the recommendation algorithm or Russia had a role in their own controversial vote, commonly known as Brexit. In response, YouTube's global head of public policy Juniper Downs defended the RS, characterizing it as a "reflection of what the viewer wants to see", and stated that "People want to watch what they want to watch. It's hard to insert something they don't want. We see an abandonment of the service when we do that" (Borger, 2018). This strange comment seems to imply that any effort to moderate the content that the system pushes somehow requires the promotion of something else, content that viewers don't want to see. In any case, Downs stated that YouTube had "conducted a thorough investigation around the Brexit referendum and found no evidence of Russian interference". As researchers lack the internal data required to do such investigations themselves, users must rely upon reports issued by Google. In 2018, Google stated that they had removed 43 YouTube channels associated with The Internet Research Agency, a Putin-backed troll-farm that exists solely to spread misinformation online (Calamur, 2018; Walker, 2018).

In an additional statement, YouTube addressed claims of misconduct related to the state-funded news agency *RT* (formerly *Russia Today*):

> Some have raised questions about the use of YouTube by RT, a media service funded by the Russian government. Our investigation found no evidence of manipulation of our platform or policy violations; RT—and all other state-sponsored media outlets— remains subject to our standard rules. ("Security and disinformation", 2017, p. 1)

The criticism surrounding YouTube and *Russia Today* started in 2013 after Robert Kyncl, now-CBO at YouTube, made comments about the channel's journalistic integrity and praised it for not promoting "agendas or propaganda" (Wakabayashi & Confessore, 2017). By the time Hillary Clinton was ramping up her presidential run, RT was one of YouTube's largest featured news organizations and used their reach to actively discredit her campaign. In a partially-declassified report by the U.S. National Intelligence Council, officials frame *RT* as being a "propaganda outlet", and detail how videos like "Julian Assange Special: Do WikiLeaks Have the E-mail That'll Put Clinton in Prison?", "Clinton and ISIS Funded by the Same Money", "How 100% of the Clintons' 'Charity' Went to...Themselves", and "Trump Will Not Be Permitted To Win" were part of an extensive plot by the Russian government to destabilize the West ("Background to assessing", 2017). After the Department of Justice forced *Russia Today* to identify as a foreign agent in November 2017 (Stubbs & Gibson, 2017), February 2018 saw YouTube take steps to identify state-sponsored media, adding a notice to videos on channels that received government funding alongside a link to the corresponding Wikipedia page ("Greater transparency", 2018).

Despite the underhanded motives behind Putin's support of *RT*, did the media company break any of YouTube's guidelines? In this case, *Russia Today* wasn't 'hacking' the system but rather taking advantage of a flawed system. *New York Times* contributors Daisuke Wakabayashi and Nicholas Confessore (2017) explain: "RT uploads videos frequently, sprinkling in buzzy viral videos of disasters — plane crashes, tsunamis, a meteor strike — to earn likes and longer watch times, which YouTube's algorithm rewards with better placement among search results and recommendations". Like YouTube channels chasing ad revenue, *RT* is 'gaming the system', or, as the owner of the YouTube-specific analytics company VeeScore.com Christoph Burseg observes, “they’re just riding the algorithms” (Wakabayashi & Confessore, 2017). As implied, the difference is that *RT* has ulterior motives. In an investigation by the New York Times, journalists found the same strategy used across various social networks, especially Facebook. Russian actors used anger and misinformation to attempt to shape the thoughts of the American public. Other than stealing YouTube videos, editing them to be more controversial, and spreading them on Facebook and Twitter, these troll farms also used their extensive networks to promote divisive videos that already existed on YouTube, boosting their engagement and therefore their ability to be naturally promoted by the recommender system as well (Confessore & Wakabayashi, 2017).

On April 25th, 2019, ex-YouTube engineer Guillaume Chaslot tweeted the following: “One week after the release of the Mueller report, which analysis of it did YouTube recommend from the most channels among the 1000+ channels that I monitor daily? Russia Today's !!!” (gchaslot). Of the videos covering Robert Mueller’s report on possible Republican/Russian collusion, Chaslot found that one video from the *RT* channel was getting over half a million recommendations from 236 separate channels. Of the 84,695 videos he included in his analysis, *RT*’s video was the clear outlier, having been recommended from nearly double the number of channels. But Chaslot doesn’t blame *RT*, and instead points his criticism at the system that is allowing the manipulation to take place: “To be clear, my problem is not with Russians, who did an amazing job at ‘optimizing YouTube’ and generated millions in advertising revenue for Google. My problem is with algorithms that were designed with little consideration for bias or abuse” (gchaslot, 2019, April 28). While Chaslot shifts blame away from those who are actively engaging in election meddling, his point still stands: If foreign interference is to be assumed, information distribution systems (particularly the most influential) should take precautions against manipulation.

## Conclusion

Context is important. Having a solid understanding of the elements that inform this thesis topic is key to being aware of how YouTube interacts with broader social and political contexts and complex systems that are black boxed like algorithms and information systems. In this chapter, the main candidates in the 2019 Canadian election have been discussed alongside cautionary tales of over-automation and foreign influence. Finding that Canada’s democracy may be susceptible to manipulation by foreign governments and other groups via YouTube, the discussion then moved to providing context regarding YouTube--a platform with enormous reach and limited oversight. Examining the values, policies, and roles of this company revealed contradictory ideals and actions, as well as reactive responses to errors in automated processes which often prioritized human moderation of video content. Some of the ways the company's policies and algorithms can be exploited are also discussed, again highlighting the importance of the subject. Marginally acceptable is what YouTube calls 'borderline content', or videos that almost break their community guidelines. As shown in the following chapter, it is apparent that the presence of this type of content is an essential part of a pipeline that takes advantage of algorithmic flaws to pull viewers to increasingly extreme ideologies.

---

[1]The term 'far-right' is used here to encompass a collection of hateful ideologies, including white supremacy, the alt-right, and neo-Nazism. The Southern Law Poverty Center (SLPC) defines white supremacy as "focusing on the alleged inferiority of nonwhites", and describes the 'alt-right' as held together by the belief that "'white identity' is under attack by multicultural forces using 'political correctness' and 'social justice' to undermine white people and 'their' civilization. SLPC defines 'neo-Nazi' as groups that "share a hatred for Jews and a love for Adolf Hitler and Nazi Germany" ("Ideologies", n.d.). The grouping 'far-right' is not exclusive to these ideologies but primarily refers to the above three.

[2]I have purposefully chosen not to identify the shooter as I do not wish to contribute towards a culture of infamy.

# CHAPTER THREE: LITERATURE REVIEW

## Introduction

In the information age, humans must make use of technology to help them retrieve and filter through the ever-increasing amount of data being produced every day. One of the technologies employed is the algorithm, essentially a step-by-step set of instructions designed to solve a complex problem. In the context of YouTube, a specific type of algorithm called a Recommender System (RS) is used to predictively serve users with suggestions of videos to watch next. This chapter will provide a survey of the existing literature on algorithms and recommender systems and will work to demonstrate the gap in which my research will plant itself. More specifically, I will be examining algorithms as a concept, how they have the potential to influence world politics, and the ways in which they may be constructed to serve users ethically. Building on this basis, I will then explore recommender systems, including their function, design, and objectives. Finally, this chapter will bring the research on these topics together to study YouTube's recommendation algorithm; I focus on the design and development of the system before moving on to survey literature discussing how that system may promote extremism online.

## Algorithms

### *Design*

Thomas Cormen, author of *Introduction to Algorithms*, offers an informal framing of the term 'algorithm' as a "tool for solving a well-specified computational problem" (2001, p. 5). Consider a basic ordering algorithm, in which a randomized list (or array) of numbers must be transformed into an ordered list. Given an input (the randomized list), the algorithm works to meet the desired output (the ordered list) by completing a series of steps. There are many different types of sorting algorithms, each which use a different set of prioritized steps to order the list. The key here is the output; this type of algorithm is referred to as deterministic or 'exact' as it is designed to provide a single, provable answer (Harris & Ross, 2006). While success for this type of algorithm is defined as 'correctness' (meaning predictable output based on input), efficiency is measured by the speed at which they are able to solve the problem (Cormen, 2001).

Another model is the approximate algorithm, or heuristic. As the name suggests, these algorithms can only attempt to meet an estimated output by making a prediction based on “certain well-known characteristics of a problem” (Harris & Ross, 2006). An example of a heuristic is an information retrieval system, which is needed to filter the overwhelming amount of data online down to a manageable size. In an information retrieval context, heuristics return and rank information based on user queries; success is measured by approximation to the defined search criteria and which qualities the specific engine prioritizes.

*Algorithms and Political Influence*

While not specifically directed at YouTube, there has been much research and public interest on the topic of algorithms and their potential influence on democracy much in part due to their pervasive and largely invisible nature. As algorithms are often used in information retrieval contexts to filter through data, concern arises regarding their ability to affect audiences by way of exposure. One of the earlier warnings came in 2004 from an American legal scholar named Cass Sunstein. In a short opinion piece in the *Communications of the ACM*, Sunstein speculated on the future of filtering, and cautioned that democracy is put at risk when people are only exposed to ideas they already agree with, which is precisely what some algorithms are designed to do. The result, he says, is that “extremist groups will often become even more extreme” (Sunstein, 2004, p. 59). This is certainly true when it comes to politics, for news agencies are incentivised to be partisan thanks to an innate bias towards information that conform with our preconceptions (selective exposure is covered in Chapter 4); furthermore a diverse information environment online provides the means to cater to any opinion regardless if it is mainstream or fringe (Perloff, 2013). Similarly, internet activist Eli Pariser uses the term ‘filter bubble’ to describe a kind of intellectual isolation formed by personalized web filters which trap users in bubbles of agreeable information (Pariser, 2012). Confirming Sunstein and Pariser’s fears in the context of news algorithms, a study by Michael Beam found that “personalized news systems usage had negative direct effects on knowledge gain” (2013, p. 31). Additionally, other research suggests that algorithmically driven personalization can “inhibit collective public outrage”; stunting the power of journalism to inspire collective action as social groups become fragmented by personalized media (Carlson, 2017, p. 1768). While the literature cited here indicates that personalization often results in negative consequences for users, the focus of this thesis is YouTube’s recommendation algorithm outside of the personalized aspect of the system. Here, I do not investigate how the RS delivers videos to users based on their watch history and other identifying signals, instead examining how the system pushes videos to new or ‘unpersonalized’ users. However, as discussed in the following sections of this chapter, there is evidence that

YouTube's algorithm suggests content based on what other users watch, making personalization complicit in the potential poisoning of every user's experience (Davidson et al., 2010).

Although many media figures have commented on the subject, few empirical studies have been conducted that examine the problem of algorithms and their political power. The academic work that has been done appears to unravel some of the assumptions in this area. For example, a 2015 study by O'Hara and Stevens examined the role of the recommendation technology in violent extremism only to find that algorithms are not solely responsible for the creation of echo chambers, nor are such filter bubbles necessarily harmful. In other words, there is no clear consensus that algorithmic filtering has negative impacts on users, even when resulting in ideological homophily. In fact, there is some sociology work to suggest that being exposed to opposing viewpoints on social media only serves to increase political polarization (Ball et al., 2018). If personalized filtering does cause the kind of intellectual isolation that Pariser and others warn about, it is not a stretch to suggest that information retrieval algorithms do in fact have the potential to affect democracy through their ability to control information exposure, but the full nature and extent of that influence is highly dependent on a number of factors, including the reach of the platform, the audience, and the information being filtered.

*Algorithms and Ethics*

An essential component of a prioritization algorithm is 'sorting criteria', organizational decisions laden with value-statements. Created by humans, these criteria necessarily result in a system biased by human values (Diakopoulos, 2015). Even something as simple as sorting and ranking the social web is an act imbued with value judgements (Mager, 2012). If a hypothetical 'neutral' algorithm could be designed, it is still the case that users "manually influence the filtering process after the algorithm has been designed" simply by interacting with it (Bozdag, 2013, p. 224). Nothing is neutral, and all content and the systems that distribute that content are shaped by human philosophies. When thinking about the human biases imbued in computer code, it is often the case that unintentional exclusion is occurring, not intentional exclusion. An example of this is facial recognition software that does not function correctly for non-white faces (Buolamwini, 2016): While developers did not purposely exclude other races when designing this technology, the choice to only use white faces to train the algorithm left people of colour unable to use it. In this way, AI is influenced by the data it is trained on, which can "unintentionally produce data that encode gender, ethnic and cultural biases" (Zou & Schiebinger, 2018). Cathy O'Neil's book *Weapons of Math Destruction: How Big Data*

*Increases Inequality and Threatens Democracy* (2016) explores these ideas further, highlighting the increasing number of decisions that are made for us by algorithms (2016). In her Ted Talk, O'Neil states that "algorithms are opinions imbedded in code" and points towards algorithms that were trained with antiquated data sets that define success or other positive metrics differently than is done today. One example used is a hiring algorithm given the applications of historically successful candidates. Without intention, the algorithm will replicate social ills simply by way of the data it was given and likely will define success as male applicants (O'Neil, 2017). Providing a more comprehensive look at the topic is Safiya Noble in her 2018 book *Algorithms of Oppression*. According to Noble, these algorithmic biases against women and people of colour are an act of willful ignorance by companies bent on pursuing their bottom line over ethical considerations, labelling Google, for example, a "broker of cultural imperialism" (Noble, 2018, p. 86). Repeatedly, Noble notes that searches for women of colour returns pornography, a practice Noble calls 'digital redlining'--a process in which technology is used to reinforce and perpetuate inequality. Throughout the book, her point is clearly conveyed: Systems are just as prone to bias as humans are because they are created by humans. It is then imperative to be mindful of the fact that YouTube's recommendation algorithm was also designed by human beings and as a result bears embedded human flaws in its code as well.

Kraemer et al. (2011) draws a similar comparison with the medical field, in which human value-judgements must be made to set diagnosis thresholds in imaging algorithms. As an example, the authors suggest that it is ultimately up to a human designer to decide if a cell is diseased or not, a preference that often lies between favoring false positives or false negatives. Wagner (2016) stresses that algorithms cannot be looked at as good or evil, but instead as great centers for power distribution which have human values inherently coded into them. Coming to a similar conclusion, Feenberg considers technology to be 'biased but ambivalent' (1991). Essentially, this means that technological institutions like the internet are influenced by their creators and the historical context of their time, and yet individual parts of that institution can still be rearranged or reconstructed (McCarthy, 2011).

## Recommender Systems

### *Information Retrieval*

As the name implies, information retrieval (IR) is the act of obtaining relevant information as it is defined by various search criteria or other limitations. As described by William Goffman, relevance is "a measure of information conveyed by a document relative to a query" (1964, p. 201). In an information retrieval context, relevance takes on a more nuanced meaning: Instead of simply meaning how well results relate to the desired information, relevance can consider other factors. For example, subject relevance measures the semantic distance between search queries and items, while situation relevance prioritizes the utility of the information. Alternatively, affective relevance measures the "capacity of the information to satisfy and please the user". (Dinet, 2014, p.30). Otherwise, factors like authority, novelty, and timeliness can help determine what is most relevant.

In *Information Retrieval in Digital Environments* (Dinet, 2014), IR is characterized in three ways: firstly, it is a composite activity, meaning that it requires multiple 'cognitively complex' activities, including reading, writing, memorizing, and making decisions. Secondly, it is dynamic as a result of a non-static information environment. The example given by Goffman is a search engine that will always retrieve slightly different items despite having conducted the same two searches seconds apart. Finally, IR is iterative due to the changing nature of the knowledge and behaviour of the searcher themselves as a consequence of each search. Here, it becomes immediately obvious why attempts to automate these cognitively complex activities seem ideal, for requiring direct input from users is unnecessary if their information needs could be algorithmically estimated with similar or greater efficacy.

Two primary components of IR are precision and recall, which are used to gauge retrieval accuracy (Sundin et al., 2017). Precision is a measurement of relevancy against total instances retrieved, while recall refers to the total number of relevant instances that have been retrieved from all the relevant resources available (Grossman, 2004). As an example, consider a search for 'Justin Trudeau' on YouTube. Of the 18 results retrieved, say 6 correctly relate to the Prime Minister: These are referred to as 'true positives'. The rest of the videos retrieved are then 'false positives', content returned by the search algorithm but are not relevant to the search. Outside of what is displayed, there is a vast array of content that is not retrieved: 'true negatives' were correctly ignored for being irrelevant, while actual relevant content that was not retrieved become 'false negatives'. The calculation of precision then is the fraction of true positives (relevant videos) returned out of all the content displayed on the results page. Conversely, recall is the amount of true positives out of all the relevant videos in the database. An information professional might use logical operations to maximize the precision or recall of a search, but rarely both: the inverse relationship between the two measurements means a search is likely to result in a small number of highly relevant resources, or a large number of resources with lower relevance (Tunkelang, 2009). This focus on relevance is indicative of the traditional pairing of search and retrieval: To retrieve something relevant a system must first be supplied with aspects of what is being sought. The relationship between searchers and the IR system is built upon a shared language of the keywords recognized by the system and the keywords used by the searcher to describe the information being sought. Effort is required from both sides to result in successful searches: searchers must utilize basic search strategies, and the system must be adjusted to better aid future searches.

Two and a half quintillion bytes of data is produced online every day (Marr, 2018). In order to effectively find and retrieve relevant parts of this information, digital systems such as search engines must be used. The use of these retrieval algorithms to rank and filter results in many cases outperforms human processes, for a 1994 study found that simple free-text queries in early search engines on average retrieved more relevant results and had better ranking for browsing than an expert using Boolean retrieval methods (Turtle). Besides functionality, the ubiquitousness of search engines is also driven by other factors. Mager (2012) posits that search engines are unavoidably tainted by capitalism. Tracing Google's trajectory from academic roots to commercial adoption, the author characterizes the modern search engine as being optimized around the concept of traffic, or flow from one website to another. Google pioneered this model by introducing AdWords, a system that presented relevant ads to the searcher based upon the search being performed. Mager argues that as Google began harvesting and storing users' searches in order to present increasingly personalized search results and ads, they began profiting from users' practices and data. Given the success of Google's search engine, it is not surprising to see them apply this model of targeted ads and personalization to other platforms they own, including YouTube.

Another of the primary focuses in information science research is human-system interaction. In the article "The search-ification of everyday life and the mundane-ification of search", the authors argue that "the internet and specifically social media have blurred the distinction between informal and formal information systems", causing the act of information retrieval to become almost invisible to everyday users (Sundin et al., 2017, p. 5). Citing literature from the field, the authors turn away from traditional models of information-seeking behaviour which require an information need or problem. Steeped in information and surrounded by complex information retrieval tools like algorithms, the modern user often happens upon information without setting out to discover it, making the complex cognitive processes of understanding, evaluating, and determining relevance mundane and overlooked.

By relegating understanding search results to the background, users rely on the services they use to be trustworthy (Haider & Åström, 2016). The term 'algorithmic authority' is used in the literature to describe the way widely accepted information retrieval systems are given undue preference, simply due to the fact that they replace the human parts of information retrieval, including determining relevance, drawing on multiple sources to retrieve results, and producing verifiable information (Lustig & Nardi, 2015; Shirkey, 2009). In their study on the mundanity of searching, Sundin et al. found that participants had faith in the sources search engines retrieved, for users often assume that the system has already assessed the quality and relevance of the information. They use Google as an example, stating that the search engine “has become part of the information infrastructure of everyday life in a way that makes us less reflective about how the rankings of search results come about and ultimately this changes how we trust technical systems” (p. 36). This level of technological authority has the power to shape and reinforce our understanding of the world as many use Google to outsource actual research.

*Design*

One kind of information retrieval algorithm is a recommender system (RS). These digital tools function to provide suggestions to users in order to aid various decision-making processes (Ricci et al., 2015) In order to provide a description of typical RS objectives, expertise is sought from Aggarwal’s comprehensive text *Recommender Systems: The Textbook* (2016). In the first chapter of his book, Aggarwal describes recommender systems as marketing tools which are used to “transform data into useful information for decision-making in industry” and states that the rise in business transactions on the web drove the development of RS. Immediately, the underlying functions of mainstream RS are clear: they are designed in order to influence decision-making, often to promote a product or concept, including commercial items, people, places, and more (Ricci et al. 2015). Aggarwal even addresses financial incentive directly, saying “increasing product sales is the primary goal of a recommender system” (2016, p. 3). In some cases, this objective is met by serving products to increase direct sales; other times the system’s content itself is the product on which advertisements can generate income. As examples, Amazon’s RS serves products perceived to be relevant to the consumer, while Facebook’s RS delivers the profiles of people users may know in order to grow a network.

Beyond profit, recommender systems are also designed to mitigate information overload. Earlier work positions RS as designed to ‘augment natural social processes’ such as making choices without sufficient personal experience (Resnick & Varian, 1997; Ricci et al., 2015). Sohail, Siddiqui, and Ali note the similarities of these systems to a person’s reliance on friends or family for advice; both serve as personalized solutions to an information-saturated environment (2017). Ricci et al. position the rise of RS as coinciding with the new information environment challenges presented by the internet and its unending horizons of choice (2015). A prime example given is a book recommendation service (Resnick & Varian, 1997), which, at least in a library environment, is created to serve an information need instead of a profit incentive.

As information retrieval systems, modern recommender systems often incorporate user signals to help make accurate recommendations, including engagement, location, and search history, among others. Three primary methods are employed: Collaborative filtering, content-based filtering, and hybrid filtering (Aggarwal, 2016). The primary objective of most RS is accuracy, defined here as how closely the system is able to rank or rate an item to match the users' wants.

Collaborative filtering takes a variety of user data, including behaviour, activity, and preferences to make a suggestion based on users with similar traits. Some will be able to recognize this recommender system in place at Amazon. Other prominent examples include Facebook, Last.fm, and LinkedIn. Relevant to this study is YouTube, which groups users together using viewing history to serve similar content to users with similar content consumption habits (Davidson et al, 2010). Content-based filtering removes the individual from the context of the rest of the consumers and instead examines the relationship between the user and the items they have previously enjoyed. Strong examples of this system in use are Goodreads, IMDB, and Rotten Tomatoes. A user profile is populated with keywords related to the type of item that the user likes; typically, preference is determined by explicit user input using signals such as 'like' buttons, rating systems, and others. However, some systems rely on less visible input and "consider the navigation to a particular product page as an implicit sign of preference for the items shown on that page" (Ricci, 2015). In YouTube's case, the amount of time a user spends watching one video (known as watch-time) is an example of an implicit user cue. Hybrid filtering occurs when these two methods are combined: this can include filtering based on demographic and consumer knowledge, among other things. The most well-known example of a hybrid system is Netflix, which recommends movies based on films users themselves have rated as well as the ratings of viewers with similar perceived tastes or demographics.

*Beyond-Accuracy Objectives*

Some RS have objectives beyond only accuracy built into them. Four relevant beyond-accuracy objectives are diversity, novelty, serendipity, and coverage. In their article "Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems" (2016) authors Kaminskas and Bridge provide some definitions to differentiate these terms. *Diversity* refers to a method used to broaden the scope of an ambiguous term. Using the example they have provided, the keyword 'Jaguar' could refer to several things, so a more diverse range of sources is needed. *Serendipity* is quite different, instead relying on the word 'surprise'. An informal definition would be a recommendation that provides a surprisingly interesting item the user might not have otherwise discovered. *Novelty* is very similar to serendipity, although this time the user is factored into the equation. In recommender systems, novelty is "an item being unknown to the user and an item being different from what the user has seen before" (p. 11). The final term to define here is *coverage*, which at the outset sounds very similar to diversity. In recommender system literature, coverage is a term used to describe what the system is able to do, not what the user experiences like the other terms. Coverage in this context "measures the degree to which recommendations cover the set of available items" (p. 14). All these terms then are factors that can play into what the end user is recommended.

In a survey on recommender systems, authors Bobadilla et al. (2013) suggest that early recommender systems prioritized accuracy above all other factors, but newer models increasingly incorporate elements such as diversity and novelty into their recommendations (Kaminskas & Bridge, 2016). In most cases, increasing diversity and novelty is only done in order to enhance the system's ability to effectively market to users, not out of a desire to expose users to a multiplicity of perspectives for more educational purposes. This idea is corroborated by other researchers (Aggarwal, 2016; Kaminskas & Bridge, 2016; Shi et al., 2012; Vargas et al., 2014; Ziegler et al., 2005) who suggest that users are more satisfied with increased diversity, despite a potential decrease in accuracy. One of the reasons that more diverse results are added to systems is that serving users redundant items that they are already familiar with or would otherwise already consume has a negative effect on the overall perceived quality of the RS (Kembellec et al., 2014). An example given by Kembellec et al. (2014) is that of a commercial enterprise that strives to increase user perception of choice by using diversity, an action that results in a "positive impact on purchasing decision". Aggarwal (2016) analyzes the operational and technical goals of an RS and identifies relevance, novelty, serendipity, and increasing recommendation diversity as common goals that may have short-term disadvantages, but long-term advantages. An important insight to observe here is that there is a financial incentive for companies to design recommender systems with beyond-accuracy objectives, and that this statement is supported by RS literature. This finding is consistent with YouTube's own practice:

Opposing research on the commercial sector (Fleder and Hosanagar, 2007), Zhou et al. (2016) discovered that YouTube's collaborative filtering system helps to increase the aggregate view diversity, or the greater range of content that is receiving views.

## YouTube's Recommender System

### *Design*

To begin to unravel this complex algorithm, the first step is to examine what it purports to do: YouTube clearly lays out the objectives of its RS in a series of videos designed to help creators succeed on their platform. In these videos, they state that "the algorithm has two simple goals: one, help each viewer find the videos they want to watch and to get viewers to keep watching more of what they like" (YouTube Creators, 2017). Although there are likely differences between the company's goals and the objectives of their algorithm, it is possible to infer that the design ultimately serves to fulfill the company's goals—namely, profit: The more time users spend watching videos, the more opportunities YouTube has to serve revenue-generating advertisements. While YouTube's statements at face value indicate that the RS is designed to help match users to content and thereby increase advertising opportunities, this objective is presented without ethical nuance. For example, what happens when viewers want to watch borderline content? In other words, how does YouTube's commitment to responsibility contrast with the way their systems are designed?

Analysis of the inner workings of YouTube's RS is limited by the fact that the algorithm itself is not a static thing, nor are the technical details available to the general public. Like any other trade secret, specifics of how the algorithm works are not publicly available, and only by piecing together documentation of previous designs with reverse-engineering attempts is it possible to begin to understand the system. One of the most central sources on this topic comes from a team of Google employees, who explain how the system worked in 2010. In the article, Davidson et al. make several important claims regarding the objectives of the system, beginning by stating that it aims to address what they call 'unarticulated want' or a user's desire to be simply entertained. To problematize the phrase 'unarticulated want', it is telling that Davidson et al. should use these words to pin entertainment demand as coming from users, when it is the algorithm itself that is creating those wants in the first place. To clarify my position, there seems to be an intentional obfuscation of what is causing users to click on recommendations.

Specifically, the words used suggest that users have an undefined entertainment need, and that recommendations are simply finding ways to fulfill those needs. In reality, it's the other way around: The system is responsible for hijacking user attention by serving recommendations on what to watch next.

In their paper, Davidson et al. (2010) also state that the system was designed to keep users engaged with high-quality videos relevant to their interests, and provide two pieces of important information regarding the algorithm in use at the time: Firstly, recommended videos are only found on the "Home" page and the "Browse" page. This revelation indicates that YouTube did not use the same algorithm in 2010 to recommend videos on the sidebar. Secondly, Davidson et al. also state that recommended videos are generated using a user's 'personal activity', which they define as watched, favorited, and liked videos. That data is combined with co-visitation counts, or the aggregated 24-hour view count of videos users have watched alongside the initial seed video in a single session. Also relevant to this discussion is the authors' statement that the algorithm is only available in a limited state to users who are not signed in. A more recent analysis of YouTube's recommendation algorithm comes from a different set of Google employees, who describe the video platform as "one of the largest scale and most sophisticated industrial recommendation systems in existence" (Covington et al., 2016, p. 191). Although the terminology can be inferred from Davidson et al. (2010), Covington et al.'s paper explicitly states that YouTube's RS makes use of collaborative filtering, a way of making predictions of user interest based on the consumption patterns of other users.

Information about YouTube's usage highlights why its cultural and social significance. Usage statistics help to reverse-engineer the system as they reveal clues as to how the company wants viewers to use the platform, assuming that the algorithm has been successfully altered over time to produce intended behaviour. Evidence suggests that the recommended videos served to the side of currently watched videos "strongly influences users' consumption pathways" (Airoldi et al., 2016, p. 2; Zhou et al., 2010). According to YouTube CEO Neal Mohan, more than 70% of the traffic to the website is driven by that recommended content (Solsman, 2018). A 2018 PEW Research study provides additional information, stating that recommendations:

1. Point towards increasingly longer content
2. Are watched by roughly 80% of YouTube users
3. Are usually videos with more than 1 million views (Atske, 2018)

The very need for this type of study paired with the dearth of technical information about the recommendation algorithm highlights the level of secrecy surrounding Google's product.

Overall, the objective of YouTube's recommendation system can be distilled to 'keep viewers watching for as long as possible', a claim repeated in Guillaume Chaslot's work with *The Guardian*. Although there has been no formal announcement of change, YouTube claims that Chaslot's understanding of the recommendation algorithm is incorrect, and recommendations do not currently prioritize watch-time. As quoted in a 2019 PC Magazine story, a YouTube spokesperson states that the algorithm is actually designed to "optimize for satisfaction and information quality," not watch time (Smith). The spokesperson does not elaborate on how 'satisfaction' or 'information quality' is defined, nor do they reveal when the change was implemented.

YouTube serves viewers with videos that they may be interested in by using a complex network of weighted signals. This neural network was called Google Brain (Covington et al., 2016), a system that was subsequently replaced in 2018 by a similar algorithm called Reinforce (Roose, 2019). In the simplest terms, a neural network is made up of highly interconnected nodes that pass along weighted dynamic information from external sources to other nodes, eventually completing tasks like recognizing patterns (Butterfield & Ngondi, 2016). This structure is modelled after the human brain and uses 'unsupervised learning' to discover beneficial relationships that software engineers were unable to (Newton, 2017). In contrast, the process of pairing videos to users is supervised, weighted and fine-tuned by software engineers to meet internal goals.

While the exact nature of these signals are corporate secrets, much of the process is known because of the papers published by the Google engineers that designed the system (Chen et al., 2019; Covington et al., 2016; Davidson et al., 2010). For example, the algorithm makes use of previously collected user data to make these suggestions: content data and activity data. Content data includes video titles, descriptions, and other metadata. The second data source, activity data, includes both implicit and explicit activities. Explicit activities include video likes, channel subscriptions, and ratings. Implicit activities are considered to be any signals that indicate interactions with videos, and includes the percentage of a video watched before the page is closed.

To make a personalized recommendation, the system takes a seed video (the video being watched) and generates a list of related videos using a process called association rule mining, or co-visitation counts. This is measured by studying a pair of videos for a set time period to analyze how many times both were watched in the same viewing session. The more often two videos are watched together, the more relevant they are deemed to be to each other, resulting in a higher relatedness score. This is an example of collaborative filtering. Combining viewing history with activity data, the algorithm creates a seed set. Using this starting set of recommended videos, potential paths of related videos are traced out to generate candidate recommendations, which are each ranked using a variety of data signals. There are three signals used at this stage to rank the videos: video quality, user specificity, and diversification. According to Davidson et al. (2010), video quality refers to the likelihood that a video will be appreciated regardless of who is viewing it. This includes view count, ratings, upload time, and commenting, favouriting, and sharing activity. User specificity signals are properties of the seed video like view count and time of watch that represent the user's preferences. Diversification in this context only refers to the balance maintained between relevancy and problems associated with over-relevancy such as recommendations from the same channel or a large amount of videos being recommended from a single seed video. It is also possible to manually correct the algorithm's behaviour by using the 'not interested' button included in the drop-down sorting button. In any case, YouTube's recommendation system can be simplified into four steps:

1. Association rule mining
2. Creation of seed set
3. Generation of candidate recommendations
4. Ranking

In their paper, Davidson et al. focus on personalized recommendations for signed-in users, stating only that the recommender system is available in a 'limited form' for signed-out users. In order to recommend videos to signed-out users, it can be assumed that the limited form of the algorithm would not use user viewing history and activity data, as those signals are tied to user accounts. Following this logic, recommendations could potentially be made by restricting the seed set to the first video that is selected as a result of a search, expanding it with the subsequently viewed videos. In other words, it is probable that all YouTube recommendations are personalized in the sense that the system requires a seed video from which to build upon, and do not suffer from what is referred to as 'the cold start problem', a situation in which an RS does not have sufficient information to build recommendations from (Lam et al., 2008). There is however no documentation outlining how the RS makes recommendations for signed-out users.

*Development*

After YouTube was founded in 2005, the way the platform has recommended videos have undergone several changes. Preceded by 'leanback', the algorithm has evolved over the years to better surface videos to users. Perhaps in an effort to mimic traditional television, 2010 saw YouTube roll out an extension of the recommended videos sidebar: the autoplay feature. Now, when a video finished playing, the highest-ranked recommended video automatically started playing. This video is identified by the small ‘up-next’ portion of the recommended videos sidebar (“Latest changes”, 2010). In March 2012, the YouTube team announced changes to the way results were ranked. As creators learned to game the system by using misleading thumbnails and exaggerated titles, YouTube realized the need for a change in signal weighting. In a 2014 interview on the YouTube channel *Computerphile*, Vice President Engineering at Google, Christos Goodrow, provided an example of why the algorithm needed to be changed:

> There were several videos that had titles about a particular boxing match, and one of them, especially, had a thumbnail of a guy like he was *just* about to be hit, and [...] his fist was right up in the other guy's face, and *everybody* was going to click on that one. What turned out [was] that the video behind that was a person talking about the fight, and it had no footage of the fight--and eventually the viewers would finally get to the one that had some footage of the fight. When we looked into that situation we realized that one way to detect the difference between these was one of these videos wasn't [...] getting watched very long and the other one was. And so instead of just ranking the videos by how often they were clicked in response to a query, we rank them for how much they were watched in response to a query. (*Computerphile*, 2014)

To combat this rise of clickbait, YouTube then adjusted the algorithm to favour watch time over clicks, ensuring that viewers were getting recommendations for videos people continued to watch (“Changes to Related”, 2012). In a separate move a month later, another major change was introduced: now all creators were permitted to run ads alongside their videos, whereas before any ad-bearing channel had to be vetted (“Being a YouTube Creator Just Got Even More Rewarding”, 2012). Suddenly, there was an enormous financial incentive to game the system.

Before Google’s Brain project, the company used a machine-learning program called Sibyl to serve users videos they want to watch (Woodie, 2014). Channeling its namesake, a mythical Greek prophetess, Sibyl used a massive dataset of user interactions with recommended videos to find click-rate probabilities, which allowed it to infinitely fine-tune its recommendations. In 2015, YouTube started using Google Brain, the more advanced big brother of Sibyl that was able to find connections too nuanced for human engineers to spot (Newton, 2017). As the machine-learning capabilities of Google Brain increased, other small changes started to take place where the algorithm had found ways to increase viewership. In 2016, 190 micro-changes were made at the suggestion of the system, including altering the length of videos recommended to different platforms, for Brain had noticed that mobile users were more likely to watch shorter videos (Newton, 2017).

The next upgrade was a new system altogether: Now YouTube uses Reinforce, an artificial intelligence recommendation algorithm that employs reinforcement learning to predict which videos will expand users' interests over time and therefore keep them using the website (Roose, 2019). The two most distinguishing factors of reinforcement learning are iterative trial-and-error and a delayed reward or 'reinforcement' signal. Like all machine learning, the idea is that the algorithm is not supplied with the necessary steps to meet those goals and can only find "which actions yield the highest reward by trying them" (Sutton, 1992, p. 1). The delayed reward aspect relates to cumulative rewards that are made up of individual success states; in other words, reinforcement learning attempts to maximize those rewards by attempting solutions and retaining the most successful. In the case of YouTube, Reinforce's reward is user engagement. It is designed to recommend videos in a way that will broaden the scope of a viewer's interest, therefore keeping them watching longer. According to some YouTube spokespeople, this new algorithm was implemented to minimize bias towards popular content and has already increased sitewide views by nearly 1% (Roose, 2019).

In 2019, a technical paper was released by YouTube engineers outlining the testing, scaling, and implementation of Reinforce (Chen et al.). The researchers confirm that the algorithm makes use of reinforcement learning to maximise the user's long-term satisfaction with the system, defined as an increase in clicks and or watch time. While immediate (short term) rewards are user interactions with videos, the long-term reward is that interaction data, aggregated over a 4- to 10-hour period. The process to recommend a video then is as follows:

> For each user, we consider a sequence of user historical interactions with the system, recording the actions taken by the recommender, i.e., videos recommended, as well as user feedback, such as clicks and watch time. Given such a sequence, we predict the next action to take, i.e., videos to recommend, so that user satisfaction metrics, e.g., indicated by clicks or watch time, improve. (Chen et al., 2019, p. 2)

In contrast to previous iterations, this version of the recommendation engine will 'reinforce' short-term gains (user engagement) into long-term rewards, in this case an overall increase in watch time. Prior to the introduction of reinforcement learning, YouTube's recommendation system focused on individual success states, i.e. which recommendations resulted in the highest engagement metrics. During live testing, researchers found that the test model (which favoured videos 'less-viewed' than the control model) did not increase ViewTime (length of time a video is watched), but did increase the total number of videos viewed by 0.53%, which, according to the authors, suggested that users were "indeed getting more enjoyment" (Chen et al., 2019, p. 8).

During the writing of this thesis, YouTube also announced major changes to how 'borderline content and content that could misinform users in harmful ways' was promoted by the algorithm. In January 2019, the platform made waves on social media after promising to use human evaluators to help train their machine-learning systems to make better, more trustworthy recommendations. Despite the gravity of their promises, YouTube stressed that only about 1% of videos on the platform would be affected by this change. As mentioned previously, YouTube's cited examples point towards conspiracy videos and other clickbait material that could mislead viewers. Alternatively, some journalists hypothesize that YouTube's true intentions are to suppress the flood of profitable far-right videos plaguing the platform (Eordogh, 2019). Given the bad press YouTube has received for hosting borderline content, it seems likely that the vague language used in the release is only public relations management and (as they admit) does not represent significant change.

*Radicalization*

As previously established, the ethically blind design of the recommendation algorithm can lead to the creation and popularity of content that rides YouTube's policy line seemingly without crossing over, all with the intent to increase user engagement. This borderline content is often a gateway to more extreme ideologies (Bahara et al., 2019). One of the newest breeds of far-right extremism has been termed the alternative right, a movement that has its roots burrowed deep into the internet (Lyons, 2017). The origins of the alt-right are murky at best. Some position the definitive beginnings of this amorphous hate group at the creation of the website alternative-right.org, brainchild of the white-nationalist Richard Spencer. Others suggest that the movement only became somewhat cohesive at the infamous Unite the Right rally in Charlottesville, an event that featured a strong neo-Nazi presence and left one counter-protester murdered. In any case, a rigid definition of the movement is difficult, as it is made up of many fragmented groups. In an excerpted report from Matthew Lyons' book *Insurgent Supremacists: The U.S. Far Right's Challenge to State and Empire* (2017), the author traces the origins of this seemingly amorphous group and offers a summary of their ideology. Lyons identifies five major ideological currents within the alt-right: white nationalism, patriarchy, male tribalism, right-wing anarchy, and neoreaction. The final group (neoreactionaries) are defined by a belief that "differences in human intelligence and ability are mainly genetic, and [...] that cultural and political elites wrongfully limit the range of acceptable discourse" (p. 12). Lyon's report highlights a few reasons why the alt-right has been so successful, citing their ability to work under a 'big-tent' philosophy that is able to contain a relatively diverse range of viewpoints, and Donald Trump's rise to the presidency. In 2017, Lyon predicted that this group was in a "strong position to pursue a 'meta-political' transformation of the political culture and thereby lay the groundwork for structural change, centered on its vision of a White ethnostate" (p. 17). Overall, if one is to understand the gist of the alt-right, it is a movement characterized by white male supremacy, spread with the power of anonymity and internet activism. Particularly relevant to the discussion here is Lyon's characterization of the alt-right as "a group that has a strong internet presence and embrace of specific elements of online culture" (p. 2). For the alt-right and other far-right groups, one of the most fruitful recruiting grounds online has been YouTube. While viewership of alt-right content on YouTube has been in decline since 2017, researchers Munger and Phillips found that the remaining audience is "more engaged than any other audience, in terms of likes and comments per view on their videos" (2019, p.7).

As YouTube is a global platform, it follows that its influence is not limited to North America. In 2019, *New York Times* reporters Max Fisher and Amanda Taub investigated YouTube's responsibility for the radicalization of Brazil, culminating in the election of the proto-fascist Jair Bolsonaro in late 2018. In the article, the authors detail how a young Brazilian man named Matheus Dominguez journeyed from guitar tutorials to conspiracy videos in one year thanks to YouTube's recommendation system. Far from an isolated example, Dominguez's experience is common in Brazil, a country in which only one television station is more widely watched than YouTube. Fisher and Taub note that even Maurício Martins, the local vice president of Mr. Bolsonaro's party in Niterói, "credited 'most' of the party's recruitment to YouTube". To back this claim up with data, the authors turned to research by the Federal University of Minas Gerais which analyzed thousands of video transcripts to find that right-wing YouTube channels in Brazil grew faster than others. The researchers also found that while Bolsonaro's popularity began to dip in the polls, positive depictions of him on YouTube only grew, a trend that Fisher and Taub claim is evidence for something more than a simple reflection of political trends. Similarly, the *New York Times* reporters also present research from Jonas Kaiser, Yasodara Córdova, and Adrian Rauchfleisch at Harvard's Berkman Klein Center, who used a Brazil-based server to follow thousands of recommendations several layers deep to explore where viewers were organically being pushed to. After watching videos about politics or entertainment, users were more likely to be recommended "right-wing, conspiracyfilled channels”. Once viewing these types of videos, the viewers were more likely to be served similar content.

Recently, a collaborative study by the Dutch news organizations *De Volksrant* and *De Correspondent* found that YouTube encourages right-wing radicalization (Bahara et al., 2019). Using data collected from 1,500 channels, 600,000 videos, 120 million responses, 15 million recommendations and 440,000 transcriptions of videos, researchers teamed up with algorithm experts, data analysts, and media scientists to develop a comprehensive picture of YouTube as a whole before focusing in on what they call the ‘reactionary right’, a group which they describe as having a “fierce aversion to progressive values”. While the study’s scope is broad, it is the interviews with members of the alt-right that humanize the research in a way that numbers alone cannot. Targeting comments left on alt-right videos, they selected 175 Dutch YouTube users who were willing to be interviewed. Lukas, one of their participants, outlined his own journey: an interest in Atheism led him to the *Amazing Atheist* channel, which soon began to also produce anti-feminism content. Thanks to the recommendation system, Lukas was soon introduced to YouTubers like ‘Sargon of Akkad’, ‘Jordan Peterson’, and ‘Stefan Molyneux’, all of whom reject progressive values and preach distrust in the ‘mainstream media’. Bahara et al. characterize Lukas’ interview responses as ‘friendly and eloquent’, contrasted with his comments on YouTube, which are described as ‘raw’ and ‘hateful’. The study also tracked the comment history of viewers over time, tracing how one user turned from supporting the progressive channel *The Young Turks* to eventually calling for the execution of left-wing politicians after becoming increasingly mired in alt-right YouTube channels. Given previous research and their analysis of 175 YouTube users and their comments, *De Volksrant* found there are three primary causes of the radicalization happening on YouTube: the technology behind the algorithm, the ‘value-free’ way in which they recommend videos, and culture of the communities that exist on the platform. Citing the Alternative Influence Network report (Lewis, 2018), they go on to suggest that guest appearances on these YouTube channels can lead viewers down dangerous rabbit-holes. When the self-described libertarian Dave Rubin interviews the ‘race-realist’ Stefan Molyneux, the algorithm will tie the two together, regardless if ideas are challenged on either side. Another participant acknowledges the power of the RS but frames the situation in a positive light: "Yes, I have been influenced by the recommendations. But that is why I have discovered what suits me, what my values are.” Although the details of the methodology used in this study are largely absent, Bahara et al. have managed to demonstrate real-world examples of individuals who have been turned to the far-right by YouTube’s recommendation system.

The Alternative Influence Network (AIN) report itself examines a web of 65 connected figures on 81 channels from a range of mostly right-wing political positions on YouTube. According to author Rebecca Lewis, these channels often collaborate with each other to form a network of adjacent ideologies, an action that often forces YouTube's algorithm to tie them together. Intentionally or otherwise, a viewer is likely to be slowly radicalized by being exposed to increasingly extreme ideas. As an example, Lewis uses an interview between Dave Rubin and his guest, Stefan Molyneux. Throughout the interview, Molyneux outlines his beliefs in scientific racism, claiming that certain racial groups are predisposed to have higher IQs than others. These ideas are not challenged or questioned by Rubin. As this collaboration increases the chances that a viewer might search for the guests' channel, the algorithm notes the crossover between viewers and begins to promote the two channels in each other's recommendation sidebar. This is an example of an ethically blind system being used to promote extremism. In the words of Lewis, "YouTube monetizes influence for everyone, regardless of how harmful their belief systems are" (Lewis, 2018, p. 43). Again, the key takeaway is responsibility. Lewis argues that YouTube should consider not only the content of a channel, but also the voices and ideas it platforms. Despite its usefulness as a resource to identify nodes of an alt-right-adjacent network on YouTube, the AIN report only measures interactions between channels, a surface-level analysis that lacks nuance. Regardless, Lewis' report provides an invaluable list of alt-affiliated actors and has been used in my own research here to measure the extent to which alt-right voices are being represented in YouTube recommendations. The channels identified by Lewis are not insignificant in terms of audience, nor are they to be considered as fringe actors. Munger and Phillips compare the viewership of AIN with CNN, MSNBC, and Fox News to demonstrate that "global hourly viewership of the AIN has consistently eclipsed the 'Big Three' cable news channels  since 2017--and that the rise of the former has been precipitous" (2019, p. 4).

Providing an expanded list of alt-right-adjacent channels on YouTube and a more comprehensive investigation into the ways viewers are radicalized through the recommendation system, Ribeiro et al. (2019) analyze 331,849 videos from 360 channels, including the 79 million comments on those videos. Their article, "Auditing Radicalization Pathways on YouTube" found that Alt-lite (a group that does not publicly endorse white nationalism and related views, but often promotes those views in less overt ways) content can often be found in recommendations on Intellectual Dark Web (I.D.W) channels. Ribeiro et al. define the I.D.W. as a group that "discuss controversial subjects like race and I.Q. without necessarily endorsing extreme views" (p. 1). Additionally, they found that Alt-right channels are present in recommendations on both I.D.W. and alt-lite channels. These three communities (the I.D.W., the alt-lite, and the alt-right) were found to share a similar user base, and that users "consistently migrate from milder to more extreme content" in these pathways. Despite these significant findings, the authors were hesitant to solely blame the recommender system for user radicalization. While the study does provide evidence of YouTube's failings, the paper itself lacks analysis of the multi-directional movement of users, as identified by Munger and Phillips (2019). For example, the authors demonstrate that I.D.W.-exclusive commenters are 3% more likely to comment on alt-right channels than the control group but fail to provide contrasting information regarding movement out from alt-right channels to more moderate ones. Like data on how users might traverse from left to far-left channels, further comparison could have given the study more weight and context. Penn State researchers Munger and Phillips specifically criticize Ribeiro et al. for failing to demonstrate that YouTube's algorithm has a "noteworthy effect on the audience for Alt-Right content", showing how Ribeiro's own research indicates that encountering a video from the alt-right only occurs no more than once every 10,000 trips (2019, p.8). However, Munger and Phillips themselves fail to take note of Ribeiro's et al.'s recognition that their research does not account for personalization, and that their already significant findings would likely be amplified by personalized recommendations. Regardless, Ribeiro et al.'s research is particularly effective at identifying specific alt-right and alt-right-adjacent communities and how members of those groups tend to become increasingly radicalized.

As increased media attention to YouTube and its algorithm has brought scrutiny on the platform, criticism is coming from journalists investigating potential problems with YouTube's recommendation system. What is lacking however, is a wide range of in-depth academic study of YouTube's algorithm (Munger & Philips, 2019). In 2014, authors O'Callaghan, Green, Conway, Carthy, and Cunningham postulated that YouTube's recommender system set consumers of extreme-right content on a path towards an ideological bubble by narrowing the range of content they were exposed to. Following their investigation of extreme-right content on YouTube, the authors framed the platform's RS as being potentially complicit in influencing political thought, and by extension, possibly action (Bahara et al., 2019; Kaiser & Rauchfleisch, 2018; O'Callaghan et al., 2014). Specifically, O'Callaghan et al. worry that exposure to extremist online content may result in ideologically motivated violence. Following the explosive investigative report from *The Guardian* (Lewis et al., 2018), other prominent figures were quick to add to the conversation, including Zeynep Tufekci, an associate professor at the School of Information and Library Science at the University of North Carolina. In her scathing opinion piece for the *New York Times*, Tufekci likens YouTube to a restaurant that only serves increasingly sugary, fatty foods, a practice that ultimately serves customers what they want but not what they need (Tufekci, 2018). Researchers Jonas Kaiser and Adrian Rauchfleisch (2018) quickly built on work done by *The Guardian* to follow the recommendations provided by Google from an initial set of 1,356 channels. From the first set, the researchers retrieved a list of 13,529 recommended channels, which they used to create a map of YouTube's universe. This visualization demonstrates a key takeaway: YouTube creates a right-wing filter bubble. Kaiser and Rauchfleisch clarify this finding by stating that the algorithm is "not creating something that is not already there", or in other words, the videos and their content exist separately from the algorithm. The recommendation algorithm however both connects far-right communities and isolates them by overwhelmingly recommending videos that populate the same ideological sphere. The question then becomes what to do about it.

Not all researchers agree that YouTube's recommendation system radicalizes its viewers. Munger & Philips argue that the 'dominant hypothesis' of recommender radicalization is not supported by enough data. In fact, their own research demonstrates that viewership of far-right videos peaked in 2017, and this lack of growth makes it more likely that the RS would now have a de-radicalizing effect (2019). This change in viewing pattern may be associated with YouTube's January 2019 statement, which vowed to crack down on conspiracy videos, although it was unclear at the time what this entailed. In February of that same year, QUT faculty of law professor Nicolas Suzor used the list of channels found in the AIN report to investigate the recommendations of 'alt-right' content over a month. Using a random sample of 3.6 million videos, Suzor found that in the first two weeks of February, YouTube was recommending videos from at least one of the listed alt-right channels on more than one in every thirteen randomly selected videos, coming to 7.8% of videos studied. In the final two weeks of February, this percentage dropped to 0.4%. Citing this dip and what they deem a lack of quantitative evidence, Munger and Phillips assert that it is not YouTube's recommendation algorithm that is necessarily turning otherwise neutral viewers into radicals, but rather that it is the environment of YouTube itself that is ideal for niche and extreme viewpoints:

> We believe that the novel and disturbing fact of people consuming white nationalist video media was not caused by the supply of this media 'radicalizing' an otherwise moderate audience. Rather, the audience already existed, but they were constrained by the scope of the ideology of extant media. The expanded supply allowed them to switch into consuming media more consistent with their ideal points. (2019, p. 12)

Munger and Phillips aren't wholly incorrect; the endless supply of viewpoints on YouTube supply an undoubtedly large pool of white nationalists around the world. However, their research positions users' ideology as static and unchangeable. Furthermore, Munger and

Phillips do not address the body of qualitative and quantitative research produced by Bahara et al. and others that all point towards the algorithm as having a direct role in radicalizing YouTube users.

## Conclusion

The research in the RS field indicates that systems are designed to serve consumers with suggestions in order to increase company profits, but also to personalize recommendations to soothe the burden of information overload. Additionally, many RS's implement beyond-accuracy objectives that ultimately still serve marketing goals. Of those objectives, diversity may be beneficial to online intermediaries and algorithm curators (Helberger et al., 2016). Some insight regarding the objectives of a system can be gleaned from the output of the algorithm; Diakopoulos (2015) finds that reverse-engineering the input-output relationships of algorithms can reveal important information aspects of its design, to some extent an endeavor undertaken here. Also underlining the need for this kind of research is Beam (2013), stating "Empirical results testing theory are needed to help us understand positive and negative consequences of the evolution of information distribution systems" (p. 1038). Many of the observations about search engines also apply to recommender systems. Not only do recommendation systems complete what would be a complex cognitive process for humans, they are also driven by financial incentives and unconsciously given authority by users. Differing from search engines, there is a troubling of the concept of relevancy, for there is no direct query to measure against. Parted from searching, the act of retrieval then is left to a complex system of associations and predictors that attempt to keep viewers watching videos, and therefore watching the accompanying advertisements.

While various researchers have discussed the inherent objectives built into RS, YouTube's own motivations, the potential real-world impact of said objectives, and the possibility of introducing ethical safeguards to algorithms, few have brought these ideas together to form an argument regarding the ethical responsibility that corporations have to regulate their recommender systems. Furthermore, few have addressed YouTube specifically in a Canadian context. As discussed in chapter 5, although the work done by *The Guardian* to investigate the American election in 2016 is indeed comprehensive, engaging, and thoughtful, their study is in some ways lacking further detail. For example, the methodology provided does not adequately define 'partisan' or how such a quality is determined. Regardless if this research is able to produce comparably unsettling findings, it will be able to show a new perspective on the use of recommender systems. In doing so, this research will meet the glaring need for an academic investigation into YouTube's role in Canadian democracy. Validation of this project's relevance to the fields of Digital Humanities and Library and Information Studies should be self-evident in its discussions surrounding ethics, technology, and information retrieval. In this literature review, significant research in the areas described above are used to support five basic claims:

1. *Recommender systems are primarily designed with profit motives*
2. *YouTube's video recommendation system prioritizes watchtime*
3. *Algorithms can negatively affect democracy*
4. *YouTube recommendations can radicalize some users*
5. *Recommender systems can be designed with ethical principles*

Using the support of peer-reviewed academic literature, it is then possible to integrate these statements into a singular argument: As YouTube's recommendation algorithm prioritizes watchtime in order to pursue profit motives, their design does not consider the potential negative consequences that may arise when people use YouTube in the way that the algorithm designers intended. Ergo, it is incumbent upon YouTube to introduce ethical principles into their system design.

## Introduction

Theory provides frameworks from which behaviour or phenomena can be explained and predicted. In this chapter, theory is used to describe selective exposure, gatekeeping theory, theories surrounding the relationship between technology and society, and finally the media theorist Marshall McLuhan's thoughts on the effects of media as technologies. By exploring and applying some of these theories surrounding information consumption and distribution, I have used theory to gain new perspectives on online behaviour and the systems designed to exploit human psychological tendencies. Theory as a structure will guide my contextual understanding and overview of previous research through to the results of my own study by continually asking the same question: How do the structures in place in this specific information environment inform the user's experience of the content? As this research is not just an investigation into bias but also the prevalence of borderline content on YouTube, it is also useful to question who holds power in information distribution, whether that be platforms or the people that use them. To help answer these questions, selective exposure theory proposes an understanding of human bias, while gatekeeping theory helps explain how the internet has caused a paradigm shift to take place on the flow of information. Combined with Marshall McLuhan's thoughts on media, theories related to the interplay between society and technology provide structure from which to build ideas of how YouTube became the social force that it is today, and how researchers might begin to think about bringing effective change.

## Selective Exposure Theory

From a psychological perspective, Joseph T. Klapper's Selective Exposure theory hypothesizes that people fundamentally tend to avoid content that clashes with their preconceptions. At the same time, the reverse is true: Humans tend to favour information that supports their worldviews (Klapper, 1960). Although there are differences, selective exposure is often a term used interchangeably with 'confirmation bias' (Hart et al., 2009), and is made up of three primary concepts: Selective exposure, selective perception, and selective retention. As mentioned previously, the core idea behind selective exposure is the push and pull of opposing forces: An avoidance of information that challenges pre-existing views is combined with an attraction to information that confirms previously held opinions. An example of selective exposure is the selection of a newspaper, where an individual might pick the outlet that most closely represents their views on the world. Similarly, Klapper's definition of selective perception is a situation where a person is less likely to notice opinions that counter their own. This bias could be represented by a person who is reading a newspaper and subconsciously only paying attention to the articles that they agree with. Conversely, selective retention explores the idea that people more accurately remember information that reflects their own beliefs. The equivalent example would see an individual struggle to recall an article that presented them with evidence contradicting their previously held beliefs (Klapper, 1960). In his book, The *Effects of Mass Communication*, Klapper suggests that mass media does not particularly influence audiences as much as it reinforces beliefs, a symptom of these underlying psychological tendencies. He refers to this phenomenon as reinforcement theory, or the idea that humans are not passive information sponges that can be swayed easily. At the end of the day, people "tend to consolidate or reinforce existing opinions rather than change them each time they are exposed to a new set of opinions presented through the media" (Heath, 2013, p. 2). The logical conclusion of these cognitive distortions of media is that some opinions are reinforced over others in ways that slowly shape an individual's perception.

This idea forms the basis of nudge theory, or the means employed by governmental regulators and private companies to influence a population's choices either towards positive change or new consumer behaviour. A nudge is defined as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler & Sunstein, 2009, p. 6). An example of the difference between a nudge and a mandate is given by Thaler and Sunstein in their book *Nudge*, where they illustrate the theory with a scenario where fruit is placed at eye-level to encourage consumption in contrast to outright banning junk food (2009). In this example, although people remain free to make any choice they wish, their consumption patterns can be influenced by optimizing food placement with the natural human tendency to select items at eye-level. As outlined in Silvia Knobloch-Westerwick's 2015 book, *Choice and Preference in Media Use*, another human bias that can be targeted is a tendency to conserve cognitive energy when searching for information when making decisions (Fischer, 2011; Kaye & Sapolsky, 1997); also referred to as the 'cognitive miser' hypothesis (Harvey, 1980). This preference overrides any notion of consistent optimal choice, for people tend to make media choices that do not meet 'rationality criteria' regardless of the perceived importance of the decision (Kahneman, 1994). In addition, Knobloch-Westerwick cites research claiming that individuals make use of cognitive shortcuts to make media choices (Wilson & Gilbert, 2005). The result then is that information or media preferences "occur mostly spontaneously and without considerable elaboration" (p. 19). Overall, the research indicates that individuals do not often make rational, measured media decisions, but are instead motivated by expending the least amount of cognitive energy possible. The tendency towards reducing cognitive effort leave information-seekers open to subtle manipulation or nudges. Similarly, Fischer (2011) proposes that seeking confirmatory information requires less cognitive effort than searching for 'decision inconsistent' information, meaning that people tend to search for information that they already agree with because it is easier. Extending this line of inquiry, research shows that selective exposure in an online environment is greatly influenced by speed (Lohr, 2012), suggesting that a more instantaneous environment may require less cognitive effort to navigate.

While traditional media like TV, radio, and newspapers 'push' static, pre-curated content to the public without the possibility of interaction, the internet and other media can simultaneously push content while allowing users to 'pull' or seek out whatever information they want directly while ignoring the rest (Bimber and Davis, 2003). This capability leads to ideological isolation online: If it is easier for users to avoid content that clashes with their preconceptions in an online environment, they will do so: "The Internet, therefore, has the potential to magnify confirmation-biased exposure patterns" (Knobloch-Westerwick and Kleinman, 2011, p. 172). These ideas about information consumption online is borne out in research that found that participants were twice as likely to pursue information that conformed to their point of view than content that directly opposed it (Garrett, 2009; Hart et al., 2009). Another study indicated that displaying digital content that opposes a user's viewpoint results in negative emotions (Graells-Garrido, 2013). Contradicting these findings, Messing and Westwood propose that it is not news media on the internet that causes group polarization (as a result of selective exposure), but more specifically social media. To Messing and Westwood, selection of content is influenced by social endorsements (opinions expressed by online acquaintances), the presence of which "reduces partisan selective exposure to levels indistinguishable from chance" (2014, p. 1). Focusing on the consumption of political information online, Knobloch-Westerwick and Kleinman investigated selective exposure in relation to the 2008 presidential election in the United States (2011). Using data collected from 205 participants, these researchers found that there were further lines to be drawn between the different characteristics of information seekers: If a user frequently consumed news online, they were less likely to succumb to selective exposure. The opposite was observed to be true as well: Infrequent browsers were more likely to demonstrate a predisposition to confirmatory information. In addition, Knobloch-Westerwick and Kleinman found that if the participant's favoured party was projected to win the election, significant confirmation bias was present. Alternatively, their results indicated that bias could be overruled by information utility if the participant's favoured party was likely to lose an upcoming election. Given the authors' definition of information utility as something that is expected to benefit the seeker, this observation can be restated: People are more likely to exhibit selective exposure bias when their preferred political affiliations are winning, while it becomes less important to consume confirmatory information at the expense of useful information if their chosen party is losing.

*Application*

Applying some of the aforementioned theories to the research problem at hand, there are clear connections to existing structures and phenomena. For example, Klapper's selective exposure has significant implications on how information is consumed on YouTube. If humans have a bias towards information that echoes their own worldviews, then a recommender system that uses co-visitation and personalization metrics to push similar content to viewers can only exaggerate this effect. In fact, one could argue that the system has been crafted to take advantage of this human flaw: If keeping viewers watching means potentially putting them in ideological bubbles, then it seems likely that a profit-driven technology would gladly take that risk. As viewers are already internally motivated to pursue, avoid, and overlook content that runs contrary to their opinions, Klapper's observation that mass media reinforces beliefs is particularly astute when applied to YouTube. Furthermore, YouTube's recommendations also play into a human tendency to conserve mental energy when making media decisions. By automatically carrying out the majority of the filtering work and automatically playing the next most recommended video, the platform's RS is capitalizing on the quick cognitive shortcuts people already make online. While the pull of a Google search requires engagement and activity from the user to complete the search task, the process of selecting a suggested video from a pre-filtered list is a push to a passive user. Even more relevant to the research questions posed in this thesis are Knobloch-Westerwick and Kleinman's findings on selective exposure and political information consumption. In their research, they found that information preferences changed as a result of how well the political party they aligned with was doing. If their party was succeeding, people were more likely to be affected by confirmation bias and only consumed information that aligned with the views expressed by that party, in addition to being motivated to seek out the same content for its informational utility. For those whose party was not winning, the usefulness of being informed about "upcoming political decisions and circumstances" overrode any preference for information that aligned with their chosen party's perspective. These findings may suggest that external factors (such as the success of a political party) may alter information preferences outside of algorithmic effects like filter bubbles. In other words, information preferences are already affected by factors outside of human control.

The extent to which algorithms and other information systems exaggerate this influence is up for debate. In an attempt to ease cognitive effort, it is possible that YouTube's recommendation system dangerously exaggerates natural information consumption habits. By using recommendations to reduce the cognitive effort involved with making information choices, the system could exaggerate typical behaviour by way of ease. In this way, YouTube's RS attempts to alleviate cognitive load pushes towards ideological homogeneity, albeit without the intent to control user's information preferences through recommendations--only the negligence of pursuing video views without considering the consequences for audiences, and by extension, society. This is not an example of nudge theory, as nudges require a greater and necessarily positive intent (e.g., a governmental policy designed to better the lives of citizens). Instead, the recommendations blindly push users to the desired behaviour: longer watch-times. In the concluding chapter, I further address YouTube's move towards the new engine Reinforce and how it may work against the dangerous effects of filter bubbles and selective exposure by finding that recommending homogeneous content to users is antithetical to information preference growth, and therefore financial gain.

**Gatekeeping**

Another useful theoretical lens to apply to my research problem is gatekeeping. This theory is concerned with questions of who controls the flow of information, and how that impacts which events and topics are considered to be more or less newsworthy than others. While gatekeeping theory is typically used to address biased political perspectives of traditional media outlets, the ideas can be applied to new information environments online. While previous publishing intermediaries like newspapers and radio stations had the power to include or exclude (emphasize or de-emphasize) information that the public consumed, the structure of Internet protocols as envisioned by Tim Berners-Lee left control of the network in the hands of users, which by-passed "traditional" information mediaries. More specifically, this meant that data packets sent over networks could not be prioritized over others (Berners-Lee, 2010). The openness of the internet brought with it a wild-west kind of anarchy to information distribution: Suddenly, information consumers had the power to decide themselves what was newsworthy, and who to consider as authoritative. This move away from established streams of information also had other consequences: As business models became increasingly focused on viewership, different tactics emerged in the battle for clicks. Attention span maximization became an exact science, and ad revenue as an income model meant that the number of visitors to a page was more important than the page's contents. For some web creators, this resulted in the creation of clickbait, the practice of using intentionally misleading headlines and images. For web users, the advertising model was accompanied by an expectation for 'free' content, leaving many users unaware that they and their personal information were the product themselves. Here, advertisers become gatekeepers for online content as well, forcing creators to find other revenue streams if their content is not deemed palatable for the advertisers' market.

It is a mistake to consider online information distribution platforms as having less bias than their human-curated predecessors simply due to their automation. Researcher Dr. Engin Bozdag warns that "online gatekeeping services are not just algorithms running on machines; they are a mix of human editors and machine code designed by humans" (2013, p. 224). Bozdag also discusses automation bias, a human tendency to trust automated decision-making systems over contradictory information as being a potential threat to a realistic understanding of digital distribution: "Because these information intermediaries automate their core operations, often, mistakenly, they are treated as objective and credible" (p. 210). For example, Google's search engine is often the go-to resource for fact-checking, but instances of error occur from time to time (see Chapter 2 for examples of the retrieval of holocaust denial websites). Source vetting, fact-checking, and other forms of curation are largely invisible acts, and a great deal of trust must be placed upon the information distributors that make claims of trustworthiness. Again, the assumption is that the mysterious 'system' is always correct. To Bozdag, the opposite is true. The syllogism would read: 'Every action or inaction by humans is biased / all technology is created by humans / therefore all technology is influenced by humans'. This concern is undoubtedly heightened by the changing role of audiences from passive consumers to gatekeepers, as well as becoming as producers themselves. As the masses now shape media in a more direct, instantaneous way due to the rise of collaborative filtering, they can also boost unconventional or fringe voices that previously were not accessible on mainstream media--from amateur filmmakers to conspiracy theorists and everything in between. This is what Bozdag means when he says "People affect the design of the algorithms, but they also can [...] manually influence the filtering process after the algorithm has been designed" (2013, p. 224); if systems make use of user metrics to help serve relevant content, the audience as a whole then has the power to draw attention towards non-traditional media sources.

Of those non-traditional sources, some lack the professional rigour required by more reputable media. What is referred to as algorithmic authority in the previous chapter has a role here too, as the assumption that an information retrieval algorithm has already assessed and vetted the sources it returns influences consumption habits. If the information retrieval algorithm is perceived as having the same standard of factual verification as an established medium like a reputable newspaper, the misconception is further complicated by personalization: Most algorithms are programmed to filter content down to what users are most likely to engage with, de-prioritizing values like reputability and impartiality. Similarly, the use of collaborative filtering in an algorithm prioritizes the viewing patterns of similar users. Guided by engagement metrics, collaborative filtering and personalization undermine professional, authoritative sources and prioritize the voice of the layperson. Because a layperson does not have the skill and knowledge of an expert, there is an increased chance that misinformation may be present:

> Some individuals, who are more information-savvy, will automatically occupy strategic positions to facilitate access to information to others. Depending on the subject matter, not everyone in a group is equally important or qualified in providing information. Those who have more knowledge will act as gatekeepers. I might trust John's competence in football, and use him as my gatekeeper in this subject, but not in the area of international politics. However, in most online services, we get to see everything published by a user, or nothing at all. We need mechanisms to assess the competency of the information sharer and determine the needed gatekeeper for a given context. (Bozdag, 2013, p. 222)

Both Google and YouTube, for example, have had to adjust their algorithms in order to return more authoritative results related to news topics (Alvarez, 2017). As audience members have become gatekeepers for themselves, the role of the journalist has changed to that of a 'gatewatcher', whereby journalists must evaluate widely used sources instead of determining who is promoted in the first place (Bozdag, 2013). The point being, gatekeeping theory provides another perspective from which to view how new technologies mediate the flow of information. In the internet age, a traditional paradigm is upturned and actors must adjust to new roles.

*Application*

On YouTube, the gatekeeping conversation turns to the implications of collaborative filtering on authoritative content. As Bozdag points out, people are more likely to trust automated systems than manual ones, and the information environment of YouTube is no different. Using this theory, it could be understood that suggestions made by the recommender system are interpreted differently by users than manual searches for specific videos. What's more, YouTube allows the audience to become creators while also influencing that automated system simply by using it. Views, clicks, time-watched, and other metrics are captured from every user to better determine what similar viewers might wish to watch, meaning a transfer of power in media distribution. The consequence is a move towards a sort of populism of distribution, as well as a move away from authoritative sources. The question of who controls the flow of information, and how that impacts which events and topics are considered more or less newsworthy than others is outsourced to an algorithm that is very much directed by the audience it serves--insofar as YouTube deems appropriate. The alt-right and other alternative media on YouTube in particular take advantage of this newfound power, and "explicitly define themselves in opposition to mainstream structures of knowledge production" (Munger & Philips, 2019, p. 2), posturing themselves as purveyors of truth who stand against dominant (and mysteriously orchestrated) media narratives.

In a way, advertisers can also act as gatekeepers on YouTube. Participating in the Partners Program, users are able to let advertisers place commercials before and/or during videos, provided that YouTube finds the content palatable for advertisers. As many creators rely upon advertising revenue to continue producing videos, it is possible to construe both YouTube and advertisers as being restrictive to the information infrastructure on the platform. After Swedish YouTube star PewDiePie (Felix Kjellberg) posted a video containing an anti-Semitic comment, the *Wall Street Journal* picked up the story and PewDiePie lost his Disney sponsorship (Winkler, 2017). This incident caused a ripple effect across YouTube, resulting in advertisers becoming overly cautious when it came to running ads on the platform (Alexander, 2019). In response, YouTube started to aggressively remove the ability to serve ads before or during videos that could be seen as problematic or brand-unsafe. In the YouTube community, this demonetization epidemic is known as the 'adpocolypse', and curtailed the earning potential for all YouTube creators. In 2018, YouTube again drastically decreased the amount of videos eligible for monetization after another star YouTuber named Logan Paul released a video containing a deceased body. Before this incident, any channel with 10,000 views was able to apply for the Partners program. Now, in order to participate in YouTube's Partner Program and qualify for monetization, creators must reach 4,000 hours of watch time over 12 months and at least 1,000 subscribers (Kyncl & Mohan, 2018). While users are still able to post guideline-abiding videos to the platform without restriction, it is still possible to consider demonetization as an impediment to distribution, therefore implicating YouTube and the advertisers they serve as gatekeepers.

**Social Shaping of Technology (SST) & Social Construction of Technology (SCOT)**

Technological determinism is a school of thought that suggests society is molded by technology. Social shaping of technology (SST) theory argues the opposite: Technology is shaped by society. To quote Dierkes and Hoffman, "Technology is deeply affected by the context in which it is developed and used. Every stage in the generation and implementation of new technology involves a set of choices between different options" (OCSE, 1998), meaning the development of technology does not occur in a vacuum, nor is that development necessarily logical or predictable. As a result, there are various choices that can be made during innovation that can influence how society is ultimately affected. MacKenzie and Wajcman, authors of *The Social Shaping of Technology* expand on this idea, stating that "As a simple cause-and-effect theory of historical change, technological determinism is at best an oversimplification. Changing technology will always be only one factor among many others: political, economic, cultural, and so on" (MacKenzie & Wajcman, 2011, p. 4). The complexities of a contextual understanding of the development of technology cannot be overlooked as markets can be created where no need exists; similarly it is naive to suggest that consumers will make practical or informed choices. In other words, there is no predictable A to B development process for any technology, as a variety of factors introduce uncertainty. Factors like access, usability, and cost can inhibit the growth of a superior technology--it can be assumed that products are generally tempered by social context in order to maximize sales. In the SST model, technology may have consequences on society, but is not developed in an environment free of human influence.

Similarly, social construction of technology (SCOT) also rejects technological determinism and suggests that there are degrees of flexibility in how technology influences society; social context can create different benefits for different social groups than originally intended. In contrast to SST, influence occurs during the adoption of technology after it is released, when users decide how it is used. In essence, SCOT defines user agency as a major factor in technological development.

*Application*

Of interest to this thesis is SST's positioning away from the reactive design philosophy that dominates many technology-centric fields. Instead of fixing existing technology, the attitude must be one that recognizes the power held by social forces on the development of a new technology: "The academic interest in the forces which shape the generation and implementation of new technologies has been fueled by the prospect of moving beyond defensive and reactive responses to technology toward a more active role" (Dierkes & Hoffman, 1992). MacKenzie and Wajcman reiterate this plea: "The view that technology just changes, either following science or of its accord, promotes a passive attitude to technological change. It focuses our minds on how to *adapt* to technological change, not on how to *shape* it" (2011, p. 5). In a similar way to previous conversations about human bias in technology, the creators of new technologies need to acknowledge their influence and use it to address societal needs during product design. In the case of YouTube, one can consider the primary objectives of the system: YouTube does not use co-visitation metrics to decide how videos are promoted simply because this is the way the technology works, these were conscious choices that were made in the development process. If technology is in some ways shaped by society, it becomes easier to understand why recommender systems echo the information preferences (e.g. instant gratification, passive consumption) of the specific populations using them. Radio listeners shaped television, television viewers shaped internet video. YouTube is successful because it is able to capitalize on natural human behaviour and the information preferences of its user base, including the combination of the always-on spoon-fed automation of TV with the unending personalization of the internet; together they create a feed of custom content that does not require cognitively complex interaction. Unlike the technological determinist, the SST theorist is aware of the two-way relationship between technology and society: "It is mistaken to think of technology and society as separate spheres influencing each other: technology and society are mutually constitutive" (MacKenzie & Wajcman, 2011, p. 23). Therefore, it is not only that society shapes technology, and technology in turn shapes society, but that platforms like YouTube *are* society. In a recent *NiemanLab* article, Dr. Mike Ananny problematizes this flawed distinction: "Platforms are societies of intertwined people and machines. There is no such thing as 'online life' versus 'real life.' We give massive ground if we pretend that these companies are simply having an 'effect' or 'impact' on some separate society" (2019). This thinking disrupts common narratives regarding YouTube's influence on users; it is a mistake to only consider the impact of the recommendation system on election outcomes or acts of terrorism--YouTube users are members of society and their activity online is no different than other activities in their lives.

Using the SCOT model, it becomes possible to look at how YouTube has adapted in response to the ways YouTubers used it. As a concrete example, Bucher (2018) argues that one of the reasons that YouTube's algorithm was tweaked in 2012 to prioritize watch-time over clicks was in response to the disruptive force of the 'reply girls', which Bucher likens to the 'feminist killjoy'. These young women were YouTube creators who exploited the recommendation algorithm by using thumbnails focused on their cleavage in order to drive clicks. In 2012 YouTube offered a feature that allowed for 'video responses', which could be found in the comment section of every video. The idea was simple: The more popular a reply video was, the more likely the RS was to promote it in the related and recommended video areas. Using this knowledge, the reply girls were able to routinely dominate the recommendation sections of most popular or trending videos by using suggestive thumbnails on their video replies. Around the time these types of videos had reached their peak in March 2012, YouTube announced that videos would no longer be recommended on the basis of how many clicks they received, specifically citing misleading thumbnails as one of the reasons for the change ("Changes to Related", 2012). This algorithmic adjustment serves as an excellent example of SCOT theory as it was a technological change undergone on the basis of social forces of both audience and creator.

### The Medium is the Message

First published in 1964, Marshall McLuhan's work *Understanding Media: Extensions of Man* was a theoretical exploration of what shaped media, and how media in turn made society. Perhaps the phrase he is most well-known for is "The medium is the message". By this, McLuhan means that the functions of a medium or technology itself pose more significant change to the people that use it than the content that it carries. For example, McLuhan urges readers to consider the introduction of the machine: "In terms of the ways in which the machine altered our relations to one another and to ourselves, it mattered not in the least whether it turned out cornflakes or Cadillacs. The restructuring of human work and association was shaped by the technique of fragmentation that is the essence of machine technology" (McLuhan, 1968, p. 8). Just as products shaped society less than the production of said products, content changes society less than media as information distribution systems.

The impact of a medium outside of content has many implications for our understanding of newer media like the web, for as McLuhan says, "The 'message' of any medium or technology is the change of scale or pace or pattern that it introduces into human affairs" (1968, p. 8). The example cited is the railway, which as a technology did not invent human movement and yet monumentally altered the development of cities and other settlements by way of long-distance travel and transport of resources. Extending this line of thinking, one could say that the grain, passengers, or other cargo transported by the train certainly had influence wherever the rails led, but the ability to efficiently move those objects in the first place is exceptionally more impactful on society. Media is no different; and criticism is all too often levied against objectionable content rather than the methods and structures that uphold them. "All media work us over completely" McLuhan says (Fiore & McLuhan, 1967, p. 26). That is, the medium will change human behaviour regardless if the consumer is aware, just as the immediacy and visual nature of the television spawned TV dinners and prompted the reorganization of schedules in a way that portable radio never could. As a lens, media shift the way we observe the content they contain. As McLuhan notes, "The effects of technology do not occur at the level of opinions or concepts, but alter sense ratios or patterns of perception steadily and without any resistance" (1968, p. 18). As a result, analysis of any content on YouTube is incomplete without considering the medium that bears it.

*Application*

As McLuhan died in 1980, he wasn't able to apply his philosophy to newer media like personal computers, the internet, and algorithms. If society has built technology in their own image, then it is possible that we have inadvertently recreated the flaws in our behaviour. Naturally, individuals seek pleasure; rarely is there a surface preference for the unpleasant things that might do some good. Recommender systems do much of the same, usually serving that delicious, sugary food it believes is loved. But when computers start to make more and more decisions about what is and what is not relevant to personal interests, questions arise about *best* interests, the broccoli and asparagus of information. As McLuhan would say, "Too much of anything, however sweet, will always bring the opposite of what you thought you were getting" (McLaughlin, 2003). In many ways, McLuhan predicted the recommender system, and imagined that "a computer as a research and communication instrument could enhance retrieval, obsolesce mass library organization, retrieve the individuals' encyclopedic function and flip into a private line to speedily tailored data of a saleable kind" (McLuhan, 1989). Similarly, one could imagine McLuhan musing over the future of television. In many ways, he was able to predict online video: "The next medium, whatever it is — it may be the extension of consciousness — will include television as its content, not as its environment, and will transform television into an art form" (McLuhan, 1967). What if TV, as a stream of synesthetic media, was personalized for the individual and not packaged for the masses? This future has arrived, and in more than one form. The ideal example here is YouTube's autoplay service, which, despite all its flaws, has been rather successful at guessing what users may want to watch next. Like TV, content is also shaped around advertising: YouTube's 2012 decision to prioritize watch time over clicks in the process of making recommendations led to creators making longer videos, a response that gave creators the ability to serve more ads, thereby increasing profitability. But the difference maker is not YouTube as a platform, it is the algorithms behind it that imperceptibly reshape and narrow the conception of reality by placing the blinders of filtering over users' eyes. Observing that children's eyes remain fixed on the faces of those on television, McLuhan is prompted to state that "TV is not so much an action, as a re-action, medium" (McLuhan, 1968). This has become blatantly true in the age of 'react videos' and 'let's play' formats. While the previous generation passively received programmed television, the internet denizen suffers from an overload of information. The resulting synthesis replaces previous technology: YouTube offers a stream of television that is catered for you, personally.

Finally, McLuhan's thoughts on automation also have a bearing on the collaborative filtering methods of the present: "Anybody who begins to examine the patterns of automation finds that perfecting the individual machine by making it automatic involves 'feedback'. That means introducing an information loop or circuit, where before there had been merely a one-way flow or mechanical sequence." While McLuhan is certainly not the first to make this observation, it is interesting to consider how a measurement of co-visitation on YouTube is essential to the automatic operation of the system that can be compared to an enormously complex reader's advisory service. To successfully curate information, information about that process must be obtained for fine-tuning.

## Conclusion

Considering the theories about human reasoning that underpin information preferences, it comes as no surprise that humans are easily manipulated actors and that media structures have been built precisely to exploit those behaviours. The theories covered here all propose the same notion: Humans shape technology, which in turn, influences society. Preferring confirmatory information, individuals turn to gatekeepers to provide that information. As traditional models of gatekeeping become upturned with the changing roles of creators and consumers, society must readjust to new concepts of authority, including algorithms. The design of these systems is then influenced, often invisibly, by the ebbs and flows of information consumption patterns. If the medium is the message, then the exploitation of natural behaviour to the detriment of society is the shaping force of our age: The recommender system is an evolution of media and deserves critical study in the face of a rise in extremism. A model example is YouTube, a tool representing “the true democratization of political media in the medium that has consistently proven the most popular and most powerful” (Munger & Philips, 2019, p. 1).

# CHAPTER FIVE: METHODOLOGY

## Introduction

In early 2018, the British news outlet *The Guardian* worked with former Google employee Guillaume Chaslot to release the results of an investigation into the videos YouTube was recommending to viewers during the 2016 American election. Their findings were remarkable: 84% of recommended videos that the researchers deemed partisan were supportive of Donald Trump, while only 16% were supportive of Hillary Clinton. According to the authors, the goal of this study was to make use of aggregate data to help reveal how YouTube promotes some videos on its platform above others and why. Here, 'how' is a technical question of algorithmic design and is dependant on the 'why', or the value judgments YouTube must necessarily make on the content it provides a platform for. By attempting to understand this system more fully, the researchers intended to expose the 'unintended biases or distortions' found within the maze of code (Lewis & McCormick, 2018). Building on this, the methods used to answer my research questions were adapted from the work done by Guillaume Chaslot and staff of *The Guardian*. My original contribution includes extending and enhancing their methodology to better suit my needs and provide further context to the sources present in the recommendations. In this chapter, I will discuss the methodology used by Chaslot and *The* Guardian staff's original study about YouTube's recommender system and identify potential gaps in their approach. Drawing on the discussed methodology, I will then describe the methodology used for this work. Given the limitations of a master's level thesis, some aspects of *The* Guardian's study will not be replicated, either due to technological restrictions or time constraints. In addition, I ask different research questions than those explored by *The* Guardian which require different methods to answer. Consequently, I have altered, tailored, and streamlined data collection processes and other tasks, including the implementation of a consistent external reference for bias determinations. Other approaches were added to better address the questions being asked. For example, an analysis of sources that were recommended alongside searches for Canadian politicians took place with a focus on the alt-right, as defined in Rebecca Lewis' Alternative Influence Network report.

## *The Guardian*'s Research and Methodology

### *Overview*

As previously outlined, an investigation into YouTube video recommendations drawn from searches for the Republican and Democratic nominees around the time of the 2016 election in America was released by *The Guardian*. Overall, the study found that YouTube was promoting more videos that supported Donald Trump than videos that supported his Democratic Party opponent, Hillary Clinton. According to the news outlet, the methodology used to produce data from YouTube's recommendation algorithm was developed by ex-Google employee Guillaume Chaslot, who contributed by creating software to compile a large database of videos and their corresponding URLs for further analysis (Lewis & McCormick, 2018). The work to draw conclusions from the data was then completed by staff of *The Guardian*, Paul Lewis and Erin McCormick. *The Guardian*'s process can be broken down into the following steps:

1. Data collection
2. Broad analysis
3. Focused investigation
4. Ranking
5. Cross-referencing

### *Data Collection*

From 2010 to 2013, then 31-year-old Guillaume Chaslot worked for Google (before the 2015 restructuring into Alphabet), where he spent some time working on YouTube's recommendation engine. According to a spokesperson at YouTube, Chaslot has "misrepresented his position" at the company (Smith, 2019). Chaslot's knowledge of algorithms is well rounded: As well as having experience with programming, Chaslot also holds a Ph.D. in artificial intelligence. His expertise and knowledge were contested by Google who fired him in 2013 over 'performance issues'. The reasons for his dismissal were contested by Chaslot, who claimed that he was let go for "agitating for change within the company" (Lewis & McCormick, 2018). This change, Chaslot says, was the need to promote diverse content to YouTube audiences.

To help researchers investigate the hidden workings of the YouTube algorithm, Chaslot set out to develop software that mimicked an average traversal of the video-sharing website.

Like many users, the program takes an initial starting video and follows the path laid out by the recommended videos. Each 'Up Next' video has its own set of recommended videos, and so on. Using a keyword to obtain the first video or 'seed', the program then moves to the first recommended video promoted in the sidebar and repeats the process thousands of times, each iteration storing valuable information about the video and its associated recommendations in a structured database. It is important to note here that the software avoids the problem of personalized results by instead surfing the website without an associated profile. Doing so, the software can peel back layers of user-affected recommendation to find what YouTube is natively promoting. Although this trait prevents researchers from accurately observing the amplifying effects of personalized filters, it does make results more generalizable; viewing habits likely vary greatly between users.

When used in *The Guardian*'s probe, the software was used to search for the keywords 'Trump' and 'Clinton' in 2016, alternating between the two to so that both terms were equally represented. Over the course of three distinct weeks (22 August; 18 and 26 October; 29-31 October; and 1-7 November), videos were collected and added to a database containing URLs, video titles, the original search term used (Trump/Clinton, Clinton, and Trump), the name of the YouTube channel the video was posted by, the number of views, likes, dislikes and the depth at which the video was found. On some dates, Chaslot took the first five videos initially recommended from a search, recorded the first five recommended videos from those seeds, and repeated the whole process five more times. Although *The Guardian*'s methodology report does not indicate why, Chaslot did in fact intentionally choose to alter the collection method on some of the days. For example, the report states that sometimes the method was adjusted to capture the first three or four videos and capture the first three or four layers of recommended videos before repeating the entire process six times. It can be hypothesized that Chaslot intended to add some degree of variability to his program in order to produce more diverse results. In the end, Chaslot's software combed YouTube to the tune of 8,052 videos, which were all made available publicly on a Google Sheet.

In order to facilitate future research, Chaslot has packaged his software into a small Python program and made it freely available on GitHub. Additionally, he has also used the program to feed into his website, algotransparency.org. Much like the study completed by *The Guardian*, his project collates the most recommended videos on a range of topics every day, including world leaders, science, and specific elections. To aid more direct inquiries, Chaslot has also added the ability to search for keywords on any given day and will return the most recommended videos for that day with that term in the title, including the number of channels that are recommending the videos.

*Analysis*

After the database was populated by Chaslot and his software, the results were sent along to the journalists Lewis and McCormick at *The Guardian* for analysis. This study included a broad investigation into all 8,052 videos and a more focused look at the 1,000 most-recommended videos in the database. 'Most-recommended' is defined as the videos that appeared on the most dates as well as the videos that were promoted the most times in the 'up-next' sidebar. An entire third (roughly 2600) of the 8052 videos were removed from the study due to being unrelated to the election, a designation defined by the authors as "politically neutral or insufficiently biased to warrant being categorised as favouring either campaign". After these videos were removed, the remaining two-thirds of the videos related to the election were analyzed using content analysis. Next, the researchers took the 500 most-recommended videos returned by the keywords 'Trump' and 'Clinton' and analyzed them for political leaning. In perhaps a crude-yet-effective way, the 'obviously partisan' nature of each video was determined by watching each video and considering their titles. Unfortunately, roughly half of the 8,052 videos were no longer accessible at the time of the analysis. The researchers say this is either due to voluntary deletion/privatization, or removal from the platform by YouTube or copyright claim. For the sake of accuracy, *The Guardian* chose to exclude these missing videos from the first round of their analysis. Later, a second analysis took place that included all the missing or deleted videos, this time making use of their titles to determine whether the videos could be considered beneficial to one campaign over the other. From the description provided in *The Guardian*'s methodology article, they position the sorting decision as having to be relatively clear, citing example titles such as "This Video Will Get Donald Trump Elected" and "Must Watch!! Hillary Clinton tried to ban this video".

After this analysis, researching staff at *The Guardian* took all 8,052 videos in the database produced by Chaslot and ranked them according to their number of recommendations. Recommendation is defined as the number of times a single video is recommended by other videos in addition to the number of days a video is recommended. The example given is a video that appears on the 'up-next' sidebar of four different videos and has been recommended by YouTube on 3 different dates. This video in this example would have a recommendation score of 7. If videos were recommended multiple times in a single day by the same channel, this would only count as one recommendation. As an aside, I believe this decision works against the goal of determining how much a video is promoted overall and may distort the results unfavorably.

Using this methodology, Lewis and McCormick were able to produce a list of the most recommended 'obviously partisan' videos born from searches for 'Trump' and 'Clinton'. Additionally, they were able to find out which channels were most recommended using the same criteria. Using the search keywords 'rally' and 'speech', they were able to identify videos containing full campaign speeches by Trump, Clinton, and various other political figures and their family members. Finally, Lewis and McCormick shared their database and findings with a data analytics firm called Graphika, who have specialized in political disinformation campaigns in the past. Using its unique Twitter database, Graphika was able to see how the videos identified in *The Guardian*'s investigation were shared across the Twitterverse, hoping to find evidence of 'automated activity', or a concerted effort by networks of bots to push certain videos to a large audience.

*Flaws*

However helpful Chaslot's software may be to future researchers such as myself and others, it does have its limitations. For example, the software does not record the specific date and time each search was completed, nor does it automatically export spreadsheets of the most recommended videos. While functional and customizable, the lack of UI may be an impediment to researchers lacking Python experience. In the final chapter, I will expand upon my experience with the software. *The Guardian's* analysis has its flaws as well: Some terms are not adequately defined in their methodology. For example, the qualifier 'obviously partisan' is underdeveloped, and clearly needs a more rigid definition. Further work should make explicit what is considered partisan and non-partisan, for multiple researchers may come to differing conclusions on the term's application. This limitation was indeed identified by *The Guardian* themselves, who admitted that the process was 'subjective' despite noting the surprising ease in which researchers were able to categorize videos. An observation worth making here is that *The Guardian*'s study is not academic, and although this is not necessarily a negative characterization, it is fair to suggest that their approach was more focused on results rather than context or methods. An example of this difference is the choice to only use partisan videos in their analysis. While it is more eye-catching to suggest that 86% of partisan videos were beneficial to Trump and only 14% to Clinton, these findings exclude the larger field of non-partisan videos that present the politicians in a more neutral light.

Another limitation of this study is the time period elapsed between when the data was gathered by Chaslot and when the actual analysis took place. Although it is not precisely clear when Lewis and McCormick's analysis took place, it is clear that some significant time had passed between the two stages, for the sheer number of missing videos clearly impacted their ability to conduct a thorough investigation into content before it was removed. By collecting the data on a single date and beginning the analysis as soon as possible, Chaslot, Lewis, and McCormick could have had a much cleaner data set to work with and therefore information to draw upon--potentially providing more detailed answers to their questions. Finally, the design of the algorithm itself may have changed since Chaslot's employment at YouTube, making some of his claims irrelevant. As quoted in a *PC Magazine* article, a spokesperson for YouTube stated that "no part of the recommendations system we had in place when Mr. Chaslot was an employee of Google is in use in the YouTube recommendations system today" (Smith, 2019).

## Adapted Methodology

### *Overview*

As discussed previously, my own methodology was built off *The Guardian* and Chaslot's methods. Exploring similar questions using the same software, their study is a useful model on which to think about my own. As I will outline here, the methodology has been expanded to answer new questions I have chosen to explore in my own research. Considering *The Guardian's* methodology used, my personal limitations, and necessary alterations, the following outlines the methodology used in this investigation into YouTube's recommendation system in the context of the 2019 Canadian election.

My process can be broken down into the following steps:

1. Data Collection
2. Source Analysis
3. Bias Analysis
4. Ranking

Similarly to *The Guardian*, I have divided the analysis into two parts: a primary and secondary analysis. However, this choice was made for different reasons; while Lewis and McCormick were compelled to conduct a secondary analysis in order to account for missing videos, I have completed a secondary analysis intentionally to separate questions of political bias from an analysis of the sources recommended alongside the original searches for the politicians. The first analysis is conducted on the channels hosting the recommendations, and categorizes them into simple source categories, including the factual accuracy and political leaning of the channel the video was posted by, as well as whether the channel is associated with the alt-right. The second depiction analysis focuses on how each candidate is represented in relevant videos: either positively, negatively, or neutrally.

*Objectives*

This methodology is crafted to collect and analyze data that will help to answer a primary question: In what ways does YouTube's recommendation algorithm promote videos that support one or more candidates in the 2019 Canadian federal election over another? In addition, what is the political leaning and factual reporting levels of the sources being recommended alongside those candidates? Are any of those channels associated with the alt-right? By applying relevant theory to the research questions, I can begin to determine the appropriate data to collect and how. Knowing that the algorithm behind the online media juggernaut is full of unknowable, moving parts, it is at first daunting to imagine how that data can be transformed into something useful to address the questions.

Similarly to the approach taken at *The Guardian*, I started with the three party leaders for the 2019 federal election: Justin Trudeau for the Liberal Party, Jagmeet Singh for the New Democratic Party, and Andrew Scheer for the Conservative Party. The names of these three party leaders served as my base search strings. These names were used to collect seed videos that pertained to the candidates using Chaslot's software. From these seed videos, the appropriate metadata was collected, including video titles, URLs, and channel names. Next, I collected the metadata on the seed's recommendations. Additional data collected included the date and time that I collected these videos.

The objectives of the source and depiction analysis are similar but decidedly different. Whereas the purpose of the source investigation is only to take the temperature of YouTube's recommendations at the particular moment of the data collection, the express goal of the depiction bias study is to attempt to answer the outstanding thesis question. While questions like 'What kind of sources is YouTube promoting when searching for these terms?' can be answered with the source analysis, it is only the investigation into depiction that can attempt to answer questions about the algorithm's role in potentially supporting one candidate more than another. While both source and depiction analysis objectives are separate issues, they are also highly interconnected, as demonstrated by *The Guardian's* research.

<u>*Data Collection*</u>

In order to collect the metadata, I used Guillaume Chaslot's open-source software YouTube-explore, a Python-based program that crawls YouTube by following search strings to seed videos and the associated recommended videos. Although it was originally created to collect data for *The Guardian*'s study, Chaslot has made the program freely available on GitHub for anyone to download. According to the documentation provided, the software follows these instructions:

1. Obtains the N first search results
2. Follows the first M recommendations
3. Repeats step (2) P times
4. Stores the results in a JSON file

The variables N, M, and P are set by the user in the initial search in python. The following is the query used in this study, ran once for each candidate:

```
python follow-youtube-recommendations.py --query="andrew scheer" --searches=5 --branch=8 --depth=4 --name="2019FE"
```

Breaking this command into its respective pieces, each plays a somewhat self-explanatory role in the search process. "python follow-youtube-recommendations.py" simply directs Python through Windows Powershell to Chaslot's software. 'Query' are the search strings directing the program, separated by commas. This field is case-insensitive. 'Searches' indicates the number of search results from the initial query (in this example, 5). If, for example, I were to search for "Justin Trudeau" setting the 'searches' to any number I choose (represented by the variable 'N'), the first N videos retrieved by YouTube in the search page would be used for the analysis. 'Branch' refers to the number of recommendations that are followed from one video and is represented by the variable M. Again, these recommendations are found on the right-side of the video currently being watched. A 'depth' of P would then designate how many times to repeat the process of digging deeper into YouTube's recommendation rabbit-hole. Other variables include ‘name’, which sets the output filename, and the optional ‘alltime’, which sorts the initial search by all-time most viewed videos. The query used above returned exactly 1000 videos per candidate at variable depths, pulling in a grand total of 3000 videos. Each title is followed by the URL only of 19 associated recommendations. By using a high branch number (8), and a relatively low depth number (4), I have selected the data collection to include a more shallow and broad sampling of recommendations, as opposed to a deep and focused exploration of which videos might be suggested much further into a browsing session. The data collected will not include videos recommended more than 4 levels deep, lending to a more typical browsing experience. The following table (5.1) presents an overview of the data collected, including the date of collection and how many videos were collected at each depth.

| SEARCH STRING | DATE | DEPTH 1 | DEPTH 2 | DEPTH 3 | DEPTH 4 |
|---|---|---|---|---|---|
| Andrew Scheer | 2019-02-16 | 900 | 88 | 8 | 4 |
| Jagmeet Singh | 2019-02-16 | 893 | 101 | 5 | 1 |
| Justin Trudeau | 2019-02-16 | 888 | 105 | 4 | 3 |

*Table 5.1: Data Overview*

After YouTube-explore has been run, the program outputs the data as a JSON file. When using a JSON viewer or similar software, the data is neatly organized in a tree structure, beginning with a category for each search string. Each video included in the entirety of the search process is then laid out in numbered sub-categories, which themselves contain metadata about the videos, including views, likes, dislikes, the title, the depth at which the video was retrieved, other videos recommended on that video’s page, and the name of the channel. Additionally, ‘nb_recommendations’ serves the number of times the video has been recommended in that set, and ‘mult’ calculates the average amount (in percentage) a given video is recommended. For the purposes of coding, each JSON file was exported to CSV file and uploaded to a Google Sheets document, sorted by number of recommendations, and coded using the parameters described in the analysis section. As represented in the CSV file, each row in the data represents one recommendation.

*Inclusion & Exclusion Criteria*

Precautionary measures were identified to ensure clean data was collected. These measures included carefully considered decisions about how the study was carried out, or more specifically, what data should and should not be included. For instance, the names of the respective political parties (Liberal Party of Canada, Conservative Party of Canada, and the New Democratic Party) were not used due to their more common reference as the ambiguous terms 'liberals' and 'conservatives'. Instead, the full names of the three candidates were used in the search to disambiguate the candidates from other unrelated persons. Another factor to consider is temporal; to ensure a uniform collection environment, all data were collected in a single day: February 16th, 2019. Although Chaslot's software allows for the initial seed videos to be sorted by 'most-viewed', YouTube's default determination of search string relevancy was maintained in order to stay consistent with the core objective of the study which is to observe not what people are watching on YouTube, but rather, what the algorithm suggests that they watch (this distinction unravels later).

*Analysis*

*Source Analysis*

Similarly to *The Guardian*'s methodology, a broad analysis of sources was first conducted on all 3000 videos (1000 per candidate). Two existing designation sources were used: The website Media Bias/Fact Check (MBFC) provided factual reporting and political bias ratings for many major YouTube channels, and Rebecca Lewis' Alternative Influence Network (AIN) report provided a list of 65 alt-right-affiliated YouTube individuals and their associated channels.

The source analysis itself was a multi-step process, consisting of converting, staging and coding. A simple online conversion tool was used to transform the JSON output to a CSV file that could be opened with Google Sheets. Staging the data retrieved from Chaslot's software was the least complex step and only involved rearranging and adding various columns of metadata to prepare it for coding. The column containing the channel's name was moved beside the column containing the name of the video, and additional columns were added for alternative influence designation, political leaning ('bias'), and level of factual accuracy ('factual reporting'). To ensure uniform data input, Google Sheets' cell validation functionality was used: For example, all 65 individuals identified in Rebecca Lewis' AIN report were added to a drop-down list within the sheet in order to avoid potential transcription errors.

Coding began with a comparison of Lewis' list of AIN members with the recommended channels in each dataset. This process is intended to measure the 65 extent to which members of the alt-right and alt-lite are recommended on seed videos. In chapter three, I criticize the methodology Lewis' uses to group her network, stating my belief that collaboration between two individuals does not necessarily signify homogenous thought. Conversely, the algorithm does not make this distinction. As such, I considered removing the Twitch streamer 'Destiny' from the AIN list as his strong ties with the other figures on the list was based on confrontational interactions (Munger & Phillips, 2019). However, further thought on the subject brought me to the conclusion that regardless of engagement or pushback against the harmful ideas that many members of this group identify with, platforming (in this case, publicly conversing with) a member of the AIN allows that person a greater audience. As a result, I did not remove the channel 'Destiny' from the AIN list. However, another alteration was made: Several of the figures (Gavin McInnes, Lauren Southern, and Faith Goldy) listed in the report were associated with the right-leaning Canadian media outlet *Rebel Media*, and thus the channel itself was added to the list of considerations. To determine if alternative influencers were recommended in my datasets, the 'find' function was used on the Google Sheets dataset using the channel names associated with the 65 identified individuals.

Finally, work was done to code the media sources as they are represented on the website mediabiasfactcheck.com. On this website, a team of 10 employees use a strict methodology to determine the bias and level of factual reporting of thousands of media sources, from large news outlets to small, independent firms. While structured, portions of the methodology are characterized by MBFC as "rather subjective". Other criticisms of the website come from the *Columbia Journalism Review*, who describe it as "amateur". The *Poynter Institute* similarly criticizes the watchdog for lacking a scientific methodology. That being said, the rankings provided by the website have been employed by researchers at both the University of Michigan and the Massachusetts Institute of Technology to create tools associated with media bias tracking. In addition, Ribeiro et al. (2019) used the websites' ratings in their in-depth study into the radicalization of YouTube users, giving the source further legitimacy. The use of the ratings in this thesis is not done without caution or criticism, and it is acknowledged that any attempt to grade the political leanings of a given news organization is necessarily a subjective task.

MBFC uses a 10-point scale to rank media sources, divided into 4 major categories:

1. Biased Wording/Headlines

2. Factual/Sourcing
3. Story Choices
4. Political Affiliation

Part of this determination is made by considering bias by omission, by labelling, by placement, by selection of sources, by spin, by story selection, confirmation bias, connotation, denotation, loaded language, and finally what they call 'purr' and 'snarl' words, or descriptive language that might indicate bias. A 'purr' is a word used to describe something that is favored, while a 'snarl' describes something negative. Political examples include the word 'democracy' as representing something positive, and the word 'fascist' as characterizing someone or something negatively (Hoffmann, 2005). Once each category has been ranked on the 10-point scale, the scores are added and divided by four. This final score places the source within one of four categories. A score of 0-2 indicates that the source is one of the least biased, while a score of 2-5 is designated as having a "Left/Right Center Bias", depending on the political affiliation of the source. Scores falling in the 5-8 range are classified as having a Left or Right bias, while sources in the final 8-10 bracket find themselves in the Extreme bias category.

The 'Factual/Sourcing' category used to make the media bias determination is similarly segregated into 5 sub-categories from 'Very high' all the way down to 'Very low', which MBFC uses to provide a 'Factual Reporting' grade. While inescapably susceptible to human interpretation, this methodology provides me with a large database of media sources and their associated biases. To determine the grade for each source, MBFC reviews a minimum of 10 headlines and 5 news stories using a numbered scale from 0 to 10. A score of 0 indicates a Very High factual accuracy rating: The source has never failed a fact check, sources to credible information, and releases immediate corrections where necessary. A source with a score of 1-3 is considered to have a High level of factual accuracy as it is almost always factual, makes immediate corrections, and has only failed 1 fact check. In addition, MBFC requires sources graded as having a High level of factual accuracy to use "reasonable language that retains context" as well as sourcing to mostly credible low-biased information. A source with a score of 3-4 will be given a Mostly Factual rating if it occasionally uses biased information sources and may have failed a fact check and do not correct mistakes in a timely manner. Additionally, the source is generally pro-science but may use misleading headlines. MBFC is still working towards implementing the Mostly Factual rating and as such this grade cannot be found in my coded data. Moving on, a score between 5 and 6 results in a Mixed factual accuracy rating, a designation denoted by a tendency to source information from other biased or unreliable sources, do not correct inaccuracies, or do not support scientific consensus topics like climate change, GMO's, vaccinations, and evolution. Any source that employs "extremely loaded language that alters context of facts" will automatically be considered as having a Mixed factual accuracy rating, regardless if it has failed a fact check or not, as will any source that fails to include a mission statement or ownership information. A Low factual accuracy grade is determined by a score of 7-9 and is given to a source if it rarely uses credible information sources. Sources within this range may be re-graded as containing fake news, conspiracy content and propaganda. Finally, a Very Low factual accuracy rating is given to sources with a score of 10, outlets that almost never use credible information sources and cannot be considered reliable in any way. The Fake News rating specifically denotes hoax websites and hate groups. The Conspiracy/Pseudoscience grade is given to sources that related to "known conspiracies" including the new world order, the illuminati, false flags, aliens and more. Otherwise, the source must include unverified health or scientific claims. The methodology used by MBFC to determine which sources classify as Propaganda is unclear.

During the next step of the coding, I manually added MBFC's factual reporting and bias ratings for each possible source in each dataset. While many of the videos did not originate from media sources, the ones that did received the corresponding MBFC designation in the spreadsheet. I attempted to make automatic calls from the spreadsheet to the corresponding page on MBFC's website, but I soon abandoned this task as various complexities came to light. First and foremost, my testing revealed that MBFC did not use a consistent data structure across their pages. I sent an API request to the webmaster but did not receive a response. Secondly, not all channels were listed consistently in my datasets. To rectify this, several sources were combined. For example, a search for BBC included subsidiaries, including BBC Newsnight, BBCPanorama, etc. After coding was complete, I was left with a dataset that identified channels associated with the alternative influence network, as well as the factual reporting and political bias level of most major news channels. As outlined in the following chapter, I was able to use these three data points to make observations about the content that YouTube recommends on videos associated with Canadian politicians.

### *Depiction Analysis*

Once the source analysis was complete, a second, more focused depiction analysis was conducted in order to determine whether recommended videos directly related to each respective candidate were critical, supportive, or neutral toward them. This analysis was constructed to determine how each video represents the relevant candidates, as interpreted by my own semantic judgement of titles and actual video content where titles were unclear. Hypothetically, what I call depiction analysis should reveal bias in the recommendation system. To begin, I collected the videos that had titles relating to Andrew Scheer, Justin Trudeau, or Jagmeet Singh. For example, the video '5 Most Isolated Communities at The End of The Earth' posted by the channel *Mind Boggler* is not related to any candidate and was not considered. If a title appeared to potentially relate to the Canadian election without mentioning the name of a candidate, I pasted the video ID into YouTube's URL structure (https://www.youtube.com /watch?v=[ID]) and sampled the video to determine relevance.

In order to code the portrayal of politically relevant videos, I read titles and watched content where necessary. In most cases, the bias was clear and the determination was simple: A video that frames a candidate in a negative or positive light was considered to be biased and was coded as such. For instance, the video "Andrew Scheer leaves Justin Trudeau speechless, he didn't see this coming" by the channel *True Liberty* is very clearly biased in its depiction of both Justin Trudeau and Andrew Scheer. In the following results chapter, a list of the most recommended news channels and how they were coded is provided.

Some coding decisions were more complex. For example, a video of a candidate interview is considered neutral, for no additional commentary has been made (Figure 5.1). In a similar way, videos by news organizations are considered neutral in most cases, as they strive to only report objective facts about the candidates. Examples of this include interviews and reports on critical speech. An example of a notable exception to this rule is the video "Maxime Bernier: The Next Prime Minister of Canada? (Full Interview)" on *The Rubin Report* channel. Although he considers himself a 'classical liberal', Dave Rubin is identified as a major node in the Alternative Influence Network report and consistently platforms other members of the network. In the opinion of this author, Rubin is to be considered as a bad actor who does not adequately challenge the hateful rhetoric his guests often display, nor does his program offer sufficient viewpoints from left-leaning individuals. In other words, *The Rubin Report* is not considered a neutral source in this thesis and therefore will not be coded as one.


*Figure 5.1: Example of a Neutrally Coded Video*

<u>*Network Analysis*</u>

Once the coding process was complete for both the source and depiction analysis, I collated the data points together in order to undergo network analysis. This stage was done with the use of Google Sheets' query formula, which I was able to employ to bring together various coded videos and run calculations on them. The results of these data queries were then translated into bar charts and graphs for easier digestion and will be discussed in the following chapter. Finally, work was done to transform each of the datasets into a format that could be interpreted by Gephi, a piece of open-source network analysis software that displays complex relationships between nodes (here, videos), and directed edges (recommendations). Preparing the edge file for Gephi took additional work: A formula was used to find the number of recommendations (=COUNTA(E2:W2)) to iterate each seed for each recommendation it was associated with, creating the source half of the directional edge. This formula made use of the VLOOKUP function:

```
=ArrayFormula(vlookup(transpose(split(query(rept(row(Sheet1!A2:A)&" ",Sheet1!D2:D),,9^9),"
")),ArrayFormula({row(Sheet1!A2:A),Sheet1!A2:D}),{2,3,4},0))
```

The text-editing software Notepad++ was then used to transpose the ~19,000 recommendations using a simple REGEX replace of \s (space) with \n (newline). Any empty lines were removed with Notepad++'s line operations tool, 'remove empty lines'. With both the node and edge files prepared, they were then saved in the CSV format.

Once loaded into the software, it was possible to colour-code any trait or tag associated with each node. For the purposes of this research, nodes associated with either one of the three primary candidates were labeled, as well as MBFC bias, MBFC factual accuracy ratings, and individuals associated with the Alternative Influence Network. This information was used to colour-code various visualizations, while null or uncoded nodes were set to display in black and dark blue. Node size was determined by In-Degree, or the number of times a node received a recommendation from another node. In other words, the size of the nodes was used to show which videos had the most recommendations. Run through the software, the 1000 videos from each dataset and their 19 respective recommendations formed a network of almost 19,000 directional relationships. In the following analysis chapter, these visualizations will be used to highlight various clusterings in the data and how users can travel through the network over time.

*Identified Assumptions & Limiting Conditions*

As stated previously, this research does not take into account personalization. An expert on this subject is Eli Pariser, author of *The Filter Bubble* (2012) and coiner of the term 'filter bubble' itself. Pariser characterizes personalization as a mid-2000s move to a web experience that catered to the individual rather than the masses (MacManus, 2009). Using information about unique users like browsing history, age, time of day, geographic location, browser and operating system information and other identifying data, websites started retrieving different content for different people. In the 2011 TedTalk that preceded his book, Pariser began to raise the alarm about the invisible over-personalization of major services like Google and Facebook. The example given in his lecture was that of a Google search conducted by two different people using the exact same keyword that returned vastly different results, each tailored for them specifically. A potential side-effect of such a difference are filter bubbles or information cocoons where people are only ever exposed to ideas they agree with and content they are familiar with. "There is no standard Google anymore", Pariser says, and the same is true for YouTube. For every user the browsing experience is different, which makes research difficult. As reiterated here, this study does not take account of viewing history on YouTube. The implications of this are twofold: First, the veil of personal suggestions will be pulled back to help expose the algorithm's underlying tendencies and will place the focus squarely on the algorithm rather than users. Second, this approach will make the results of this study more generalizable. Avoiding personalized results is a simple task with Chaslot's Python software YouTube Explore as it does not access YouTube while signed into an account. Each search is not influenced by the others as viewing history is not retained. To be clear, there is no evidence that YouTube uses two separate versions of its recommendation algorithm for signed-in and signed-out users; instead it is likely that the system simply uses fewer data points to reach its conclusions. If YouTube's video algorithm has negative effects on users outside of personalized recommendations, a major assumption is that any potential negative effects are amplified when user information is used to personalize results. In their own research, Ribeiro et al. (2019) also propose that the problematic aspects of YouTube's RS are exaggerated by personalization.

One of the problems with collecting a set of data on three different individuals is the context that separates them from one another. While Andrew Scheer and Jagmeet Singh are both relatively new party leaders (May and October 2017, respectively), Liberal Party leader Justin Trudeau has been Prime Minister of Canada since 2015. Knowing this, the number of videos that are related to Trudeau (outside of his 2019 re-election campaign) is vastly higher than any other candidate. Given his position in society, more content revolves around him and would skew the results. To combat this issue, only videos published after May 2018 were considered in the close analysis. Another parameter of this study is that I am analyzing videos related to three candidates (party leaders) rather than all six major leaders who ran in the 2019 Federal election. However, the three other unincluded party leaders only made up roughly 8% of all seats countrywide pre-election combined (Parliament of Canada, n.d.). It is for this reason that they will not be considered in either analysis. It is also possible that there are titles in my data that do not refer to party leaders by name, despite containing content related to them. These videos will not be included in analysis. Additionally, spelling errors in video titles may result in videos being omitted from my analysis.

At one point, I was confronted with the same problem YouTube currently has: The sheer amount of content overwhelms any human fact-checker. YouTube loosely defines 'borderline' content as containing "unsubstantiated conspiracy theories, or verifiably inaccurate information" (Menegus, 2019), as determined by human evaluators. If I were to attempt to filter out videos into the incredibly vague 'borderline' category that YouTube has promised to crack down on, independently investigating each video for factual correctness would take an enormous amount of time and would ultimately be a subjective process. There is a great deal of content in the dataset that I would certainly personally identify as fake news, and it is not feasible to determine which videos would or would not belong to this category. An example of a video that I find factually dubious comes from *Rebel Media*'s Lauren Southern, a video in which she claims to have been 'thrown out' of a no-go zone in Sydney, Australia. Even from her own video, it is clear that her claim is false. This example would not be included in any determination of depiction bias as it does not relate to any candidate. However, while I do not know which videos YouTube would consider to be 'borderline' content, it is useful to explore how the spectrum of factual accuracy is represented in YouTube's recommendations. Thus, the source analysis is intended to supplement the lack of focus on borderline content. In order to make my research feasible, I have outsourced the determination of factual reporting and political bias levels to mediabiasfactcheck.com (MBFC), and similarly have used Rebecca Lewis' AIN report to discover how the alternative network is represented on YouTube.

Additionally, it is important to consider the possible technological aspects the data collection may be influenced by. For example, although the sample area is non-geographical by the very nature of research on the internet, it does not mean that my own location does not impact the search itself. Although Chaslot purposefully created a program that is not tied to an account with associated viewing history, searches can still be filtered by the IP address used to carry out the search in the first place. This is discussed in further detail in the following chapter. Knowing that Google uses over 200 different data signals to filter results on their search engine (Sullivan, 2010) and that a signed-in search on YouTube is influenced by users' viewing history, it is likely that the searches conducted with YouTube-explore are in some way filtered by factors including location and time of day. For reference, the software was run from an IP address located in Alberta, Canada. It can also be assumed that the YouTube algorithm in use at YouTube is to a varying degree somewhat different than it was in 2016, for as an ex-Google employee himself, Chaslot attests to the ever-shifting nature of the algorithm. The weighting of the signals used to rank videos is constantly being adjusted in an attempt to meet YouTube's internal goals. Many of those goals and signals are corporate secrets, but technical documentation published by Google engineers help reveal what is being prioritized and how. For example, Covington et al. divulge some of these signals in a 2016 paper: Click-through rate, search history, watch history, demographic information, and watch time are all used to help rank recommendations.

To their advantage, *The Guardian* worked within a two-party political system, giving them the ability to make general assumptions about how a video's attack of one politician almost certainly meant support for the opposition. There is no such case in Canada, where a YouTube video criticizing Andrew Scheer does not automatically signal support for either Trudeau or Singh. Another major advantage *The Guardian*'s study holds over mine is the period of data collection. For their study, Chaslot ran his software in three distinct time periods during 2016; once in August, several times in October, and for the first seven days of November. On the eighth day, Donald Trump was elected president of the United States of America. While *The Guardian* is quick to admit that their data was "probably influenced by the topics that happened to be trending", the timing of the data collection was impeccable. Assessing the state of YouTube's recommendations up to the very eve of the election certainly provided invaluable insight towards answering questions about YouTube's role in the election and the possibility of external influence. Given my own intended graduation date and the time it takes to collect and analyze the appropriate data as an individual researcher, I am not able to collect information in the months leading up to October 2019. Instead, I must embrace the positive aspects of this fact. For one, the data captured 9 months prior to the actual election comes at a point before the campaigns have gained traction, before the media has drummed up the voting public into a frenzy, and perhaps before any kind of external force (or indeed, country) have fully operationalized viral botnets and troll farms. Hopefully, this research will show a clearer picture of what YouTube's algorithm is promoting than data collected amid a battle over people's hearts and minds. Finally, the sample size used here is significantly smaller than the one used in *The Guardian*'s expose. While the dataset used in Lewis and McCormick's investigation covered 8052 videos, it was determined that an appropriate set for a lone researcher would amount to 3000 videos, given the time it would take to code and analyze each one.

**Conclusion**

This chapter has served to detail how my methodology was crafted, how data was collected, and how that information was analyzed to answer my research questions. After an overview of a 2018 study by *The Guardian* that serves as the basis for my own methodology, I discuss Guillame Chaslot's software YouTube-explore and what it is designed to do. Next, I explore the analysis that *The Guardian* uses in their research to draw claims from their data, identifying flaws that could bear improvement. Building upon and adapting *The Guardian's* methodology, I lay out the details of my own approach, beginning with my guiding objectives. I cover the specifics of how my data collection is conducted, accompanied by an explanation of the factors that influence which data is used in analysis. Analysis itself is the next topic discussed, split into three categories; Source analysis, depiction analysis, and network analysis are all described in detail. Finally, I also acknowledge some of the factors that limit my study, as well as some of the assumptions made by this research approach. Discussed in the following findings chapter are various patterns that emerged out of the content recommended from videos related to the three candidates, including prominent channels, topics, and videos. Also included is a look into the level of factual reporting and political bias of those sources.

# CHAPTER SIX: RESULTS & DISCUSSION

## Overview

In this chapter, I compile the results of the analysis described in the previous chapter and provide additional commentary on potential interpretations of those findings. Analysis of candidate representation, candidate depiction, Alternative Influence Network presence, factual accuracy, media bias, common sources, and network visualization are all described and discussed here. The first point of discussion will center around the number of videos related to the three candidates that were recommended in each dataset and the way they are presented in those videos. For clarification, the word 'dataset' refers to the data obtained from each keyword search. As such, there are three datasets, one pertaining to each candidate. Secondly, I will report the findings of my analysis of bias towards or against the three candidates, as well as discuss how these results may help answer my research questions. Thirdly, the extent to which members of Rebecca Lewis' Alternative Influence Network (AIN) appear in each dataset is discussed. Fourthly, my investigation into the factual accuracy and political bias of sources using mediabiasfactcheck.com (MBFC) is discussed. Fifthly, I present my analysis of the most common sources in the datasets. Finally, the chapter concludes with observations regarding the node clustering of recommended videos as drawn from the network analysis visualizations, more of which are found in Appendix A. For further context, a sample of a full dataset can be found in Appendix C.

## Candidate Representation

### *Results*

To begin, I counted the number of unique titles in each dataset (Scheer Singh, Trudeau) that related to **any** of the candidates (Table 6.1.1). The Trudeau dataset saw 21 relevant videos, while the Singh dataset similarly had 24. Scheer's dataset stood apart from the others, containing more than double the amount of titles relating to Justin Trudeau, Jagmeet Singh, or Andrew Scheer: 51 unique recommendations related to the candidates (for 1000 unique videos, this makes up only 5.1% of the entire dataset). It is noted here that this sample size is too limited for concrete conclusions to be made. I will suggest that future studies run multiple iterations of searches and aggregate the data to compensate.

From this point onwards, these unique recommendations will be referred to as 'relevant videos', as there is a meaningful distinction between a video title and the amount of times it is recommended within a single dataset. For example, the single video "SNC-Lavalin probe is fifth ethics investigation for Trudeau's cabinet" by *CBC News* is recommended 16 times in the Trudeau dataset. The terms 'video' and 'title' are used interchangeably.

| DATASET (1000 TITLES) | RELEVANT VIDEOS | % OF DATASETS |
|---|---:|---:|
| Scheer | 51 | *5.1%* |
| Singh | 24 | *2.4%* |
| Trudeau | 21 | *2.1%* |

*Table 6.1.1: Candidate Representation per Dataset*

In order to comply with the date restrictions outlined in the methodology, these numbers had to be adjusted to remove videos that were posted before May 2018 (Table 6.1.2). The updated results are again proportionally similar for Singh and Trudeau, while the Scheer dataset continued to contain the most videos relevant to the candidates. To reiterate, the number of relevant videos in each dataset refers to videos relevant to *any* candidate. The correct interpretation of Table 6.1.2, for example, is that the Scheer dataset contained 38 videos that were relevant to Scheer, Singh, or Trudeau.

| DATASET (1000 TITLES) | RELEVANT VIDEOS | % OF DATASETS |
|---|---:|---:|
| Scheer | 38 | *3.8%* |
| Singh | 18 | *1.8%* |
| Trudeau | 18 | *1.8%* |

*Table 6.1.2: Candidate Representation per Dataset (adjusted for post May 2018 date restriction)*

Next, I looked only at which candidates the relevant videos pertained to (Table 6.1.3). In other words, I counted how many of the relevant videos mentioned Trudeau in the title, and so on. In order to accurately represent the presence of these titles over all three datasets, I compiled a list of all pertinent videos per candidate and removed duplicates, as many of the titles appear in all three datasets. The total saw an overwhelming tilt towards Trudeau, with 69.1% of the pertinent titles across all datasets. Only 20.6% of the relevant videos relate to Scheer, and 10.3% to Singh. It is also worth noting here that some of these videos are pertinent to multiple candidates, resulting in overlap.

| CANDIDATE | VIDEOS PERTAINING TO CANDIDATE | % OF TOTAL |
|---|---|---|
| Scheer | 14 | *20.6%* |
| Singh | 7 | *10.3%* |
| Trudeau | 47 | *69.1%* |

*Table 6.1.3: Pertinence to Candidate over All Datasets*

Again adjusting the numbers to only account for date-eligible videos, Trudeau's prominence only grew, accounting for 72.5% of all the videos pertaining to the candidates (Table 6.1.4). In other words, once videos that were posted before May 2018 were removed, the overwhelming majority of videos that related to the candidates in the datasets still pertained to Trudeau. As noted in Table 6.1.2, the Scheer dataset contained the most relevant videos, highlighting the first result of interest: Videos pertaining to the three politicians were far more likely to be found in the Scheer dataset. The prevalence of Trudeau-pertinent videos over all datasets combined with how the Scheer dataset contained the most relevant videos to candidates indicates that recommendations favored Trudeau, even when searches occurred using the keywords 'Andrew Scheer'. This prevalence of Trudeau-related content was observed universally across all three datasets, as the presence of videos related to him vastly outweighed any other candidate.

| CANDIDATE | TITLES PERTAINING TO CANDIDATE | % OF TOTAL |
|---|---|---|
| Scheer | 9 | *17.6%* |
| Singh | 5 | *9.8%* |
| Trudeau | 37 | *72.5%* |

*Table 6.1.4: Pertinence to Candidate over All Datasets (adjusted for post May 2018 date restriction)*

Breaking down pertinence by dataset, a small pattern emerges. As expected, the number of recommended titles that pertained to a candidate was tied to the initial search performed (Table 6.1.5). For example, the most unique titles pertaining to Andrew Scheer were found in the Scheer dataset. Similarly, the only videos pertaining to Singh were found in the Singh dataset. In contrast, the Scheer dataset was an outlier as it contained the most videos pertaining to Trudeau. In fact, both the Scheer and Singh datasets had more titles pertaining to Trudeau than Scheer or Singh themselves, again highlighting Trudeau's universal prevalence.

| DATASET (1000 TITLES) | TITLES PERTAINING TO SCHEER | TITLES PERTAINING TO SINGH | TITLES PERTAINING TO TRUDEAU |
|---|---|---|---|
| Scheer | 9 | 0 | 31 |
| Singh | 1 | 5 | 14 |
| Trudeau | 1 | 0 | 18 |

*Table 6.1.5: Titles Pertaining to Candidates per Dataset (adjusted for post May 2018 date restriction)*

Counting how many times relevant titles were recommended serves to provide a better representation of what a user would actually experience, as any unique title could appear multiple times in the recommendations. The data in Table 6.1.6 shows how many recommendations over all the datasets were relevant to the three candidates. As shown, there is not a significant difference in percentage between the number of videos related to each candidate and the share of recommendations those videos received over all three datasets: Videos pertaining to Scheer make up 17.6% of all relevant titles (Table 6.1.4), and 14.8% of all relevant recommendations (Table 6.1.6). While videos related to Singh were fairly rare (9.8%) in all the datasets, those videos were recommended far fewer (1.7%) in comparison to the other two candidates, leading to a drop in visibility. The opposite occurred for videos related to Trudeau, which already made up 72.5% of all relevant titles. When these titles were recommended, they received a higher share of percentage (83.5%) due to their higher recommendation numbers.

| CANDIDATE | RELEVANT RECOMMENDATIONS | % OF TOTAL |
|---|---|---|
| Scheer | 86 | *14.8%* |
| Singh | 10 | *1.7%* |
| Trudeau | 485 | *83.5%* |

*Table 6.1.6: Relevant Recommendations per Candidate over all Datasets (adjusted for post May 2018 date restriction)*

It is also useful to consider the data that was not analyzed (Table 6.1.7). Ineligible videos are videos that relate to the candidates but cannot be considered towards totals as they were posted before May 2018. To reiterate the rationale behind the methodology, this was done in order to help compensate for the advantage Justin Trudeau had with existing content due to his prior role as Prime Minister. In addition, videos that were no longer on YouTube at the time of analysis were also removed as their date posted could not be verified. Looking purely at unique titles, more videos related to Trudeau were removed than the other two candidates, a finding that supports my decision to remove them from consideration in analysis. However, a higher proportion (35.7%) of all videos relevant to Andrew Scheer also had to be removed from analysis.

| CANDIDATE | INELIGIBLE VIDEOS | % OF RELEVANT VIDEOS | ELIGIBLE VIDEOS | % OF RELEVANT VIDEOS |
|---|---|---|---|---|
| Scheer | 5 | *35.7%* | 9 | *64.3%* |
| Singh | 2 | *28.6%* | 5 | *71.4%* |
| Trudeau | 10 | *21.3%* | 37 | *78.7%* |

*Table 6.1.7: Eligibility of Videos as a Percentage of Total Relevant Videos per Candidate*

For a visual representation of the total amount of videos removed, refer to figure 6.1 below. As shown, the percentage of videos removed from the Singh and Trudeau datasets is roughly proportional, while a larger portion of the videos pertaining to Scheer had to be removed.

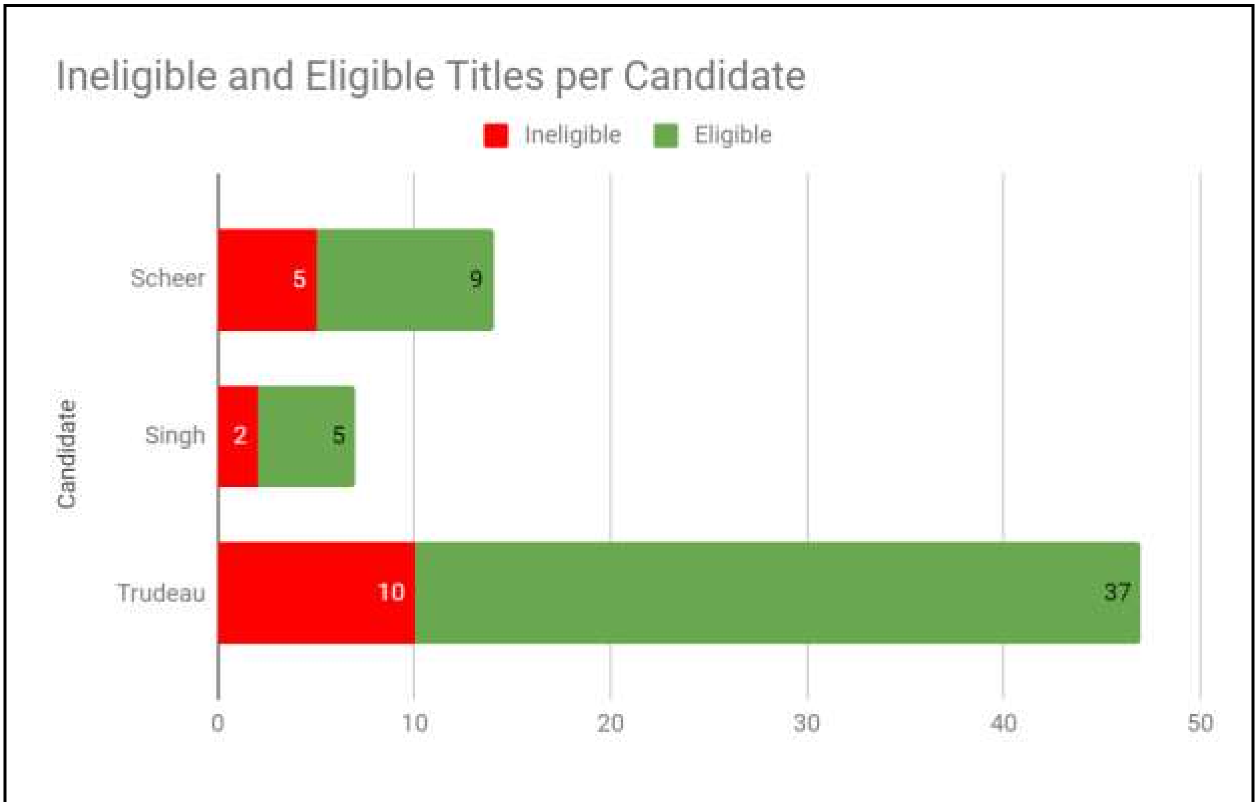



*Figure 6.1: Ineligible Titles out of Total Titles per Candidate*

When looking at the number of recommendations each ineligible video receives, the ratios become more moderate (Table 6.1.8). As expected, removing the most videos pertaining to Trudeau resulted in the most recommendations removed (90). 21.3% of videos pertaining to Trudeau were removed, but as videos are often recommended multiple times, only 15.7% of recommendations pertaining to Trudeau were removed. The ineligible videos pertinent to Scheer have low recommendation counts, as evidenced by the fact that 35.7% of titles pertinent to him were removed, but only 19.6% of recommendations.

| CANDIDATE | INELIGIBLE RECOMMENDATIONS | % OF RELEVANT RECOMMENDATIONS TO CANDIDATE |
|---|---|---|
| Scheer | 21 | *19.6%* |
| Singh | 6 | *37.5%* |
| Trudeau | 90 | *15.7%* |

*Table 6.1.8: Ineligible Recommendations out of Total Recommendations per Candidate*

*Analysis*

As demonstrated in the results section above, Scheer's dataset had significantly more relevant videos than the other two datasets. The explanation for this anomaly is likely complex: It is possible that even a single seed video could yield a particularly relevant path of recommendations and skew the relevant videos in one dataset, but these aggregate numbers suggest something else is going on. Combined with the sheer number of Trudeau-related videos in this dataset, it can be suggested that this anomaly in the recommendations is reflective of user viewing habits. That is, the collaborative filtering aspect of the recommendation engine reflects which videos other users are most likely to click on. As a refresher, collaborative filtering occurs when a recommender system makes predictions about a user's interests by aggregating preference behaviour from other users, in a very basic sense assuming that users with similar taste are good predictors for what each other will enjoy. In the context of the YouTube recommendation algorithm, this collaborative process is called co-visitation.

Here, it is possible to suggest that any difference between datasets is a reflection of viewing patterns, for recommendations based on 'Andrew Scheer' seed videos are influenced by what other users watching those same videos watched next. Although Andrew Scheer is certainly a lesser-known candidate compared to Trudeau, it would seem that users searching for videos related to him are simply more likely to watch nationally relevant videos, a trend recognized and exaggerated by the algorithm. I will then hypothesize that the number of videos related to each candidate and who is being recommended the most within each dataset is directly correlated to audience. The term audience is used here to simply represent the group of users searching for specific keywords; in this instance, the keywords are the names of the major political party leaders in Canada. As Justin Trudeau is an internationally recognized figure, I propose that the audience viewing videos related to him is broader and has less interest in or knowledge of the other candidates in the election. This is reflected in the Trudeau dataset as only videos related to Trudeau himself were recommended, other than a single video pertaining to Andrew Scheer. Conversely, users searching for videos on Andrew Scheer are more likely to have prior knowledge of Canadian politics, and therefore are recommended more videos related to both Scheer himself as well as Trudeau. The absence of Singh in both datasets is possibly a reflection of his relatively low stature in Canadian politics at the time of data collection. I will also suggest that Singh is a less controversial figure than either Trudeau or Scheer, resulting in fewer searches and recommendation clicks.

## Candidate Depiction

*Results*

As outlined in the methodology chapter, each related video was coded in neutral, critical, and supportive categories (Table 6.2.1) according to how the relevant candidate was depicted. The determination of each depiction was made on the basis of sentiment in titles, and where necessary, video content. If a title portrayed the candidate in a negative light, it was coded as being critical, and vice-versa. If the title did not indicate any bias towards or against a candidate, it was watched to confirm this categorization, and coded as neutral. When counting the recommendations in all three datasets combined, the majority (44.4%) of recommendations were critical of Trudeau, who was the only candidate to receive critical recommendations at all. The only candidate to receive supportive recommendations was Andrew Scheer, while all three Candidate were depicted in a neutral light to varying degrees. The recommendations depicting Andrew Scheer positively can all be traced back to a singular video that was found in all three datasets: “Andrew Scheer leaves Justin Trudeau speechless, he didn't see this coming” hosted by the channel *True Liberty* (now deleted). Similarly, all 10 of the recommendations for Singh himself were in a neutral light and came from a variety of major Canadian news sources, including the *CBC News, The Fifth Estate, The Vancouver Sun,* and *Global News*.

| CANDIDATE | NEUTRAL DEPICTION | % | CRITICAL DEPICTION | % | SUPPORTIVE DEPICTION | % |
|---|---|---|---|---|---|---|
| Scheer | 27 | *4.6%* | 0 | *0%* | 59 | *10.2%* |
| Singh | 10 | *1.7%* | 0 | *0%* | 0 | *0%* |
| Trudeau | 227 | *39.1%* | 258 | *44.4%* | 0 | *0%* |

*Table 6.2.1: Depiction of Recommendations per Candidate over all Datasets*

Breaking down these recommendations, the data can be used to closely study how each candidate was depicted in each dataset (Table 6.2.2). With exceptions, results indicate a somewhat uniform depiction of candidates regardless of initial search. All three datasets contained recommendations supportive of Scheer, while none contained critical recommendations. Similarly, all three datasets contained recommendations that were either neutral or critical of Trudeau, while none were supportive of Trudeau. Despite these similarities, the Scheer dataset contained significantly more neutral and critical recommendations related to Trudeau.

| RECOMMENDATION DEPICTION | SCHEER DATASET | SINGH DATASET | TRUDEAU DATASET |
|---|---|---|---|
| Neutral Towards Scheer | 27 | 0 | 0 |
| Critical of Scheer | 0 | 0 | 0 |
| Supportive of Scheer | 34 | 13 | 12 |
| Neutral Towards Singh | 0 | 10 | 0 |
| Critical of Singh | 0 | 0 | 0 |
| Supportive of Singh | 0 | 0 | 0 |
| Neutral Towards Trudeau | 130 | 39 | 58 |
| Critical of Trudeau | 161 | 59 | 38 |
| Supportive of Trudeau | 0 | 0 | 0 |

*Table 6.2.2: Candidate Depiction per Dataset*

*Analysis*

Overall, the evidence suggests a weak correlation between the number of unique videos returned from a search for a political candidate and the variance in that candidate's depiction. In other words, the greater number of videos related to or about a candidate, the greater likelihood that those videos will express a wide range of opinion about the candidate. The sample size for Singh-pertinent titles is small, and consequently there are only neutral depictions of that candidate. These findings simply indicate that a larger sample size creates a more representative pool of videos from which to make observations on.

The larger goal here is to determine if these findings are in any way indicative of some larger trend or bias towards or against any candidate. Using the recommendation data from all three datasets to capture a more comprehensive view of recommendations at the time of collection, it is possible to observe trends that may help answer this question.

1. Every search resulted in recommendations supportive of Scheer. None were critical.
2. Recommendations pertaining to Singh were almost non-existent.

3. Every search resulted in recommendations both supportive and neutral towards Trudeau. None were supportive.

These three observations allow me to state that each candidate was **not** represented equally in YouTube's recommendations. While the search being performed does appear to influence how each candidate is depicted, the overall depiction of these politicians is biased towards Andrew Scheer and against Justin Trudeau. To explain these findings in more detail: I am suggesting that a YouTube-wide bias towards the promotion of pro-Scheer videos is likely as supportive recommendations are also found when searching for either Singh or Trudeau. Similarly, all three searches lead to recommendations critical of Justin Trudeau, while no other candidate received a single critical recommendation. Using this data, the primary research question posed by this thesis can be answered: There is a degree of political slant present in the videos recommended by YouTube on searches using the names of individual party leaders Jagmeet Singh, Andrew Scheer, or Justin Trudeau. As mentioned previously, direct comparison to *The Guardian*'s findings is difficult given the vastly different contexts. However, I am also able to state that the slant present in my data is nowhere near as pronounced as it is in Lewis and McCormick's data. The recommendations that were critical of Trudeau made up 52.8% of all relevant recommendations, while the recommendations that were supportive of Scheer made up 12.1%. Comparing only partisan videos as *The Guardian* did, 100% of videos were supportive of Scheer, while 0% were supportive of Justin Trudeau. For contrast, 88% of the recommended videos in *The Guardian*'s study were supportive of Trump, and only 16% supportive of Hillary Clinton. I will suggest that *The Guardian*'s choice to exclude non-partisan videos may lend to better headlines but does not tell the whole story. Also contributing towards my hesitation in making conclusive statements is the small sample size afforded by the data. After all, titles I deemed eligible to use in candidate depiction only made up 7.2% of all titles overall. For example, *The Guardian* found 643 partisan videos in their data, while I found only 317.

When a recommendation that is critical of Trudeau or supportive of Scheer is present in the recommendation sidebar, it is put there because the algorithm has noticed a correlation between the current video being watched and what other users tend to watch next. The choice of what to watch next reinforces and informs those relationships. As a result, the videos recommended by YouTube reflect the viewing habits of users, thanks to the recursive nature of the recommender system. As the three datasets contain fairly consistent depictions of the candidates, I will suggest that viewers are simply more likely to click on recommended titles that are critical of Trudeau than those that are supportive of him.

Otherwise, this pro-Scheer and anti-Trudeau bias is up for debate; Foreign electronic interference in Canadian affairs is possible, but not probable as there is no evidence to suggest so. Again, I offer this explanation: These recommendations are reflective of the viewing habits of users. If the journey from one YouTube video to another is a path, then a recommendation represents evidence of busy traffic on that path, in the same way that a rut or groove in a road would. It will be argued here that YouTube users interested in Canadian political leaders are increasingly politically divided. It is possible that this divisiveness has been exacerbated by ongoing attempts at political intervention by the Russian government, as warned by several experts, including the Canadian Centre for Cyber Security and Canadian Public Safety Minister Ralph Goodale (Canadian Centre for Cyber Security, 2018; Guignard, 2019). Alternatively, it is possible that content creators are less likely to create content critical of Scheer or Supportive of Trudeau, either due to perceived poor reception and engagement or a bias towards or against objective reporting.

**Alternative Influence Network Presence**

*Results*

Next, I analyzed the titles of videos in the datasets associated with Rebecca Lewis' Alternative Influence Network (AIN), a report which outlined a web of interconnected individuals and channels that led towards alt-right content on YouTube. Lewis calls these content creators 'political influencers'. In short, I measured the presence of AIN members in each dataset by counting titles associated with or from AIN members, and counting the total amount of recommendations those titles received in each dataset (Table 6.3.1). In addition, I also compiled a list of the most recommended AIN figures across all three datasets to get a better picture of who YouTube was recommending alongside my searches for political figures (Table 6.3.2).

The Scheer dataset led the pack with 66 titles from 13 unique members of the network, resulting in 351 recommendations. Jordan Peterson was recommended the most in this dataset, at a total of 138 times. In the Singh dataset, Peterson was also recommended the most of the AIN figures, although this dataset saw a decline in both unique titles and members represented. Finally, the Trudeau search only retrieved videos from 7 unique members which were associated with 36 unique titles, just over half the amount found in the Scheer dataset. However, comparing the raw number of unique titles from each dataset can be misleading, as the representation experienced by the end user is actually dependant on recommendations. For example, if a unique video (title) appears in the dataset, it is not representative of how many times that title actually appears in aggregate. In this case, it is true that Trudeau's dataset saw the fewest number of unique AIN-affiliated titles, but this is not representative of how many times a video from a member of the network was promoted. Adjusting the numbers to account for recommendations, the Trudeau dataset recommended AIN videos almost as much as the Scheer dataset.

| DATASET | UNIQUE AIN MEMBERS | AIN ASSOCIATED TITLES | AIN ASSOCIATED RECOMMENDATIONS |
|---|---|---|---|
| Scheer | 13 | 66 | 351 |
| Singh | 9 | 33 | 196 |
| Trudeau | 7 | 36 | 317 |

*Table 6.3.1: AIN Presence per Dataset*

Overall, 15 of the 66 identified political influencers in Lewis' Alternative Influence Report could be found in the three datasets, accounting for 901 recommendations (Table 6.3.2). Of those recommendations, 43.6% directed to videos from or about Jordan Peterson, 27.4% to Ben Shapiro, and 8.4% were associated with the *Rebel Media* channel. The rest of these recommendations were associated with the remaining 12 individuals. Across each individual dataset, appearances by these political influencers are low: Measured against all videos, unique titles from members of the AIN only made up 0.03% of the Singh dataset, 0.04% of the Trudeau dataset, and 0.07% of the Scheer dataset, meaning that relatively speaking, the AIN was not well represented in the data collected. Adjusted for recommendations, the percentage remains similar: 0.03% of the recommendations in the Singh dataset related to the AIN, while making up 0.04% of the Trudeau dataset. The Scheer dataset contained the most AIN recommendations, but they still only made up 0.06% of all the recommendations.

| ALTERNATIVE INFLUENCE NETWORK MEMBER | RECOMMENDATIONS (ALL DATASETS) | PERCENTAGE OF AIN RECOMMENDATIONS |
|---|---|---|
| Jordan Peterson | 393 | *43.6%* |
| Ben Shapiro (The Daily Wire) | 247 | *27.4%* |
| Rebel Media | 76 | *8.4%* |
| Steven Crowder | 67 | *7.4%* |
| Dave Rubin | 44 | *4.9%* |
| Paul Joseph Watson | 21 | *2.3%* |
| Stefan Molyneux | 15 | *1.7%* |
| Lauren Southern | 13 | *1.4%* |
| Milo Yiannopolous | 8 | *0.9%* |
| Dennis Prager (PragerU) | 4 | *0.4%* |
| Joe Rogan (PowerfulJRE) | 4 | *0.4%* |
| Candace Owens (Red Pill Black) | 3 | *0.3%* |
| Gavin McInnes | 3 | *0.3%* |
| Dave Cullen (Computing Forever) | 2 | *0.2%* |
| Tommy Robinson | 1 | *0.1%* |

*Table 6.3.2: Most Recommended AIN Members*

*Analysis*

While the presence of members of the Alternative Influence Network is relatively low in my data, their appearance at all is noteworthy. There is no doubt that if a YouTube employee was pressed, they could not reconcile some of the titles recommended with their terms of service. For example, Dave Rubin's conversation with Stefan Molyneux entitled "Stefan Molyneux on Race and IQ (Pt. 2)" is an open exploration of scientific racism, packaged neatly by high production values as an honest conversation about science. In June 2019, the video's title was altered to "Stefan Molyneux on Controversies (Pt. 2)", likely in an attempt to disguise its contents. The video has over 1.2 million views, much in thanks to the algorithm's preference for videos with already high viewcounts. In contrast, the first portion of the interview was not found recommended in my research, and presumably as a consequence, only has 219,000 views. While the video from the *Rubin Report* was only recommended a total of 15 times across all three datasets, it was indeed found in all three, suggesting some sort of universal prevalence. Further work must be done with less politically relevant keywords to determine just how widespread videos like this are in YouTube's recommendation cycle.

At 95 recommendations, the title "Jordan Peterson Calmly EDUCATES Baiting Host on Her Own Show", was by far the most recommended video in any dataset and is hosted on the channel *Crysta*. This video is a prime example of two often overlapping trends in AIN videos identified in the datasets: single-word capitalization, and 3rd person singular present simple forms of verbs. Of 92 unique AIN titles, 32 contain a capitalized word, 32 and use the 3rd person singular present simple forms of a verb. Some examples include words like 'EDUCATES', 'DESTROYS', 'dismantles' and 'CONFRONTS'. This style is used to grab the attention of viewers: The capitalization of these verbs serves to emphasize the extent to which the individual in question has achieved a perceived victory over their opponents. In an interview with the Dutch publisher *De Correspondent,* alt-right YouTube user Marcus Wouter explains that these verbs are used to hold the attention of viewers (Bahara et al., 2019). To underline the prevalence of this title style on YouTube, I turn to AlgoTransparency.org's list of the most recommended verbs on the platform. Drawing from 8.9 million recommendations from 1000 U.S. based channels over a period of 6 months, Guillaume Chaslot collected the 25 most-recommended 3rd person singular present verbs on YouTube. As a sample, I have included the top 10 in Table 6.3.3.

| VERB | RECOMMENDATIONS PER VIDEO |
|---|---:|
| Dismantles | 70.3 |
| Educates | 68.6 |
| Debunks | 54 |
| Snaps | 37.2 |
| Realizes | 36.2 |

| | |
|---|---|
| Screams | 35.8 |
| Obliterates | 34.4 |
| Shreds | 34.1 |
| Defies | 33.1 |
| Owns | 31.9 |

*Table 6.3.3: 10 Most Recommended 3rd-person Singular Present Verbs on YouTube (Most recommended verbs on YouTube, n.d.)*

I propose that this dramatic language is obviously indicative of bias and furthermore has been employed successfully as evidenced by the high recommendations per video.

Conversations surrounding the popularity of the three most recommended AIN members (Jordan Peterson, Ben Shapiro, and *Rebel Media*) perhaps belong elsewhere, but there is certainly something to be said regarding their somewhat uniform prevalence in all three of these datasets: The occurrence of these political influencers across the Scheer, Singh, and Trudeau datasets indicate site-wide recommendation, and that YouTube's algorithm was and likely still is promoting these channels regardless of the search performed, or at the very least on videos of a political nature.

There are other commonalities, too. While both Peterson and Shapiro are typically praised in various titles for their quick-wittedness ("Student Tries to FRAME Jordan Peterson! INSTANTLY DISPROVEN (Lafayette University)", or "Ben Shapiro FORCES Leftists and SJW to SIT DOWN! Q&A at YALE - MGTOW"), the titles from *Rebel Media* hold their own unique trait: they are Canadian. This trend towards nationally related content is not accidental, and likely tied to the viewers who are typically searching for Canadian political candidates. As for the higher number of AIN titles in the Scheer dataset, I will suggest that this reflects a greater number of viewers with conservative ideologies. As before, I believe this trend is indicative of user viewing behaviour. There is an undisputable overlap between conservative ideology and the views often expressed by the AIN, so it comes as no surprise that viewers seeking out videos related to Scheer are more likely to click on recommendations from the AIN.

Regardless of volume, the presence of these videos in the data often contradicts YouTube's stances on borderline content. The majority of the AIN titles have millions of views, and do not appear to face any of the restrictions placed on borderline content (reduced recommendation, removal of likes, comment sections, suggested videos, and warning messages). Given their high viewcounts, I consider it unlikely that the videos in question have never received any flags in the course of their existence on the platform. I will therefore conclude that these videos are not considered to contain borderline content in the eyes of YouTube staff. But do the AIN titles in my data contain hate speech, or even borderline content? Although subjective, I consider the answer to be more complicated than a binary 'yes' or 'no'. Some videos clearly violate community guidelines: Dave Rubin's interview with Stefan Molyneux includes suggestions that African Americans are genetically predisposed to have a lower IQ than other races. This video is available to watch on YouTube, although Rubin's channel has re-labeled it as "Stefan Molyneux on Controversies (Pt. 2)". Steven Crowder's video "52-Year-Old Transgender Man Becomes 6-Year-Old Girl?" on the Louder with Crowder channel clearly violates YouTube's policies by promoting hatred against a protected group. However, there is good news: Upon revisiting the video nearly 9 months after collection, the video has been replaced with the following message: "This video has been removed for violating YouTube's policy on hate speech. Learn more about combating hate speech in your country." Other videos are more difficult to categorize: Lauren Southern's video "Thrown Out of Sydney No Go Zone" spreads the false narrative that she is not permitted to access an area near a mosque in Australia due to it being "conquered land", when in reality she is asked to leave the area by a police officer concerned about her open Islamophobia contributing towards a breach of the peace.

Despite the examples given of overt hate speech, I believe that most of the AIN titles seeing success in recommendation do so much in part due to their careful positioning at the border of permitted content. This idea is supported by Yonatan Zunger, a privacy engineer at Google: "Bad actors quickly get very good at understanding where the bright lines are and skating as close to those lines as possible" (Bergen, 2019). As obvious hate speech is banned from the platform, these individuals skirt the rules by implying their opinions without directly stating them, a tactic made all the more effective as YouTube actively avoids enforcing their already vague policy positions in order to distance themselves from the moderation process, and therefore, responsibility. In addition, Rebecca Lewis has also identified the primary tactic of this group: Using collaboration to create a self-supporting network to increase views and piggy-back the co-visitation metric. By consistently featuring other members of the network in their videos, these channels form a chain of co-visitation markers that manipulates the recommendation algorithm into leading viewers down an increasingly dark path. Rebecca Lewis is clear: "The Alternative Influence Network provides a pathway for the radicalization of audience members and content creators alike" (Lewis, 2018, p. 35). Dubbed by some as the 'reactionary right', members of the AIN produce targeted content by bracing against the wall of rules imposed by YouTube and interweaving amongst each other to climb to the top of recommendations like poison ivy. This strategic amplification intentionally rides the line between permitted and disallowed content to manipulate audiences and by extension the system into perpetuating an increasingly larger audience.

It is possible to see weak evidence of a radicalization pathway within my own data, as the number of recommendations of AIN political influencers appear to inversely correlate with how radical their ideology is. Due to the interconnected nature of the network, a user who watches a recommendation by or related to more 'mainstream' influencers like Ben Shapiro or Jordan Peterson would then be more likely to receive recommendations for slightly more alternative voices like Steven Crowder or Paul Joseph Watson. As openly white-nationalist figures like Stefan Molyneux and Tommy Robinson are not as well represented in the recommendations, this might indicate a general audience's disinterest in overtly racist content--it could be tentatively suggested that the only reason such content appears at all in the recommendations is its relationship to the seed video, as the prevalence of a recommendation is essentially a measure of what similar users watched after the seed video. In future work, I will investigate radicalization further by making use of the depth data collected, anticipating that recommendations at deeper levels (recommendations found on recommendations) are more likely to contain AIN content.

## Factual Reporting & Media Bias

### *Results*

Finally, the ratings from the website mediabiasfactcheck.com (MBFC) were ran against the datasets to determine the overall level of factual reporting and bias of the sources present in the recommendations. Over all three datasets, 7330 recommendations were coded for media bias, 2651 coming from the Scheer dataset, 2638 from the Singh dataset, and 2041 from Trudeau's (Table 6.4.1). All three sets were dominated by left-center sources, combining to a total of 64% of coded sources overall. In both the Singh and Trudeau datasets, sources designated as Left came in second, 12.6% and 19.7% respectively. In the Scheer dataset, the second-highest occurrence of media sources were coded as Right. An example of a source considered as Right by

MBFC is *Fox News*. The following table shows the 10 most prevalent channels recommended in all three datasets and the number of recommendations they received.

| MEDIA BIAS | SCHEER DATASET | SINGH DATASET | TRUDEAU DATASET | TOTAL |
|---|---|---|---|---|
| Left-Center | 1798 | 1723 | 1169 | 64% |
| Left | 225 | 332 | 402 | 13.1% |
| Right | 292 | 235 | 251 | 10.6% |
| Least Biased | 138 | 141 | 71 | 4.8% |
| Questionable Source | 66 | 74 | 9 | 2% |
| Right-Center Bias | 83 | 68 | 35 | 2.5% |
| Pro-Science | 32 | 65 | 86 | 2.5% |
| Conspiracy / Pseudoscience | 17 | 0 | 18 | 0.5% |
| TOTAL | 2651 | 2638 | 2041 | 100% |

*Table 6.4.1: Media Bias in Recommendations per Dataset*

As shown in Table 6.4.2, Left-Center sources dominate the recommendations, with only *Fox News* representing Right-coded sources in the top 10. Here, there is a prevalence of Canadian sources, as both *CBC News* and *CTV News* are represented well.

| CHANNEL | MEDIA BIAS | RECOMMENDATIONS |
|---|---|---|
| *CBC News* | Left-Center | 930 |
| *Fox News* | Right | 535 |
| *VICE News* | Left-Center | 500 |
| *ABC News* | Left-Center | 487 |

| | | |
|---|---|---|
| *PBS NewsHour* | Left-Center | 455 |
| *MSNBC* | Left | 380 |
| *Vox* | Left | 248 |
| *CNN* | Left | 213 |
| *CTV News* | Left-Center | 166 |
| *The Guardian* | Left-Center | 143 |

*Table 6.4.2: Top 10 Channels Recommended over all Datasets*

Factual Reporting as graded by MBFC also showed strong consistencies across the datasets (Table 6.4.3). To reiterate the methodology chapter, sources were assigned levels of factual accuracy by mediabiasfactcheck.com using a scale from 0-10, resulting in a factual accuracy rating between Very High and Very Low. MBFC also designates some sources as containing fake news, conspiracy, and propaganda. Per dataset, both the Singh and Scheer datasets mostly contained recommendations with a High level of factual accuracy, followed by sources with a Mixed rating. The remaining recommendations were coded as having Very High factual accuracy, with a scattering of sources MBFC has designated as Fake News, Conspiracy, and Propaganda. This time, the outlier is the Trudeau dataset, leading with a lower number of sources classified as having a High level of factual accuracy, and a higher percentage of sources having Mixed factual accuracy. Over all three datasets, these numbers show that the majority of sources recommended have a High level of factual accuracy, and yet 23% of the sources are coded as having a Mixed factual accuracy. The very small portion of channels that were coded as containing fake news or propaganda include *Occupy Democrats*, *RT* (formerly *Russia Today*), *PragerU*, and a handful of others. For the curious, the sources coded as Right had a lower average factual accuracy rating than the overall dataset. Of the 83 unique titles that were coded as Right across all three datasets, 100% of the 11 sources had a factual accuracy rating of Mixed. In contrast, the 80 unique titles across all three datasets coming from the 10 Left-coded sources are a mixed bag: half were considered to be sources with a Mixed factual accuracy rating, while the other 50% were coded as being sources with a High level of factual accuracy.

| FACTUAL ACCURACY | SCHEER DATASET RECOMMENDATIONS | SINGH DATASET RECOMMENDATIONS | TRUDEAU DATASET RECOMMENDATIONS | TOTAL |
|---|---|---|---|---|
| High | 1655 | 1772 | 1210 | 65% |
| Mixed | 476 | 544 | 624 | 23% |
| Very High | 245 | 248 | 180 | 9.4% |
| Propaganda | 31 | 43 | 4 | 1.1% |
| Fake News | 32 | 27 | 5 | 0.9% |
| Low | 15 | 0 | 18 | 0.5% |
| Conspiracy | 0 | 4 | 0 | 0.1% |

*Table 6.4.3: Level of Factual Accuracy per Dataset Recommendations*

*Analysis*

YouTube's bias towards the center-left is not an expected result, given the seemingly overwhelming warnings from both academic and conventional media sources against the far-right's growth and prevalence on YouTube. This is a misconception. It is not contradictory to state that YouTube has a highly interconnected rabbit-hole of far-right sources (as evidenced by the Alternative Influence Report and others) while at the same time primarily recommending Center-Left media. The revelation does however pose a major question: How is it possible to point the finger at systemic, structural biases when the recommender system itself favours moderate sources? Would it not instead indicate that there are small bubbles of content that persist on the platform, propped up by a small but fervent viewer base? The answer is of course no. As demonstrated in both the context chapter and the literature review, there exists a highly interconnected 'dark side' of YouTube that takes advantage of an ethically blind recommender system. While the same system does typically promote content that is not damaging to society, it has been commandeered by fearmongers to push hateful rhetoric to a curious audience that could interpret recommendation as a tacit approval. It is also worth wondering if YouTube is simply more likely to recommend titles from left-leaning channels due to their being more of that kind of content on the platform. Alternatively, YouTube's crackdown on fake news and corresponding promotion of authoritative sources may also favour the left, as my own data showed that Right-leaning sources were more likely to be factually inaccurate. Perhaps, as McLuhan might suggest, the sea of videos on YouTube is highly influenced by the audience watching them. Audience shapes media, and vice-versa. The difference is that this shift in content not only happens passively and reactively as before but is now built directly into the system.

To return to the data, the departure of Scheer's second-most promoted media bias from the two other datasets combined with the fact that Andrew Scheer is a right-leaning candidate, it can be cautiously suggested that there is some correlation between the political affiliation of the search being performed and the videos that are being recommended as a result. This however is not enough data to prove such an assertion and needs further research. The relatively lower number of videos coded with a High level of factual accuracy in the Trudeau dataset is another point of contention. As before, I will offer the argument that searches for videos related to Trudeau typically garner a more diverse viewership, and therefore more diverse sources. Overall, I think that it is fair to say that the 2.6% most dangerous and misleading sources making up the Propaganda, Fake News, Conspiracy, and Low factual accuracy MBFC designations in my data should not be recommended at all, particularly since the algorithm is unlikely to recommend a video without a few million views in the first place. At the point of recommendation, both the system and the human administrators have already failed the public and are pushing dangerous content to a wide audience.

## Common Sources

### *Results*

Coding the sources found in each dataset allowed the opportunity to observe some of the most recommended videos and channels resulting from the seed videos (Table 6.5.1). Firstly, the Singh dataset was comprised of 428 unique channels, 84 of which (20%) were news sources as recognized by MBFC. In this dataset, the channel *TEDx Talks* presented the most unique titles. In terms of channels that MBFC recognized as journalism, *CBC News* had the most unique titles. The news channel with the second-most unique titles was *Fox News*. Despite having fewer unique titles, both the channels *Vice News* and *PBS Newshour* were recommended more times than *Fox News*. The Trudeau dataset contained 526 unique channels and was dominated by the channel *LastWeekTonight* with the most unique titles. As for news sources, the Trudeau dataset contained 72 journalistic sources, and *CBC News* once again had the most unique titles. Here, *MSNBC* was the second-most prevalent news channel. Finally, the *CBC News* channel held the title of having both the most unique titles and news media titles overall in the Scheer dataset. Like the Singh dataset, *Fox News* had the second-most unique titles.

| | SCHEER DATASET | SINGH DATASET | TRUDEAU DATASET |
|---|---|---|---|
| # OF UNIQUE CHANNELS | 463 | 428 | 526 |
| MOST PREVALENT CHANNEL | CBC News | TEDx Talks | LastWeekTonight |
| # OF NEWS CHANNELS | 90 | 84 | 72 |
| MOST PREVALENT NEWS CHANNEL | CBC News | CBC News | CBC News |
| SECOND MOST PREVALENT NEWS CHANNEL | Fox News | Fox News | MSNBC |

*Table 6.5.1: Most Common Sources per Dataset*

*Analysis*

Some of the videos and channels in the datasets do not relate to the search terms in any way, and I consider these to be some of the most recommended sources on YouTube at the time of data collection. Although a control group would have been helpful to more concretely establish this supposition as fact, I will suggest that examples of universally recommended channels would include the *LastWeekTonight* and *TEDxTalks* channels due to their uniform prevalence on all datasets. The abundance of titles from the *CBC* channel echoes earlier sentiments of bias towards Canadian content as a reflection of the viewing habits of users conducting similar searches. It is also possible that the amount of Canadian political content on YouTube is higher due to the pre-election season and accompanying rise in public interest.

Other observations relate to different aspects of the videos recommended. Corroborating the research conducted by PEW (Smith et al., 2018), the average view count of a seed video was almost always in the millions. For example, the mean view count for the videos in the Trudeau dataset was 3,840,216. The Singh dataset was slightly smaller at 2,316,656, and Scheer smaller still at 2,050,563. The 2010 report published by some of the engineers who designed the recommendation algorithm provides an easy explanation for this preference towards videos with high view counts: The co-visitation metric accounts for what other users who watched the same video went on to watch next. Therefore, it logically concludes that a video with a low view count is less likely to have been watched by a large number of other users.

The heavy presence of the infamous *Fox News* channel in YouTube's recommendations also deserves further conversation; Discussions regarding the quality of the news network and the level to which the channel is complicit in spreading disinformation are critical to exploring the extent to which YouTube itself is responsible for not only hosting, but promoting controversial content. For reference, MBFC rates *Fox News* as having mixed factual accuracy, citing consistent failures to fact-check, spreading conspiracies with poor correction response, and inadequate information sourcing. As discussed in the final chapter, the solution to the information overload problem presents a new set of challenges as automation is automatic and undiscerning.

## Network Visualization

### *Results*

In these network visualizations, I have again deliberately separated the terms 'videos' and 'recommendations', for a single unique video can be exaggerated over a dataset with multiple recommendations. For example, the fact that the video "Whitaker clashes with lawmaker over donations" posted by the *CNN* channel only appears once in the Trudeau dataset is misleading: In reality, the title has actually been recommended 29 times. For this reason, recommendations will be given greater consideration in these results as they are more representative of overall bias.

#### *Factual Accuracy*

In the visualization of the factual accuracy ratings of the Singh dataset, clustering around High and Mixed ratings frequently occur, suggesting that recommendations are more likely to lead to videos with a similar factual accuracy rating (Figure 6.2). As those ratings typically correlate with MBFC's determination of political bias, similar clustering occurs. The general mass of nodes at the center of the graph are shown to have a majority High level of factual accuracy, while a cluster to the right bearing a great deal of political titles is populated with nodes mostly marked as having a Mixed level of factual accuracy. Outside of differing orientations, the

Trudeau and Scheer networks display similarly, showing factual accuracy clusters, which are likely at least in part a product of strong relationships between videos from the same channel.

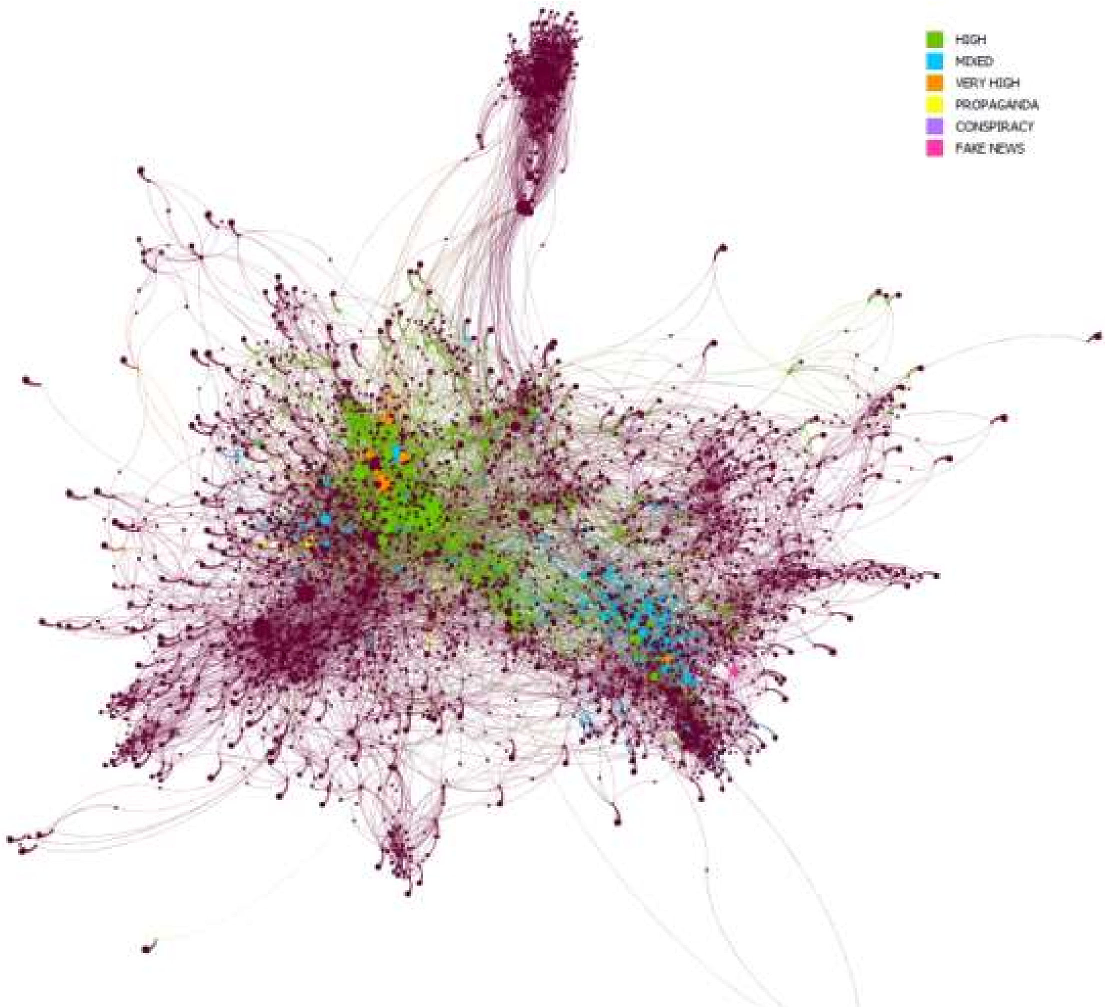



*Figure 6.2: Factual Accuracy of Sources in Singh Dataset*

*Media Bias & Topic Clustering*

The Trudeau dataset is used as an example to study the titles of videos found in the most noticeable clusters reveal some of the most insular and closely networked topics (Figure 6.3, Figure 6.4). While the great mass in the center of the graph is a tangled web of almost unrelated topics (sports, automotive, entertainment), some nodes are thrust out of the noise by the visualization algorithm (Force Atlas 2) to display communities that are highly independent, mostly only recommending each other. Right above the general mass is a jungle of political videos, comprising of the separated halves, videos with a Right bias, and those with a Left. The top-left of the graph shows a dense cluster of videos from the channel *Last Week Tonight with John Oliver*, while the top-right has a grouping of videos related to the TV show *Cant' Pay, We'll Take It Away*. The middle-right side of the graph shows a less-connected but substantial clustering of titles that can generally be regarded as clickbait, containing many list-based videos and other similar content. Just below this cluster is a small group of videos related to poker. Finally, there is a noticeable clustering of videos to the left on the graph containing videos from the channel *TEDx Talks*.

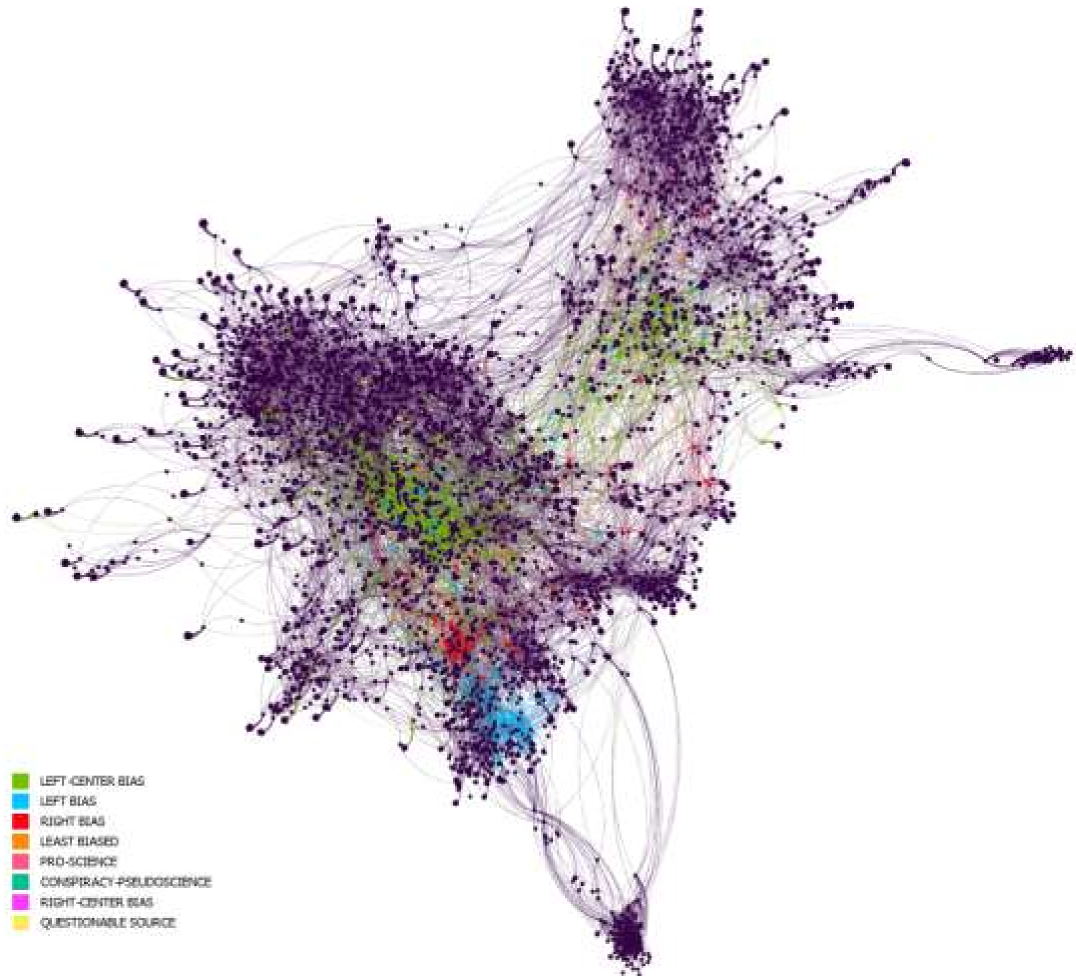



*Figure 6.3: Media Bias of Sources in Trudeau Dataset*

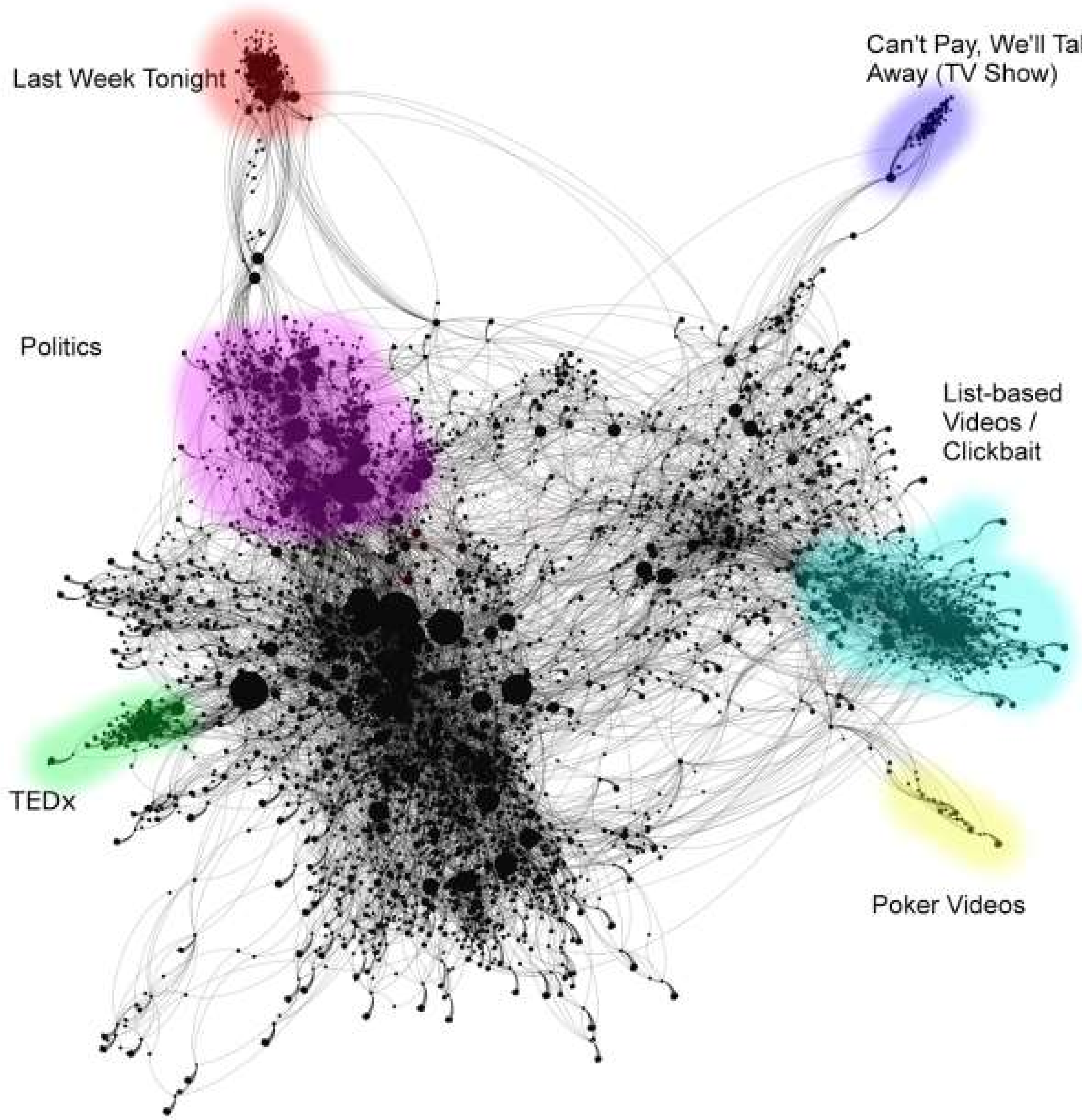



*Figure 6.4: Topic Clustering in Trudeau Dataset*

<u>*Channel Networking*</u>

Finally, the Scheer dataset can be used to demonstrate a visualization of the ten most occurring channels (Figure 6.5). Heavy clustering took place in almost every single instance, indicating a relationship between videos from the same channel. Two outliers, CBC News and ABC News, could be found in several clusters on the graph.

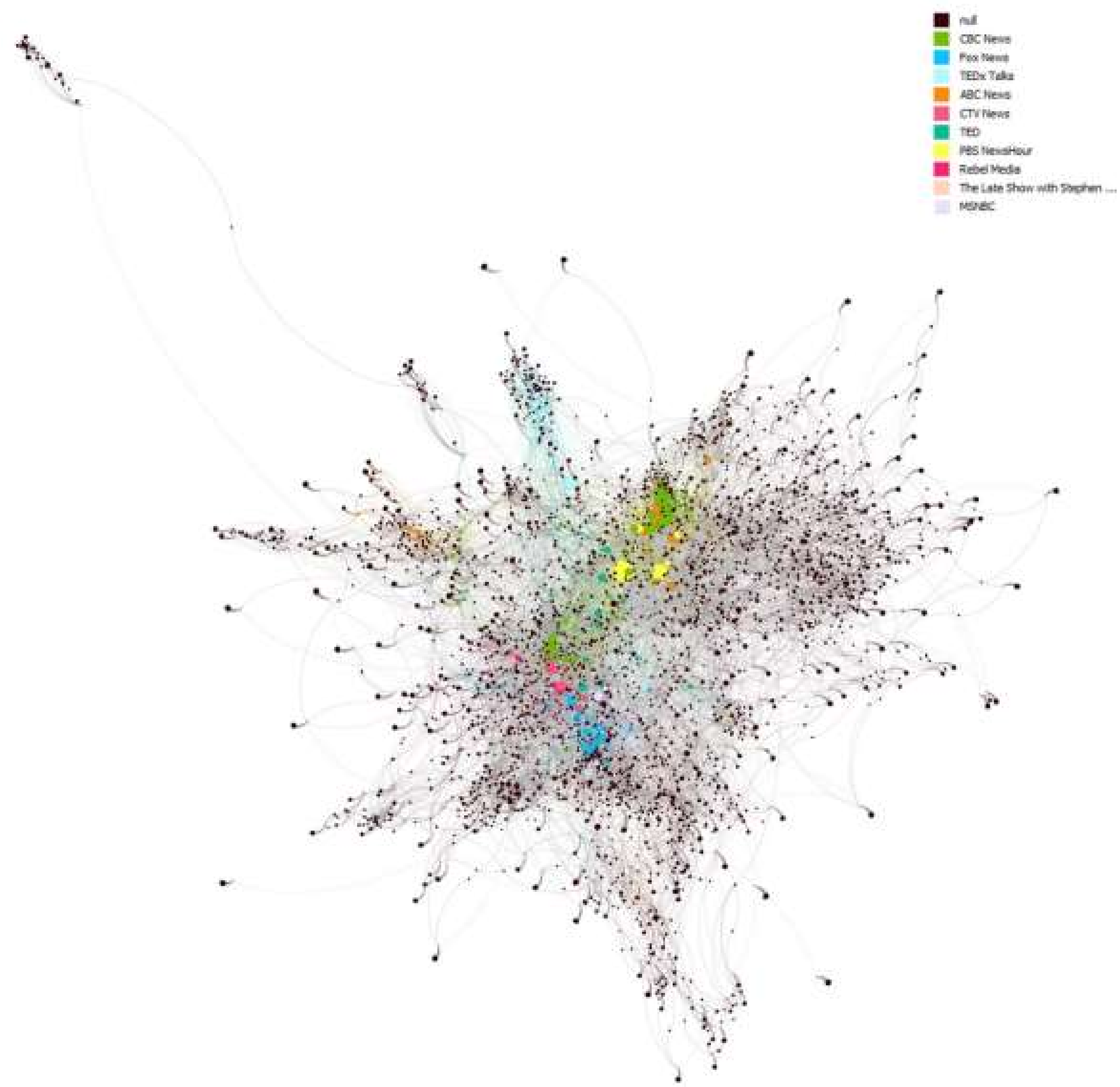



*Figure 6.5: Channel Networking of Scheer Dataset*

*Analysis*

As revealed above, titles with similar factual accuracy and bias ratings were often grouped together. This is not unexpected, but perhaps is better explained by the general clustering of channels. That is, a recommendation is most likely to lead to a video from the same channel, thereby explaining the clustering of factual accuracy ratings. If this were always true, however, one could expect small pockets of these groupings dotting over the graph, instead of the observed general clustering of both factual accuracy and bias ratings. This suggests that YouTube does group ideologically similar channels together. Or, put another way, viewers are more likely to click on content that matches their own ideology, a preference that ultimately teaches the algorithm how to create the most profitable viewer 'paths'. As previously discussed, this behaviour is expected by selective exposure, a theory that suggest humans tend to avoid content that contradicts their pre-existing beliefs. This model helps to explain the relationship between ideologically similar channels--as users actively pursue recommendations that do not challenge their preconceptions this shapes what the recommendation system is most likely to recommend in the future.

Looking at the mapping of political bias in the Trudeau dataset, I notice a trend that does not appear to be replicated across all three datasets. As the network is directional and recommendations are shown radiating outwards from the 'center' of what YouTube recommends, I can say that the sources that YouTube recommends generally have a Left-Center bias, a claim that has already been borne out in analysis. As stated previously, the clustering into various communities of interconnected topics is shown neatly by clumps of nodes. In this dataset, the second and third most visible biases are Left Bias, and Right Bias, which both cluster independently. The trend observed is that videos with a Left Bias cluster just beyond the videos with a Right Bias, indicating that YouTube is more likely to recommend a video that leans to the right before it serves one that leans to the left in this particular case. In other words, the average trajectory of a user using these keywords might consist of watching a video with a Left-Center bias, be recommended a video with a Right Bias, and only then be recommended a video with a Left Bias after that. This is where the visualization provides an advantage, as this kind of observation is difficult to make when comparing raw numbers.

## Conclusion

In order to collect recommendation data, I used Guillame Chaslot's Python software YouTube-explore. As a result of searching for the three primary candidates in the 2019 federal election in Canada, the software generated three databases each containing 1000 titles and 19,000 associated recommendations. By coding the 1000 recommendations in each dataset according to candidate representation, candidate depiction, Alternative Influence Network presence, factual reporting, and media bias, I was able to draw out observations and claims about YouTube's recommendation system and the people that use it. I was also able to use my work to study which sources YouTube has been promoting with its algorithm, including those that may negatively affect society.

An immediately obvious point of interest is content related to Canadian politics. Given the significantly higher number of videos related to the candidates in the Scheer dataset, I postulated that the difference between videos recommended in the Trudeau and Scheer datasets was a result of the recommendation algorithms' measurement of audience. As explored in Chapter 3, Google engineers refer to the collaborative signals the system makes use of as co-visitation (Davidson et al., 2010). While the seeds (personalized recommendations) tune the data for individuals, the activity data of other users (clicks, watch time, likes, dislikes, etc.) also plays a part in what everyone is recommended. Pragmatically, this means that audiences have a large part to play in recommendations. In the context of this research, the viewers searching for videos related to Justin Trudeau are not necessarily the same viewers searching for videos related to Andrew Scheer, and the algorithm reflects this difference in audience by way of what it is recommending. This observation provides insight to other irregularities in the data as well. For example, it would suggest that the proportionally higher number of videos related to members of the Alternative Influence Network in the Scheer database is again reflective of audience. As nearly all the AIN figures can be described as 'right-leaning' (some to an extreme extent), it follows that an audience watching videos related to a right-leaning candidate are also more likely be watching videos from or about these individuals. More overtly, the measurement of source bias reflected audience influence as well, showing that the Scheer dataset contained the most sources coded as 'right'.

Additionally, filtering by co-visitation also has an impact on what is not shown. For example, collaborative filtering may help explain the lack of strong bias in my datasets in comparison to the overwhelming slant as observed in *The Guardian*'s findings. It is not only differences in topics that reflect differences in recommendation, but also differences in audiences. To further emphasize this point, the relative prominence of both the *Rebel Media* and *CBC News* channels in my analysis show a clear preference for Canadian content in searches that relate to Canadian politicians. To reiterate, the recommendation system does not 'know' that Andrew Scheer is a Canadian politician - it only knows that videos bearing his name are associated with certain keywords, and that users are more likely to click on videos that are more relevant to them and their own context. Knowing that searches for Canadian political leaders are more likely to come from Canadians, a less-extreme tilt towards Andrew Scheer and away from Justin Trudeau may reflect a less-polarized political climate in Canada.

As stated in the introductory chapter of this thesis, there are two questions which this research seeks to answer, the first being 'Does YouTube's recommendation algorithm display bias towards any one of the three primary candidates in the 2019 Canadian election?' The answer to this question lies in the results of the 'Candidate Depiction' section of this chapter: In the research environment at the time of collection, YouTube's recommendation algorithm was more likely to promote videos that were critical of Justin Trudeau than it was to recommend videos that portrayed him in a positive or neutral light. Over all three datasets, both Singh and Scheer did not receive a single recommendation that presented them in a negative light. In contrast, videos critical of Trudeau were served over 200 times. Scheer was the only candidate to receive supportive recommendations, getting a total of 13.

However, context adds complexity. For example, only 40 recommendations in all three datasets referred to Scheer at all, only 10 to Singh. In contrast, 439 recommendations related to Trudeau, the current Prime Minister. 53% of recommendations related to Trudeau were critical of him, 47% portrayed him neutrally. Because Trudeau as a topic has a stronger presence in my data than the other candidates, it would follow that the greater field of videos relating to him might contain increased criticism, particularly since he was and is now the incumbent Prime Minister. Given the significant difference in recommendations between the candidates, it is very difficult to say whether either Scheer or Singh would receive a similar number of critical recommendations had videos related to them been more prevalent in the datasets. The relatively small sample size and contextual differences between these candidates leave me with an inability to responsibly make broad generalizations. In other words, the findings expressed here apply only to the data analyzed here--other searches will result in different recommendations.

However, I can draw conclusions from the data regarding the specific context in which the question was situated. As a reminder, I seek to find out whether YouTube's recommender system (with the technology that it employed at the time of collection) contributed towards a biased representation on their platform against or for Trudeau, Singh, or Scheer. For reference, *The Guardian* found that 84% of the videos they deemed relevant were supportive of Trump, with only 16% being supportive of Clinton in the lead up to the 2016 American election. In the analysis of my own data, I found evidence of a similar candidate bias, albeit to a lesser extent: 44.4% of all relevant recommendations were critical of Trudeau, and the only candidate to receive positive coverage was Andrew Scheer.

The second question that this research explores relates to context: What types of sources are being recommended alongside videos related to the three primary candidates in the 2019 federal election in Canada? As demonstrated by my wide range of analysis, there are many ways to answer this question. Taking stock of YouTube's recommendations at this point in time allowed me to make claims regarding many different aspects of the videos recommended as a result of searches for Canadian political candidates. Some of the findings are clearly specific to the situation, particularly those that demonstrated a bias towards Canadian content. Clearly, not every search on YouTube will net Canadian content from the seed videos. Still, this specific finding points towards a more generalizable truth: YouTube's recommendation system is highly influenced by viewers. This finding highlights the importance of guideline enforcement for YouTube; If users dictate what is being recommended, YouTube's responsibility towards ethical platforming is all the more paramount as the content permitted on the website can similarly influence users in a circular way.

Furthermore, other findings also revolve around user-system interaction: The bias towards Left-Center media sources is likely representative of a general, tech-savvy audience, the clustering of almost every metric examined on the network visualization graphs show fragmented user bases, and the AIN's reliance on collaboration demonstrates how the algorithm can be abused to push niche content to a mainstream viewership. In the end, this is ultimately what is problematic about this algorithm and others like it: Co-visitation or other markers of assumed interest can cause filter bubbles and be manipulated by bad faith actors and producers of hate speech. The following and final chapter will provide a critique of the collaborative co-visitation model employed by the algorithm and will focus on the bigger picture of how these findings can be useful, including the implications it could have on various parties, ideas related to possible future work, and potential solutions to the problems identified.

# CHAPTER SEVEN: CONCLUSION

## Introduction

Drawing from my own research and that of others, there are some generalizable claims I can make about bias in YouTube's recommendations. However, I do not find sufficient evidence to indicate a corporate conspiracy to boost or prioritize a particular political worldview. In fact, the bias found in the recommender system proves not to be overtly political, but merely human. YouTube's decision to hand the wheel over to users when it comes to recommendations is dangerous--and if anyone can influence what is being recommended--*anyone* can influence what is being recommended. Regardless of the cause of bias in YouTube's recommendations, it is clear that the flaws in the system have a negative impact on society. In the context of my research, the repercussions include the promotion of mis- and disinformation and an unequal representation of political candidates, both of which result in uninformed voters. Uninformed voters are not able to make decisions that are in their best interest.

The failure of YouTube's recommendation algorithm is a reminder of McLuhan's insistence on the power of systems over content: Individual videos may cause societal harm; the system that pushes harmful videos to hundreds of millions of viewers is infinitely more damaging. Modeling my own argument after his, it is possible to build a series of statements that lead to a logical conclusion--if YouTube as a platform has a great deal of influence over the public (this is self-evident), it can be said that the moderators of this platform have an ethical responsibility to the public not to promote videos that can negatively affect the public as a whole. By actively pushing harmful content to users, the recommendation system goes further than only the practice of hosting those videos in the first place, and by extension bears more ethical responsibility. Again, harmful content on its own is not as problematic as a system that promotes it, even done so without malicious intent.

In this chapter, I will pose five primary critiques of YouTube and their video recommendation system, as well as providing several suggestions for each stakeholder involved, including YouTube as a company, the public, and governmental regulators. For YouTube itself, I explore various solutions provided by expert sources, concluding that a push towards transparency policies are the most pragmatic way forward. For the public, I suggest two approaches, both of which assume that YouTube continues to ignore its problematic system: One, that the public must work to educate themselves and others about YouTube's RS, and two, that YouTube creators could attempt to exploit the system in a similar way to the alt-right, assuming that political content can be engaging regardless of ideology. Finally, I stress the role that governments must play in creating modern laws that address the problems associated with modern technology.

## Findings

The most significant findings that I am able to draw from my research relate to influence: The nature of collaborative filtering (here, co-visitation) implicates the individual in the shaping of the broader experience on the platform. As I postulated that different audiences account for the differing results per political candidate in my data, I will also conclude that YouTube's use of the co-visitation metric to make recommendations make way for those who wish to exploit it. In some ways, the algorithm's use of collaborative filtering to help improve watch-time is a way of crowd-sourcing their automated marketing: There is no human gatekeeper deciding what is recommended. YouTube's impact on society is mixed; There are many positive aspects of YouTube, and the weaknesses highlighted here can also be seen as strengths. For example, just as it is easy to say that the erasure of lines between creator and consumer could have negative effects on how viewers perceive truth, it is simultaneously important to acknowledge that YouTube's break from established power structures has given a platform to many groups that have been marginalized by traditional gatekeepers, including LGBTQ+ communities, people of colour, and those living with disabilities. Regardless of the popularity of these groups on YouTube, at the very least they have the opportunity to express their opinions on video in a public forum without relying upon traditional structures to platform them. If YouTube's community guidelines are enforced to their fullest extent, the freedom of expression afforded by the website is invaluable to those seeking representation. With minor exceptions, anyone can post a video on the platform that anyone can watch. Similarly, fears concerning the use of YouTube as an educational platform can be positioned as a positive as well. While 15% of adult YouTube users in the U.S. encountered information that 'seemed obviously false or untrue', it is also true that 51% of those same users said that the website was 'very important when it comes to figuring out how to do things they haven't done before' (Atske, 2018). While countless examples of fake news can be found on the platform, it is perhaps too easy to dismiss the overwhelming weight of useful information that is *not* talked about.

However, my research is not concerned with the positive aspects of YouTube, and instead focuses on their problematic recommendation system and its negative impact on the public. The primary cause for concern is the promotion of dis- or misinformation. As discussed in the theoretical framework chapter, YouTube's recommendation system is designed to target human biases for profit. Beyond the inherent evils of such exploitation is the potential for bad actors to hijack the system to achieve their own ends. In the book *Post-Truth,* author Lee McIntyre has thoughts that could easily apply to the festering proliferation of alt-right-adjacent YouTube personalities: "Our inherent cognitive biases make us ripe for manipulation and exploitation by those who have an agenda to push, especially if they can discredit all other sources of information" (2018, p. 62). By using an ethically blind system, YouTube facilitates the degradation of truth, a fact that is made evident by numerous incidents described in Chapter 2. This dangerous implication has a ripple effect that could threaten larger social structures, as identified by Eli Pariser and others. The noted sociologist Jürgen Habermas puts it plainly: "A 'post-truth democracy' [...] would no longer be a democracy" (2006, p. 18). Or, as outlined in the ALA's core values of librarianship, "A democracy presupposes an informed citizenry" ("Core values"). All in all, it is not difficult to see that the flaw in YouTube's recommendation system could negatively impact society, and by extension, governments. There are several implications for YouTube as a company as well: The difference is that YouTube is unlikely to change course unless a more financially advantageous path becomes clear or if governmental regulation is imposed. YouTube's past deflection of bias in recommendations as being a "reflection of viewer interest" (Lewis and McCormick, 2018) is indicative of their approach to the problem: As long as YouTube are able to shift blame to the very audiences they help to create, they will do so.

As evidenced by many others and through my own research, YouTube still hosts and recommends videos from individuals who actively promote hate. It is understandable that the daily influx of videos simply makes fully human moderation impossible, and furthermore it is acknowledged that this is not an easy problem to solve, nor am I the first to confront it. Armed with my own evidence and the findings of others, it is possible to put forward a series of potential solutions to YouTube's problematic content recommendation issue.

## Problems & Potential Solutions

### *Primary Critiques*

The following are the major issues identified with YouTube as a platform and its recommender system, accompanied by my own recommendations on how to address those issues.

1. Factually blind recommendations

There is currently no distinction between authoritative and non-authoritative sources. As an information source (and increasingly, a primary source of news for many), it will be argued here that YouTube has a responsibility towards factual accuracy when making recommendations. This means clearly defining what they consider authoritative journalistic sources and boosting their visibility by surfacing those videos more often in recommendations. A similar effort involving the addition of print sources to breaking news events on YouTube is already underway (Kyncl & Mohan, 2018). As discussed earlier, the case of hosting is quite different. When a user uploads a video to YouTube, the content is only 'approved' if it abides by the rules set out by YouTube's community guidelines and is not flagged by other users or removed by automated services. A video is given greater weight when it is promoted by the recommendation algorithm, regardless if that approval is understood to be implicit or explicit; implicit in the semantic sense, and explicit in its literal promotion of the video to a larger audience. Moving forward, YouTube must mark 'verified' news outlets and agencies that meet a rigid and uniform list of journalistic standards. These channels should be given prioritization in the recommendation system.

2. Over-reliance on community policing / Inconsistent enforcement of rules

Unpaid groups and individuals are asked to make many of the rule enforcement decisions that should instead come directly from YouTube. While it may seem hypocritical to simultaneously criticize the platform for having a machine-centric distribution model and at the same time an overly 'human' approach to content removal, there is a unifying problem: both are reflections of the company's hands-off approach to curation. Instead, YouTube needs to employ more human/algorithm hybrid methods of removing content in a way that takes responsibility for what they host. For example, Google should turn towards solutions proposed by the academic community, including Agarwal & Sureka's "focused-crawler based approach for identification of hate and extremism promoting videos on YouTube", a method that uses social network analysis to expose central nodes of hate-filled communities (2014). Instead of focusing on individual offences, YouTube should consider undertaking a more holistic approach that roots out channels and individuals that play a large role in spreading hate on their platform.

Furthermore, all efforts should be made to become more proactive about rule enforcement - YouTube should not wait for public outcry to delete the channels and videos of problematic content creators. The examples proffered in Chapter 2 highlight YouTube's double-standards when it comes to their community guidelines. One could tentatively suggest that YouTube is less likely to remove popular content that they can ultimately profit from. For example, the channel of the white nationalist Stephan Molyneux (with nearly a million subscribers) should be terminated immediately for repeated violations of community guidelines.

3. Use of co-visitation metrics and ease of manipulation

YouTube's use of collaborative filtering effectively outsources advisory to a perceived greater public; in reality the power shifts to hyper-engaged users and those who know how to exploit the algorithm. The recommendation system can and has been exploited by various parties, many of which who ultimately are damaging to society. Extremists, foreign governments, and other actors have been able to affect what is being recommended to the general public by creating borderline content and finding ways to create automated pathways of recommendations. As a tool with such a large reach is an ideal target for 'autopropaganda', YouTube has a unique responsibility as they find themselves in an information gatekeeping role. The solution to this problem is complex, and it is likely that an engineer with intimate knowledge of the algorithm's inner workings may be able to provide a better answer. Personally, I believe that borderline content will always exist on the platform, regardless of where the line is drawn. Controversial material will continue to generate attention online. Nothing can be done about this content other than determining what content fits in this category and to restrict it from being recommended, as YouTube has begun to do. While I recognize the benefits of using collaborative filtering to source recommendations from like-minded viewers, I personally believe an effort must be made to introduce a degree of serendipity to what is being recommended, even beyond what Reinforce is purported to do. Primarily, the problem with using a collaborative recommendation model is that it will slowly introduce content to an audience not actively searching out for it. For example, I will suggest that many young men in particular do not seek out white supremacist or anti-feminist content on YouTube, but instead are pulled towards it by a system that takes its cues from other radicalized users. If recommendations serve as paths to other videos, then their weighting is determined by the ruts and grooves created by the viewing patterns of other users.

4. Lack of transparency

Without necessary information, users are unable to make informed choices regarding their use of the platform and researchers cannot effectively make conclusions about how the system might affect society. Using the detailed transparency guidelines as supplied by Chaslot and Diakopoulos in the ‘suggestions for stakeholders’ section following, I advocate for the implementation of a standard transparency release law for recommender systems that ensures the public is able to access information regarding the systems they use every day. Alternatively, it is possible to imagine a user setting dedicated to customizing how content is recommended. For example, users could choose if content was recommended based on channels they have watched, topics they have expressed interest in, or even more serendipitous content like trending videos. In any case, users currently do not have enough information about what is being recommended to them and why, making the algorithm more comprehensible and giving users a chance to choose how they are being marketed to would be a huge improvement to transparency.

5. Corporate motivations

Far from a feasibly addressed criticism, my final qualm is with the much larger structure of capitalism that directs the company’s future. The inherent need for constant growth means that YouTube will never make adjustments to their platform unless it ultimately results in financial gain. Further ethical considerations will only be a result of public image issues, highlighting the need for continued research and journalistic investigation. Improving YouTube regardless of revenue loss is necessary, and a small price to pay in the face of radicalized users. Munger and Phillips (2019) suggest that this solution is wishful thinking and ‘flattering’ to journalists and academics, sarcastically implying that there are ulterior motives behind this line of thinking: “If only Google (which owns YouTube) would accept lower profits by changing the algorithm governing the recommendation engine, the alternative media would diminish in power and we would regain our place as the gatekeepers of knowledge” (2019, p. 3). Instead, Munger and Phillips suggest that YouTube as a platform has special affordances like the ease of content creation and ability to form parasocial relationships with creators that make it ideal for alternative media, and any content issues cannot be fixed by ‘algorithmic tweaks’. However, these researchers do not reject the possibility of radicalization by the recommender system, but instead minimize its reach and potential effects. While I will concede that Munger and Phillips have sufficiently demonstrated a decline in alternative media on YouTube, I will maintain that radicalization has occurred directly as a consequence of a corporate decision to pursue profit without considering negative outcomes. Admitting that intentional profit loss is unlikely, the only other feasible adjustment comes externally from regulatory bodies.

*Suggestions for Stakeholders*

Crucial to any proposed improvement of YouTube is the concept of ethical algorithms. Using Merrill's media ethics definition, ethics is "the study of what we ought to do" (2011, p. 3). Author of "Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness", Mike Annany states explicitly why users should care about how algorithms decide what humans ought to do: "Algorithmic categories raise ethical concerns to the extent that they signal certainty, discourage alternative explorations, and create coherence among disparate objects" (2015, p. 103). Addressing these concerns are systems with built-in considerations for ethical values. One of the most prominent examples is Google's Search algorithm, which in 2017 had to be adjusted to boost 'authoritative content' to prevent holocaust denial websites from being promoted (Roberts, 2016; Sullivan, 2017). Similar weighted filters have recently been put in place on YouTube, for much of the same reasons (Herrman, 2018). As for other potential solutions, arguments have been made for transparency (Helberger et al., 2016). Similarly, Kraemer et al. advocate for leaving ethical issues up to users where possible, or, at least make the ethical assumptions built into the system as easy to identify as possible (2011); Helberger et al. argue for a 'diversity-sensitive' design principle to counter filter bubbles.

Furthermore, Middlestadt et al. survey literature on ethics and algorithms to conclude that "a mature 'ethics of algorithms' does not yet exist, in part because 'algorithm' as a concept describes a prohibitively broad range of software and information systems" (2016, p. 14). This is elaborated on by Ananny (2015), who gives three reasons why such a framework is difficult to put together:

1. Algorithms are created by professionals with shifting boundaries
2. Outputs cannot be understood or controlled by a single individual
3. Increasing personalization make results difficult to measure

While it is tempting to make broad generalizations from literature on algorithms, Ananny's warning of assumptions regarding the stability of technology give reason to tread carefully. Despite this caution, each of the three primary stakeholders in YouTube's continued success are able to make changes that can help to address the identified problems with YouTube and its recommendation system. In an ideal world, YouTube would abandon collaborative filtering and switch to content-based filtering, recommending videos on the basis of individual preference; similarity between channels and keywords could be used to indicate what other videos a user might be interested in. By placing every user in a vacuum, the algorithm could avoid basing the individual's recommendations on other user's viewing patterns. However, as recommended videos constitute more than 70% of YouTube's daily traffic (Zhou et al., 2010), a shift away from collaborative filtering likely means profit loss due to decreased ad revenue. Without external pressure from regulators, a change to the way the system recommends videos is unlikely. As a result, I believe Diakopoulos' push for transparency the most pragmatic move forward.

1. *YouTube*

The ability to make positive social impact with algorithms is highly dependent on transparency (Lepri et al., 2018). On July 14th, 2019, ex-YouTube engineer Guillame Chaslot tweeted a list of recommendations for various groups to combat how YouTube can "amplify our worst inclinations" (gchaslot). Expanding upon his recently published *Wired* article, Chaslot provided two recommendations for platforms like YouTube: "Be more transparent about what your AI decides[.] Align your 'loss function' on what users really want, not pure engagement" (gchaslot, 2019, July 14). In a *Medium* article co-written with Andrea Gorbatai, an assistant professor at UC Berkeley, Chaslot is more specific about what information the platform should supply to the end-user. First, users should be given the option to see a random recommendation each day. This would "enable the public to assess YouTube's A.I. general alignment" (Chaslot & Gorbatai, 2016). In other words, the function would give users the ability to see the 'neutral' version of the algorithm, without factoring in personalization. Second, the public must be given access to three data points: "the number of views, the number of views resulting from AI recommendations, and the total number of recommendations". Armed with this information, users would have the ability to discern bias in what is being recommended. Chaslot's suggestions towards transparency echo the solutions proposed by other academics. In his paper "Algorithmic Accountability", author Nicholas Diakopoulos argues that the hidden design of algorithms is at odds with transparency, for trade secrets "seek to hide information for competitive advantage" (2014, p. 403). Going on, he suggests that transparency does not necessarily translate to informed decision-making, since source code availability is of little use to those without adequate technical expertise. Despite this, Diakopoulos sees this algorithmic accountability as "a mechanism for elucidating and articulating the power structures, biases, and influences that computational artifacts exercise in society" (p. 398). Like Chaslot, Diakopoulos has a list of information that he believes algorithmic overseers should release as part of a transparency policy:

A. The criteria used to rank things in the algorithm
B. What data act as inputs to the algorithm
C. The accuracy including the false positive and false negative rate of errors
D. Descriptions of training data and its potential bias
E. The definitions, operationalizations, or thresholds used

As stated prior, incentive is key. YouTube will not release information about their trade secrets if they are not forced to through legislation or public outcry.

Another alternative solution for YouTube is to increase counter-messages. In 2017, YouTube started redirecting searches for violent extremist content to playlists of videos that debunked and challenged those ideas ("Bringing new Redirect", 2017). Currently, YouTube also automatically adds fact-checking links to some breaking news stories and videos relating to controversial topics such as the holocaust and vaccination (Glaser, 2018; O'Donovan, 2019; Paul, 2019). To decrease radicalization on the platform, YouTube could further implement counter-messages, defined as "anti-extremist messages in the same environment in which extremist messages occur" (Schmitt, 2018, p. 2). By proactively challenging common 'alternative facts' before users are further down radicalization pathways, it may become possible to reduce the propagation of hateful ideology.

*2. The Public*

Chaslot also has suggestions for the public: "Stop trusting Google/YouTube blindly[.] Their AI is working in your best interest only if you want to spend as much time as possible on the site. Otherwise, their AIs may work against you, to make you waste time or manipulate you" (gchaslot, 2019 July 14). While this solution may be easier said than done, it does highlight the public's information/digital literacy challenges and should be addressed by those in learning environments, including libraries.

As author and audience become increasingly one and the same, there is also an alternative solution: left-leaning voices could also make use of exploitative strategies to 'game' the recommendation system as the right has. As identified in the *New York Times* article "The

Making of a YouTube Radical", there is a small but growing faction of left-wing creators known as 'BreadTube' who employ similar tactics as the far-right: "The core of BreadTube's strategy is a kind of algorithmic hijacking. By talking about many of the same topics that far-right creators do--and, in some cases, by responding directly to their videos--left-wing YouTubers are able to get their videos recommended to the same audience" (Roose, 2019). Evidence of this crossover is also found in Rebecca Lewis' Alternative Influence Network report, in which the left-leaning YouTuber *Destiny* (Steven Bonnell) is lumped in with alt-right voices due to his frequent interactions with other members of the network. To play devil's advocate, it is also possible that 'Breadtubers' will never be able to reach the same levels of popularity as other groups as there is a correlation between audience engagement and borderline content (Maack, 2019).

*3. Government*

In some cases, corporations can bend to national regulators. Citing IBM's recent call for algorithmic legislation (Hagemann, 2019), Guillame Chaslot also provided some suggestions on how governments should approach YouTube's failings: "Create a special legal status for algorithmic curators[.] Demand some level of transparency for recommendations. This will help understand the impact of AI, and boost competition & innovation (gchaslot, 2019 July 14). Although vague, Chaslot's suggestions for legislative oversight serve as another reminder of how far lawmakers have fallen behind technology. Far from being overly restrictive or damaging to profits, these proposed solutions are a step towards ethical technology and should be fleshed-out by advocacy groups and presented to government agencies.

## Additional Remarks

*Evaluation of Python Software*

Guillame Chaslot's Python script YouTube-explore was an effective tool that made this study possible. The ability to adjust how many branches were crawled and how many levels were explored was a boon to my needs, and it is easy to see how the option to sort by most-viewed might be valuable in other situations. Despite its usefulness, the software faces many accessibility issues and deserves further development. For example, the initial setup of YouTube-explore involved installing the Python interpreter itself, learning the basics of the Python language, and installing requisite dependency packages. While this task may be easy for those experienced with the language, the choice to not release this software as an executable may be daunting and become a barrier to access for some researchers. Furthermore, little in the way of documentation could be found in the installation package, and any additional information that existed had to be supplied from the *GitHub* page for the project. Accessibility is also an issue given that the software does not have a user interface and must be operated from the command line. Finally, some work should be done to expand the options available within the program itself; data formats outside of JSON like CSV may reduce the need for conversion and allow less tech-savvy researchers to immediately access the data. Additionally, it would be ideal if YouTube-explore could also scrape the date each video was published in the metadata it returns.

A fully fleshed model for this software already exists in the form of *algotransparency.org*, Chaslot's own website dedicated to tracking what YouTube is recommending to its users. Upon opening the website, an informative series of statistics and visualizations help highlight the importance of the project. Once on the main page, the most recommended videos of the current day are listed, including viewcounts, thumbnails, and other distinguishing factors. A search bar at the top of the page allows users to filter those most recommended videos by keywords. Of particular use is the 'Themes' page, which allows users to browse various controversial topics, including searches for world leaders and elections. Although custom searches cannot be made, Chaslot's website serves as a clean, accessible, and informative introduction to YouTube's recommendation algorithm and how it can promote a variety of misinformation. Given the resources, *algotransparency.org* could be modeled after to create a visually improved and user-friendly application.

*Future Work & Undiscussed Topics*

Another database of videos was created in October of 2018 with the same search terms in an attempt to test Chaslot's software. Future work could be done to conduct a similar analysis of this database in order that the two points in time could be compared. Another database was created a few days after YouTube's announcement that they would start cracking down on recommended videos that contained borderline content. This also could provide some contrast to the February 2019 dataset, and perhaps help answer the question of whether YouTube's promises were fulfilled. However, there is one caveat: These two earlier searches were only conducted on about 800 videos. Furthermore, it is suggested that future work in this field could alter the methodology used to focus on different aspects of the research question. For example, running multiple searches at multiple dates might produce more useful aggregate data, making results more generalizable. In addition, focus on average experience trips could be achieved with more in-depth analysis of depth data captured by the software, particularly enhancing research on the way videos from the alt-right and other groups are being recommended.

Another direction that future study could take is made possible by an interesting feature coded into Chaslot's software: sorting by most viewed. While research on the most viewed videos on any given topic do not necessarily contribute towards an understanding of how the recommender system works, it does highlight viewership bias and reveals popular sentiments of the time period. In a more time-sensitive way, comparative work could and should be done closer to the October 2019 election date. A database was created on October 20th, the day before the election. This additional information may shed light on the state of the recommendation algorithm, and certainly might provide insight to the ever-looming question of external interference or manipulation. As the analysis of sources revealed a bias towards Canadian content with a prevalence of both *Rebel Media* and *CBC News*, further research could be done to replicate a search for Canadian political candidates with a variety of IP addresses to determine how much of the Canadian perspective is brought about by the search keywords in contrast to the search location. Regardless of focus, it is also recommended that any future work should include a control dataset, drawn from a search using less politically charged keywords.

Additionally, future work should be done on the leader of the newly founded People's Party of Canada (PPC), Maxime Bernier. Given the perceived alignment between the new right-wing populists' political positions and the type of video currently under the spotlight on YouTube, an investigation into how his platform is propagated amongst right-leaning channels should yield interesting results. However, Bernier's loss in his own riding is likely to mean bad things for his new party and may not be worthy of study in the near future.

Research should also be done to shed light on YouTube's promotion of the self-labeled 'Intellectual Dark Web', a group of rebellious thinkers who pride themselves in their goal to discuss the 'undiscussable'. Notable members include Dave Rubin, Sam Harris, Ben Shapiro, Jordan Peterson, Joe Rogan, and more. Citing the data I have collected here, these figures are heavily recommended on YouTube and many often find themselves part of Rebecca Lewis' Alternative Influence network. By extension, the Intellectual Dark Web hold a great deal of influence on YouTube at the moment and that power should be questioned as well as YouTube's role in the dissemination of those ideas. As outlined in the previous discussion about AIN, many of these individuals openly display racist, sexist, or homophobic sentiments but choose to defend these opinions as merely being exercises in free speech. Finally, an expanded list of figures associated with the alt-lite, the alt-right, and the intellectual dark web could be drawn from Ribeiro et al.'s 2019 paper "Auditing Radicalization Pathways on YouTube", which details 349 channels that could be used for a more comprehensive study of radicalization pathways on YouTube.

## Conclusion

As a self-publishing platform that has enormous reach, little-to-no factual oversight, and a recommender system that relies on collaborative filtering, YouTube is a prime target for manipulation. There is an unspoken but mutually beneficial relationship underway; poorly disguised members of extremist movements are given a soapbox on which to spread their disinformation, and in turn YouTube profits from all the advertisement views said messages generate. For long, YouTube's position on this content has been ambivalent, for taking on a more editorial role would undoubtedly cut into established revenue streams. Recent promises indicate an upcoming change. Despite the harsh criticism levied against YouTube from all fronts, there have been several positive adjustments in the past two years. The decision to identify and exclude borderline content from recommendations is an excellent idea on paper, and if implemented correctly could mitigate much of the influence-related concerns outlined here. There is also hope in the latest iteration of the algorithm: Reinforce. If the system works as intended and does in fact introduce more serendipity into video recommendations, then this pursuit of variety in the name of more addictive viewing patterns could push against the rabbit-holes of hate that formed previously.

As established in the admittedly surface-level analysis of the group dubbed the 'Alternative Influence Network', these communities are strongly interconnected with interactions, features, and guest appearances between members (Lewis, 2018). In a 2014 study on Germany's far-right YouTube community, researchers drew similar conclusions and warned of an extremist filter bubble (O'Callaghan et al.). Despite YouTube's previously hands-off approach to borderline content, there has been a growing sentiment from some right-wing thinkers that YouTube is actively silencing conservative opinions on their platform. In the opinion of this author, there are two conflations present. Firstly, restricting both the audience and promotion of certain videos is not equivalent to censorship. Or, as Microsoft's *Data & Society* president Danah Boyd puts it, "Choosing what to amplify is not the same as curtailing someone's ability to speak" (Boyd, 2018). Despite YouTube's status as a private company and not a public platform for free speech, it still exists as a space for *everyone*, as evidenced in their own marketing. Secondly, the restriction or even removal of content that is intentionally misleading or hateful does to equate to censoring conservative opinions, although the overlap may be significant. Outcry related to the removal of noted conspiracy theorist Alex Jones is reflective of a strange dissonance that binds the removal of hate speech to political discrimination.

While the first chapter in this thesis worked to sketch out an introduction to the topic at hand, the second painted in the contextual foundations needed for a more comprehensive understanding. The literature review that followed illustrated how and why YouTube's recommendation system is designed to keep users watching. Similarly, the theoretical framework chapter laid out how recommender systems are built to exploit natural human behaviour, much like the media that came before it. Noting the everlasting cycle between human and technological influence, my own methodology in Chapter five was formed with the intent of separating the recommender system from the personalized aspects that dominate it. Finally, analysis of the results I gleaned from my own research primarily found that YouTube's recommendations tended to favour Andrew Scheer and cast Justin Trudeau in a negative light. Using my data, I was also able to conclude that the recommender system is susceptible to influence by audiences, which I propose is a result of collaborative filtering.

While freedom of speech, responsible platforming, and access to information are all fascinating topics that certainly have interplay with the topics discussed here, my focus has instead been placed on systemic failures in relation to information retrieval. However, I will acknowledge that YouTube's potential to radicalize audiences is not only a result of an ethically blind recommender system, but also a problem associated with content moderation. All said, the situation surrounding YouTube and its recommender system is a constantly evolving one, and the company's response to well deserved criticism over the past three years has been slow and reactive, yet promising. The temptation to propose or implement simple or static solutions ignores the reality of a complex problem that will always require an adaptive and proactive approach. But make no mistake: As algorithms become increasingly used to combat information overload, continued work must be done to study potential pitfalls in the paths ahead, for recommender systems have the power to foster informed and responsibly entertained audiences - without being used maliciously to push hate, propaganda, and other harmful content on a user base that is still suffering the growing pains from adjusting to a new information medium.

2017 Public report on the terrorist threat to Canada. (2017). Public Safety Canada. Retrieved from https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/pblc-rprt-trrrst-thrt-cnd-2017/index-en.aspx

ABC News (Australia). (1977). Marshall McLuhan - 'The medium is the message' - ABC Radio National [Video File]. Retrieved from https://www.youtube.com/watch?v=1YZzwvjwiLs

Agarwal, S., & Sureka, A. (2014). A focused crawler for mining hate and extremism promoting videos on YouTube. In Proceedings of the 25th ACM conference on Hypertext and social media (pp. 294-296). ACM. Retrieved from https://dl.acm.org/citation.cfm?id=2631776

Aggarwal, C. C. (2016). Recommender systems: The textbook. Cham: Springer, 2016.

Airoldi, M., Beraldo, D., & Gandini, A. (2016). Follow the algorithm: An exploratory investigation of music on YouTube. *Poetics*, *57*, 1-13.

Albright, J. (2018). #NotOKGoogle search suggestions: 2018 edition. Medium. Retrieved from https://medium.com/@d1gi/notokgoogle-search-suggestions-2018-edition-ba09eaf49fc2

Albright, J. (2017). 📺FakeTube: AI-generated news on YouTube. Medium. Retrieved from https://medium.com/@d1gi/faketube-ai-generated-news-on-youtube-233ad46849f9

Alexander, J. (2019). The golden age of YouTube is over. Verge. Retrieved from https://www.theverge.com/2019/4/5/18287318/youtube-logan-paul-pewdiepie-demonetization-adpocalypse-premium-influencers-creators

Alexander, L. (2016). Inside the strange world of million-view ‘surprise egg’ YouTube videos. Intelligencer. Retrieved from https://nymag.com/intelligencer/2016/04/inside-the-strange-world-of-million-view-surprise-egg-youtube-videos.html

Alvarez, E. (2017). After Las Vegas shooting, Facebook and Google get the news wrong again. Engadget. Retrieved from https://www.engadget.com/2017/10/02/facebook-google-fake-news-las-vegas-shooting

Ananny, M. (2016). Toward an ethics of algorithms: Convening, observation, probability, and timeliness. *Science, Technology, & Human Values*, 41(1), 93-117.

Background to ‘Assessing Russian activities and intentions in recent US elections’: The analytic process and cyber incident attribution. (2017). National Intelligence Council. Retrieved from https://www.dni.gov/files/documents/ICA_2017_01.pdf

Bahara, H., Kranenberg, A. & Tokmetzis, D. (2019). Aanbevolen voor jou op YouTube: racisme, vrouwenhaat en antisemitisme. *De Correspondent*. Retrieved from https://decorresp ondent.nl/9149/aanbevolen-voor-jou-op-youtube-racisme-vrouwenhaat-en-antisemitisme/445528853-0f710148

Bail, C. A. et al. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216-9221.

Beam, M. A. (2014). Automating the news: How personalized news recommender system design choices impact news reception. *Communication Research*, 41(8), 1019-1041.

Being a YouTube Creator just got even more rewarding. (2012). YouTube Creator Blog. [Blog Post]. Retrieved from https://youtube-creators.googleblog.com/2012/04/being-youtube-creator-just-got-even.html

Bergen, M. (2019). YouTube executives ignored warnings, letting toxic videos run rampant. Bloomberg.com. Retrieved from https://www.bloomberg.com/news/features/2019-04-02/youtube-executives-ignored-warnings-letting-toxic-videos-run-rampant

Berners-Lee, T. (2010). Long live the web. Scientific American, 303(6), 80-85.

Bimber, B., & Davis, R. (2003). Campaigning online: The Internet in U.S. elections. New York, NY: Oxford University Press.

Bobadilla, J., Ortega, F., Hernando, A., and Gutiérrez, A. (2013). Recommender systems survey. *Knowledge Based Systems*. 46(July), 109-132.

Boesveld, Sarah. (2017). A beer with Andrew Scheer: CPC leader, popcorn addict… feminist? *Macleans.ca*. Retrieved from https://www.macleans.ca/politics/ottawa/andrew-scheer-beer-interview

Borger, J. (2018). YouTube: we've found no evidence of Russian interference in Brexit vote. The Guardian. Retrieved from https://www.theguardian.com/media/2018/feb/08/youtube-says-no-evidence-of-russian-interference-in-brexit-vote

Boyd, D. (2018). *Media manipulation, amplification and responsibility*. Speech presented at the Online News Association 2018 Conference, Austin, TX. Retrieved from https://ona18.journalists.org/sessions/openingfeaturedconversation/

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, 15(3), 209-227.

Bridle, J. (2017). Something is wrong on the internet. Medium. Retrieved from https://medium.com/@jamesbridle/something-is-wrong-on-the-internet-c39c471271d2

Bringing new Redirect features to YouTube. (2017). Official YouTube Blog [Blog Post]. Retrieved from https://youtube.googleblog.com/2017/07/bringing-new-redirect-method-features.html

Britneff, B. (2017). Singh releases tax and income security agenda ahead of debate. iPolitics.

Retrieved from https://ipolitics.ca/2017/06/11/singh-releases-tax-policies-income -security-agenda-ahead-of-debate

Brown, C. (2018). How to understand the YouTube algorithm in 2018. Octoly Magazine. Retrieved from https://mag.octoly.com/how-to-understand-the-youtube-algorithm-in -2018-c435136abb97

Bucher, T. (2018). Cleavage-control: Stories of algorithmic culture and power in the case of the YouTube 'reply girls'. In Z. Papacharissi (Ed.), *A Networked Self and Platforms, Stories, Connections* (pp. 125-143). New York: Routledge.

Buolamwini, J. (2016). How I'm fighting bias in algorithms. [Video file]. Retrieved from https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in _algorithms?language=en

Butterfield, A., Kerr, A., & Ngondi, G. Ekembe. (2016). *A dictionary of computer science.* Seventh edition.

Calamur, K. (2018). What Is the Internet Research Agency? Atlantic. Retrieved from https://www.theatlantic.com/international/archive/2018/02/russia-troll-farm/553616

Camp Fire incident information. (2019). CAL FIRE. Retrieved from http://cdfdata.fire.ca.gov /incidents/incidents_details_info?incident_id=2277

Canadian Centre for Cyber Security (2018). *National Cyber Threat Assessment 2018*. Retrieved from https://cyber.gc.ca/sites/default/files/publications/national-cyber-threat -assessment-2018-e_1.pdf

Carlson, M. (2018). Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism. *New Media & Society*, 20(5), 1755-1772.

Changes to related and recommended videos. (2012). YouTube Creator Blog. [Blog Post]. Retrieved from https://youtube-creators.googleblog.com/2012/03/changes-to-related-and-recommended.html

Chaslot, G. (2017). YouTube-explore [Software]. Retrieved from https://github.com/pnbt/ youtube-explore

Chaslot, G., & Gorbatai, A. (2016). YouTube's A.I. was divisive in the US presidential election. Medium. Retrieved from https://medium.com/the-graph/youtubes-ai-is-neutra l-towards-clicks-but-is-biased-towards-people-and-ideas-3a2f643dea9a

Chen, M., Beutel, A., Covington, P., Jain, S., Belletti, F., & Chi, E. H. (2019, January). Top-k off-policy correction for a REINFORCE recommender system. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (pp. 456-464). ACM.

Computerphile. (2014, April 24). *YouTube's Secret Algorithm - Computerphile* [Video File]. Retrieved from https://www.youtube.com/watch?v=BsCeNCVb-d8

Confessore, N., & Wakabayashi, D. (2017). How Russia harvested American rage to reshape U.S. Politics. N. Y. Times. Retrieved from https://www.nytimes.com/2017/10/09 /technology/russia-election-facebook-ads-rage.html

Continuing our work to improve recommendations on YouTube. (2019). YouTube Creator Blog. [Blog Post]. Retrieved from https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html

Core values of librarianship. (n.d.) American Library Association. Retrieved from http://www.ala.org/advocacy/intfreedom/corevalues

Cormen, T. H. (2001). *Introduction to algorithms*. 2nd ed. Cambridge, Mass.: MIT Press.

Covington, P., Adams, J., & Sargin, E. (2016, September). Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (pp. 191-198). ACM.

Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., ... & Sampath, D. (2010, September). The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 293-296). ACM.

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, *3*(3), 398-415.

Dinet, J. (2014). *Information Retrieval in Digital Environments*. Hoboken: Wiley.

England, R. (2017). Facebook, Google and others add trust icons to tackle fake news. (2019, January 25). Retrieved from https://www.engadget.com/2017/11/16/facebook-google-and-others-add-trust-icons-to-tackle-fake-news

Eordogh, F. (2019). YouTube stops recommending conspiracy videos, finally. Forbes. Retrieved from https://www.forbes.com/sites/fruzsinaeordogh/2019/01/28/youtube -stops-recommending-conspiracy-videos-finally/#625ba26b77cf

FeldmanHall, O., Dalgleish, T., Evans, D., Navrady, L., Tedeschi, E., & Mobbs, D. (2016). Moral chivalry: Gender and harm sensitivity predict costly altruism. *Social psychological and personality science*, *7*(6), 542-551.

Fife, R., Chase, S., and Fine, S. (2019). PMO pressed Wilson-Raybould to abandon prosecution of SNC-Lavalin. *The Globe and Mail*. Retrieved from https://www.theglobeandmail.com /politics/article-pmo-pressed-justice-minister-to-abandon-prosecution-of-snc-lavalin/

Fingas, J. (2017). Google won't show news from sites that hide their country of origin. Engadget. Retrieved from https://www.engadget.com/2017/12/16/google-bans-news-

sites-which-hide-country-of-origin
Fiore, Q., & McLuhan, M. (1967). *The Medium is the message*. New York: Bantam.
Fischer, P. (2011). Selective exposure, decision uncertainty, and cognitive economy: A new theoretical perspective on confirmatory information search. *Social and Personality Psychology Compass*, 5(10), 751-762.
Fisher M., & Taub, A. (2019). How YouTube radicalized Brazil. *The New York Times*. Retrieved from https://www.nytimes.com/2019/08/11/world/americas/youtube-brazil.html
Fisher M., & Taub, A. (2019). On YouTube's digital playground, an open gate for pedophiles. *The New York Times*. Retrieved from https://www.nytimes.com/2019/06/03/world /americas/youtube-pedophiles.html
Fleder D., Hosanagar K. (2007). Recommender systems and their impact on sales diversity. In: Proceedings of the 8th ACM conference on Electronic commerce, p 199. ACM.
Full transcript: Sally Yates and James Clapper testify on Russian election interference. (2017). Washington Post. Retrieved from https://www.washingtonpost.com/news/post-politics /wp/2017/05/08/full-transcript-sally-yates-and-james-clapper-testify-on-russian-election-interference/
Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication*, 14(2), 265-285.
gaywonk. (2019, May 31). Since I started working at Vox, Steven Crowder has been making video after video "debunking" Strikethrough. Every single video has included repeated, overt attacks on my sexual orientation and ethnicity. Here's a sample: [Tweet]. Retrieved from https://twitter.com/gaywonk/status/1134264395717103617
gchaslot. (2019, July 14). My first op-ed in @WIRED: how the AI feedback loops I helped build at YouTube can amplify our worst inclinations, and what to do about it. [Tweet]. Retrieved from https://twitter.com/gchaslot/status/1150486810730143748
gchaslot. (2019, April 25). To be clear, my problem is not with Russians, who did an amazing job at 'optimizing YouTube' and generated millions in advertising revenue for Google. My problem is with algorithms that were designed with little consideration for bias or abuse [Tweet]. Retrieved from https://twitter.com/gchaslot/status/ 1122554807024140288
Ge, M., Delgado-Battenfeld, C., & Jannach, D. (2010). Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 257-260). ACM.
Gillespie, T. (2010). The politics of 'platforms'. *New Media & Society*, 12, 347–364.

Gilmore, R. (2018). Andrew Scheer rules out free trade deal with China. CTVNews. Retrieved from https://www.ctvnews.ca/politics/andrew-scheer-rules-out-free-trade-deal-with-china-1.4228729

Glaser, A. (2018). YouTube is adding fact-check links for videos on topics that inspire conspiracy theories. Slate Magazine. Retrieved from https://slate.com/technology/2018/08/youtube-is-adding-fact-check-links-from-wikipedia-and-encyclopedia-britannica-for-videos-on-topics-that-inspire-conspiracy-theories.html

Goffman, W. (1964). On relevance as a measure. *Information Storage and Retrieval*, *2*(3), 201-203.

Graham, Jefferson. (2005). Video websites pop up, invite postings. USA TODAY. Retrieved from https://usatoday30.usatoday.com/tech/news/techinnovations/2005-11-21-video-websites_x.htm

Greater transparency for users around news broadcasters. (2018). YouTube Official Blog. [Blog Post]. Retrieved from https://youtube.googleblog.com/2018/02/greater-transparency-for-users-around.html

Grenier, E. (2019). Canada poll tracker. *CBC*. Retrieved from https://newsinteractives.cbc.ca/elections/poll-tracker/canada/

Grossman, D. A, & Frieder, O. (2004). *Information retrieval : algorithms and heuristics*. Second edition. Dordrecht: Springer Netherlands.

Gruzd, A., Jacobson, J., Mai, P., & Dubois, E. (2017). The state of social media in Canada 2017. Ryerson University Social Media Lab. Retrieved from https://doi.org/10.5683/SP/AL8Z6R

Guignard, J. (2019). Public safety minister Ralph Goodale discusses national security priorities. The Globe and Mail. Retrieved from https://globalnews.ca/news/4851043/public-Safety-minister-goodale-discusses-national-security-priorities

Habermas, J. (2006). Religion in the public sphere. European journal of philosophy, 14(1), 1-25.

Hagemann, R. (2019). A Precision regulation approach to stopping illegal activities online. IBM. Retrieved from https://www.ibm.com/blogs/policy/cda-230

Haider, J. and Åström, F. (2016). Dimensions of trust in scholarly communication: problematizing peer review in the aftermath of John Bohannon's 'Sting' in science. *Journal of the Association for Information Science and Technology*, Vol. 68 No. 2, pp. 450-467.

Harassment and cyberbullying policy (n.d.). Retrieved June 10, 2019, from YouTube Help website: https://support.google.com/youtube/answer/2802268?visit_id=

1-636215053151010017-1930197662&rd=1&hl=en

Harris, S., & Ross, J. (2006). *Beginning algorithms*. Indianapolis, IN: Wiley.

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychological Bulletin*, 135(4), 555.

Harvey, J. H. (1980). *Cognition, social behavior, and the environment*. Hillsdale, N.J: L. Erlbaum.

Haskins, C. (2018). YouTube lets California fire conspiracy theories run wild. Vice. Retrieved from https://motherboard.vice.com/en_us/article/43937d/youtube-lets-california-fire-conspiracy-theories-run-wild

Heath, R. L. (2013). Reinforcement theory. In *Encyclopedia of public relations*. Los Angeles: SAGE Reference.

Helberger, N., Karppinen, K., & D'Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, *21*(2), 191-207.

Herrman, J. (2018). The Making of a No. 1 YouTube Conspiracy Video After the Parkland Tragedy. N. Y. Times. Retrieved from https://www.nytimes.com/2018/02/21/business /media/youtube-conspiracy-video-parkland.html

Hoffmann, G. (2005). Rhetoric of Bush Speeches: Purr words and snarl words. et Cetera, 62(2), 198.

Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J., & François, C. (2018). *The IRA, social media and political polarization in the United States, 2012-2018*. University of Oxford. Retrieved from https://media1.s-nbcnews.com/i/today/z_creative/IRAReport17Dec.pdf

Hutchins, A. (2017). Andrew Scheer's path to leadership of the Conservative Party. *Macleans.ca*. Retrieved from https://www.macleans.ca/politics/ottawa/andrew -scheers-path-to-leadership-of-the-conservative-party

Ideologies. (n.d.) Southern Poverty Law Center. Retrieved from https://www.splcenter.org /fighting-hate/extremist-files/ideology

Imminent federal election to be costliest, longest in recent Canadian history. (2015). *Toronto Sun*. Retrieved from https://torontosun.com/2015/07/29/imminent-federal-election-to-be-costliest-longest-in-recent-canadian-history/wcm/8d4e7c83-bcad-42ad-a36a-84d75 5920026

Jones, A. (2017). Ontario politician Jagmeet Singh launches bid for federal NDP leadership. Globe and Mail. Retrieved from https://beta.theglobeandmail.com/news/politics/ ontario-politician-jagmeet-singh-launches-bid-for-federal-ndp-leadership/article350015

59
Kaiser, J., Rauchfleisch, A. (2018). Unite the right? How YouTube's recommendation algorithm connects the U.S. far-right. Medium. Retrieved from https://medium.com/@MediaManipulation/unite-the-right-how-youtubes-recommendation-algorithm-connects-the-u-s-far-right-9f1387ccfabd
Kaminskas, M., & Bridge, D. (2016). Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-Accuracy objectives in recommender systems. *ACM Transactions On Interactive Intelligent Systems*, 7(1).
Kassam, A., & Lartey, J. (2017). Québec City mosque shooting: six dead as Trudeau condemns 'terrorist attack'. The Guardian. Retrieved from https://www.theguardian.com/world/2017/jan/30/quebec-mosque-shooting-canada-deaths
Kahneman, D. (1994). New challenges to the rationality assumption. *Journal of Institutional and Theoretical Economics*, 150 (1), 18–36.
Kaye, B. K., & Sapolsky, B. S. (1997). Electronic monitoring of in-home television RCD usage. *Journal of Broadcasting & Electronic Media*, 41 (2), 214–228.
Kembellec, G., Chartron, G., & Saleh, I. (2015). *Recommender Systems*. John Wiley & Sons.
Khosrow-Pour, M. (2013). *Dictionary of information science and technology*. 2nd ed. Hershey, PA: Information Science Reference.
Kilpatrick, S. (2014). Justin Trudeau says anti-abortion candidates can't run as Liberals. *The National Post*. Retrieved from https://nationalpost.com/news/politics/justin-trudeau-says-anti-abortion-candidates-cant-run-as-liberals.
Klapper, J.T. (1960). *The effects of mass communication*. New York: Free Press.
Knobloch-Westerwick, S. (2014). *Choice and preference in media use: Advances in selective exposure theory and research*. Routledge.
Knobloch-Westerwick, S., & Kleinman, S. B. (2012). Preelection selective exposure: Confirmation bias versus informational utility. *Communication Research*, 39(2), 170-193.
Kraemer, F., Van Overveld, K., & Peterson, M. (2011). Is there an ethics of algorithms? *Ethics and Information Technology*, 13(3), 251-260.
Kyncl, R. & Mohan, N. (2018). Additional Changes to the YouTube Partner Program (YPP) to Better Protect Creators. YouTube Creator Blog. [Blog Post]. Retrieved from https://youtube-creators.googleblog.com/2018/01/additional-changes-to-youtube-partner.html
Kyncl, R. & Mohan, N. (2018). Building a better news experience on YouTube, together.

YouTube Creator Blog. [Blog Post]. Retrieved from https://youtube.googleblog.com /2018/07/building-better-news-experience-on.html

La Monica, Paul. (2006). Google to buy YouTube for $1.65 billion. CNN Money. Retrieved from https://money.cnn.com/2006/10/09/technology/googleyoutube_deal/index.htm?cnn=yes

Lam, X. N., Vu, T., Le, T. D., & Duong, A. D. (2008). Addressing cold-start problem in recommendation systems. *In Proceedings of the 2nd international conference on Ubiquitous information management and communication* (pp. 208-211). ACM.

Latest changes to video page: new playlist experience, integrated comments & more. (2010). Retrieved from https://youtube.googleblog.com/2010/02/latest-changes-to-video-page-new.html

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, 31(4), 611-627.

Lewis, R. (2018). Alternative influence: broadcasting the reactionary right on YouTube. Data & Society. Retrieved from https://datasociety.net/wp-content/uploads/2018/09/DS_Alternative_Influence.pdf

Lewis, P. (2018). 'Fiction is outperforming reality': how YouTube's algorithm distorts truth. The Guardian. Retrieved from https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth

Lewis, P., & McCormick, E. (2018). How an ex-YouTube insider investigated its secret algorithm. The Guardian. Retrieved from https://www.theguardian.com/technology/2018/feb/02/youtube-algorithm-election-clinton-trump-guillaume-chaslot

Limited features for certain videos. (n.d.). YouTube Help. Retrieved from https://support.google.com/youtube/answer/7458465?hl=en-GB

Lohr, S. (2012, February 29). For impatient web users, an eye blink is just too long to wait. New York Times, p. A1.

Lovett, I & Nicas, J. (2017). PragerU sues YouTube in free-speech case. WSJ. Retrieved from https://www.wsj.com/articles/prageru-sues-youtube-in-free-speech-case-1508811856

Lustig, C. and Nardi, B. (2015). Algorithmic authority: the case of Bitcoin. *Proceedings of the Annual Hawaii International Conference on System Sciences*, March, pp. 743-752.

Lyons, M. N. (2017). Ctrl-alt-delete: The origins and ideology of the alternative right. *Political Research Associates*. Retrieved from https://www.politicalresearch.org/wp-content/uploads/2017/01/Lyons_CtrlAltDelete_PRINT.pdf

Maack, M. M. (2019). 'YouTube recommendations are toxic,' says dev who worked on the algorithm. Next Web. Retrieved from https://thenextweb.com/google/2019/06/14/youtube-recommendations-toxic-algorithm-google-ai

MacKenzie, D. A., & Wajcman, J. (2011). *The social shaping of technology*. Maidenhead: Open University Press.

MacManus, Richard. (2009). Top 5 web trends of 2009 - personalization. NYTimes.com. Retrieved from https://archive.nytimes.com/www.nytimes.com/external/readwriteweb/2009/09/09/09readwriteweb-top-5-web-trends-of-2009-personalization-67911.html

Mager, A. (2012). Algorithmic ideology: How capitalist society shapes search engines. *Information, Communication & Society*, *15*(5), 769-787.

Making our strikes system clear and consistent. (2019). YouTube Creator Blog. [Blog Post]. Retrieved from https://youtube-creators.googleblog.com/2019/02/making-our-strikes-system-clear-and.html

Marr, B. (2018). How much data do we create every day? The Mind-blowing stats everyone should read. Forbes. Retrieved from https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#3ca5e1ec60ba

McCarthy, D. R. (2011). Open networks and the open door: American foreign policy and the narration of the internet. *Foreign Policy Analysis*, 7(1).

McCracken, Harry (2010). YouTube Leanback: YouTube that looks like TV. PC World. Retrieved from https://www.pcworld.com/article/200704/youtube_leanback_tv.html

McIntyre, L. C. (2018). *Post-truth*. Cambridge: MIT Press.

McKelvey, F., Frizzera, L. (2019) The Political Narrative of YouTube Recommendation System: 2018 Ontario election. Canadian Communication Association (CCA). Vancouver, Canada. Retrieved from https://luciano.fluxo.art.br/the-political-narrative-of-youtube-recommendation-system-2018-ontario-election/

McKenna, R. (2019). Far-right, neo-Nazi, white supremacist groups an increasing concern, threat to Canadians: Goodale. Globe and Mail. Retrieved from https://www.theglobeandmail.com/politics/article-far-right-neo-nazi-white-supremacist-groups-an-increasing-concern

McLaughlin, K., McMahon, M., McMahon, K., Sobelman, D., Donaldson, C., Primitive Features Inc., & National Film Board of Canada. (2003). *McLuhan's wake*. Montreal, Quebec: Primitive Entertainment/National Film Board of Canada.

McLuhan, M., & Powers, B. R. (1989). *The global village: Transformations in world life and*

*media in the 21st century*. New York: Oxford University Press.
McLuhan, M. (1967). The invisible environment: The future of an erosion. *Perspecta*, 163-167.
McLuhan, M. (1968). *Understanding media: The extensions of man*. London: Sphere Books.
Menegus, B. (2019). YouTube is going to bury 'borderline' content. It won't tell us what that means. Gizmodo. Retrieved from https://gizmodo.com/youtube-is-going-to-bury-borderline-content-it-wont-te-1832162383
Merrill, J. C. (2011). Theoretical foundations for media ethics. In *Controversies in media ethics*, 3rd ed., edited by A. D. Gordon, J. M. Kittross, J. C. C. Merrill, W. Babcock, and M. Dorsher, 3-32. New York: Routledge.
Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication research*, 41(8), 1042-1063.
Mission, priority areas, goals. (n.d.) American Library Association. Retrieved from http://www.ala.org/aboutala/governance/policymanual/updatedpolicymanual/section1/1mission
Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2053951716679679.
Most recommended verbs on YouTube. (n.d.). Algotransparency.org. Retrieved from https://algotransparency.org/verbs.html
Munger, K., & Phillips, J. (2019). A Supply and Demand Framework for YouTube Politics. *OSF*. Retrieved from https://osf.io/73jys/
Murthy, D., Powell, A. B., Tinati, R., Anstead, N., Carr, L., Halford, S. J., & Weal, M. (2016). Automation, algorithms, and politics| Bots and political influence: A sociotechnical investigation of social network capital. *International Journal of Communication*, 10, 20.
National cyber threat assessment 2018. (2018). Canadian Centre for Cyber Security. Retrieved from https://www.cyber.gc.ca/sites/default/files/publications/national-cyber-threat-assessment-2018-e_1.pdf
Neidig, H. (2018). Judge dismisses lawsuit alleging Google censorship of conservative YouTube videos. TheHill. Retrieved from https://thehill.com/policy/technology/380455-judge-dismisses-lawsuit-alleging-google-censorship-of-conservative-youtube
Newton, C. (2017). How YouTube perfected the feed. The Verge. Retrieved from https://www.theverge.com/2017/8/30/16222850/youtube-google-brain-algorithm-video-recommendation-Personalized-feed
Noble, S. U. (2018). *Algorithms of oppression: Data discrimination in the age of Google*. New

York: New York University Press.

O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2014). Down the (white) rabbit hole: The extreme right and online recommender systems. *Social Science Computer Review*, 33(4), 459-478.

OCSE. (1998). 21st century technologies: Promises and perils of a dynamic future. Paris: Organisation for economic co-operation and development.

O'Donovan, C. (2019). YouTube just demonetized anti-vax channels. BuzzFeed News. Retrieved from https://www.buzzfeednews.com/article/carolineodonovan/youtube-just-demonetized-anti-vax-channels

O'Hara, K., & Stevens, D. (2015). Echo chambers and online radicalism: Assessing the Internet's complicity in violent extremism. *Policy & Internet*, 7(4), 401-422.

O'Neil, C. (2017). The era of blind faith in big data must end. [Video file]. Retrieved from https://www.ted.com/talks/cathy_o_neil_the_era_of_blind_faith_in_big_data_must_end?language=en

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. London: Penguin Books.

Orr, C. (2019). Twitter bots boosted the trending #TrudeauMustGo hashtag. *The National Observer*. Retrieved from https://www.nationalobserver.com/2019/07/18/news/twitter-bots-boosted-trending-trudeaumustgo-hashtag

Papadamou, K., Papasavva, A., Zannettou, S., Blackburn, J., Kourtellis, N., Leontiadis, I., ... & Sirivianos, M. (2019). Disturbed YouTube for kids: Characterizing and detecting disturbing content on youtube. Retrieved from https://pdfs.semanticscholar.org/4820/2c77e4eae1dcf48dfe9c5c45d13026c03417.pdf

Parliament of Canada. (n.d.) Current party standings. House of Commons. Retrieved from https://www.ourcommons.ca/parliamentarians/en/partystandings

Pariser, E. (2011). Beware online "filter bubbles". [Video file]. Retrieved from https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles

Pariser, E. (2012). *The filter bubble: What the Internet is hiding from you*. London: Penguin Books.

Paul, K. (2019). YouTube algorithm adds 9/11 explainer to Notre Dame fire video. The Guardian. Retrieved from https://www.theguardian.com/world/2019/apr/15/notre-dame-fire-youtube-panels-show-9-11-attacks

Perkel, Colin. (2018). Trudeau says 2019 federal election likely to be the nastiest one yet. The Star. Retrieved from https://www.thestar.com/news/canada/2018/10/02/trudeau-says

-2019-election-likely-to-be-the-nasiest-one-yet.html

Perloff, R. M. (2013). Political persuasion. In James Price Dillard and Lijiang Shen (Eds.), *The SAGE handbook of persuasion: Developments in theory and practice*. Retrieved from the Gale Virtual Reference Library database.

Perry, B., & Scrivens, R. (2015). Right-wing extremism in Canada: An environmental scan. *Ottawa, ON: Public Safety Canada*. Retrieved from https://www.publicsafety.gc.ca/cnt/ntnl-scrt/cntr-trrrsm/r-nd-flght-182/knshk/ctlg/dtls-en.aspx?i=116

Policies and Safety (n.d.). YouTube. Retrieved from https://www.youtube.com/yt/about/policies/#community-guidelines

Recode. (2019, June 10). "Are you really sorry for anything that happened to the LGBTQ community? Or are you just sorry they were offended?" Watch @SusanWojcicki's response at #CodeCon: [Tweet]. Retrieved from https://twitter.com/Recode/status/1138231255445598208

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, *40*(3), 56-58.

Resolution on dis information, media manipulation & the destruction of public information. (2005). American Library Association. Retrieved from http://www.ala.org/aboutala/sites/ala.org.aboutala/files/content/governance/policymanual/updatedpolicymanual/ocrpdfofprm/52-8disinformation.pdf

Restricted by YouTube. (2018). PragerU. Retrieved from https://www.prageru.com/playlists/restricted-youtube

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A., & Meira, W. (2019). Auditing Radicalization Pathways on YouTube. *arXiv preprint arXiv:1908.08313*.

Ricci, F., Rokach, L., Shapira, B. (2015). Introduction to recommender systems handbook. In: Ricci F., Rokach L., Shapira B. (eds) *Recommender systems handbook*. Springer, Boston, MA.

Ricci, F., Rokach, L., Shapira B. (2015). Recommender systems: introduction and challenges. In: Ricci F., Rokach L., Shapira B. (eds) *Recommender systems handbook*. Springer, Boston, MA.

Riga, A. (2018). I didn't incite mosque shooter, conservative pundit Ben Shapiro insists. *Montreal Gazette*. Retrieved from https://montrealgazette.com/news/quebec/i-didnt-incite-mosque-shooter-conservative-pundit-ben-shapiro-insists

Roberts, J.J. (2016). A Top google result for the holocaust is now a white supremacist site. Fortune. Retrieved from http://fortune.com/2016/12/12/google-holocaust

Roose, K. (2019). The Making of a YouTube radical. *N. Y. Times*. Retrieved from https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html

Saul, H. (2015). Justin Trudeau: The rise of the feminist and pro-choice Canadian Prime Minister who wants to legalise marijuana 'right away'. *Independent*. Retrieved from https://www.independent.co.uk/news/people/justin-trudeau-the-self-declared-feminist-and-pro-choice-prime-minister-of-canada-who-wants-to-a6700976.html

Schindler, P. (2018). The Google News Initiative: Building a stronger future for news [Blog Post]. Google. Retrieved from https://www.blog.google/outreach-initiatives/google-news-initiative/announcing-google-news-initiative

Schmitt, J. B., Rieger, D., Rutkowski, O., & Ernst, J. (2018). Counter-messages as Prevention or Promotion of Extremism?! The Potential Role of YouTube: Recommendation Algorithms. Journal of Communication, 68(4), 780-808.

Search and discovery on YouTube (n.d.) YouTube. Retrieved from https://creatoracademy.youtube.com/page/lesson/discovery#strategies-zippy-link-8

Security and disinformation in the U.S. 2016 election: What we found. (2017). YouTube. Retrieved from https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/google_US2016election_findings_1_zm64A1G.pdf

Shi, Y., Zhao, X., Wang, J., Larson, M., and Hanjalic, A. (2012). Adaptive diversification of recommendation results via latent factor portfolio. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 175–184.

Shirky, C. (2009). A speculative post on the idea of algorithmic authority. [Blog post]. Retrieved From www.shirky.com/weblog/2009/11/a-speculative-post-on-the-idea-of-algorithmic-authority/

Smith, A. (2019). Does YouTube's Algorithm Lead to Radicalization? Retrieved from https://uk.pcmag.com/youtube/123362/does-youtubes-algorithm-lead-to-radicalization

Smith, A., Toor, S., & van Kessel, P. (2018). Many turn to YouTube for children’s content, news, how-to lessons. Pew Research Center: Internet, Science & Tech. Retrieved from http://www.pewinternet.org/2018/11/07/many-turn-to-youtube-for-childrens-content-news-how-to-lessons

Smith, J. (2017). Trudeau abandons promise for electoral reform. Macleans.ca. Retrieved from https://www.macleans.ca/politics/ottawa/trudeau-abandons-promise-for-electoral-reform

Sohail, S. S., Siddiqui, J., & Ali, R. (2017). Classifications of recommender systems: A review.

*Journal Of Engineering Science & Technology Review*, 10(4), 132-153.
Solsman, J. E. (2018). YouTube's AI is the puppet master over most of what you watch. CNET. Retrieved from https://www.cnet.com/news/youtube-ces-2018-neal-mohan
Stubbs, J. & Gibson, G. (2019, March 02). Russia's RT America registers as 'foreign agent' in U.S. Retrieved from https://www.reuters.com/article/us-russia-usa-media-restrictions-rt/russias-rt-america-registers-as-foreign-agent-in-u-s-idUSKBN1DD25B
Sullivan, D. (2017). Google's 'Project Owl' - a three-pronged attack on fake news & problematic content. Search Engine Land. Retrieved from https://searchengineland.com/googles-project-owl-attack-fake-news-273700
Sullivan, D. (2010). Schmidt: Listing Google's 200 ranking factors would reveal business secrets. Search Engine Land. Retrieved from https://searchengineland.com/schmidt-listing-googles-200-ranking-factors-would-reveal-business-secrets-51065
Sullivan, M. C. (2018). Why librarians can't fight fake news. Journal of librarianship and information Science, 0961000618764258.
Summers, N. (2017). Google will flag fake news stories in search results. Engadget. Retrieved from https://www.engadget.com/2017/04/07/google-fake-news-fact-check-search-results
Sundin, O., Haider, J., Andersson, C., Carlsson, H., & Kjellberg, S. (2017). The search-ification of everyday life and the mundane-ification of search. *Journal of Documentation*, 73(2), 224-243.
Sunstein, C. R. (2004). Democracy AND FILTERING. *Communications Of The ACM*, 47(12), 57-59.
SusanWojcicki. (2019, May 2). My #1 priority is responsibility, even if that comes at the expenses of growth. [Tweet]. Retrieved from https://twitter.com/susanwojcicki/status/1124021310466736129
Sutton R.S. (1992) Introduction: The challenge of reinforcement learning. In: Sutton R.S. (eds) *Reinforcement Learning. The Springer International Series in Engineering and Computer Science (Knowledge Representation, Learning and Expert Systems)*, vol 173. Springer, Boston, MA
Swartz, D. (2015). Federal election voter turnout 68.3 per cent, highest in 22 years: official vote count. *CBC News*. Retrieved from https://www.cbc.ca/news/politics/multimedia/federal-election-voter-turnout-68-3-per-cent-highest-in-22-years-official-vote-count-1.3302064
TeamYouTube. (2019, June 4). (2/4) Our teams spent the last few days conducting an

in-depth review of the videos flagged to us, and while we found language that was clearly hurtful, the videos as posted don't violate our policies. We've included more info below to explain this decision: [Tweet]. Retrieved from https://twitter.com/TeamYouTube/status/1136055351885815808

TeamYouTube. (2019, June 5). To clarify, in order to reinstate monetization on this channel, he will need to remove the link to his T-shirts. [Tweet]. Retrieved from https://twitter.com/TeamYouTube/status/1136356046887313408

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

theJagmeetSingh. (2018, January 6). Trudeau has said a $15 federal minimum wage is off the table - I believe it's the minimum that workers deserve. Canadians deserve better - an NDP government would implement a federal minimum wage of $15 because nobody should be working and living under the poverty line. [Tweet]. Retrieved from https://twitter.com/theJagmeetSingh/status/949739233664679936

Tobias, M. (2018). Why Alex Jones was banned from Apple, Facebook, YouTube. Retrieved from https://www.politifact.com/truth-o-meter/article/2018/aug/07/why-infowars-alex-jones-was-banned-apple-facebook-

Trending on YouTube (n.d.). YouTube Help. Retrieved from https://support.google.com/youtube/answer/7239739?hl=en

Trudeau formally announces he'll run again in next year's election. (2018). *CBC News*. Retrieved from https://www.cbc.ca/news/canada/montreal/trudeau-run-Papineau-1.4791248

Tufekci, Z. (2018). YouTube, the great radicalizer. N. Y. Times. Retrieved from https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html

Tunkelang, D. (2009). Faceted search. *Synthesis lectures on information concepts, retrieval, and services*, 1(1), 1-80.

Turtle, H. (1994). Natural language vs. Boolean query evaluation: a comparison of retrieval performance. In *SIGIR'94* (pp. 212-220). Springer, London.

Vargas, S., and Castells, P. (2014). Improving sales diversity by recommending users to items. In *Proceedings of the 8th ACM Conference on Recommender Systems*. 145–152.

Volz, D. (2017). Google uncovered Russia-backed ads on YouTube, Gmail : source. Retrieved from https://www.reuters.com/article/us-usa-trump-russia-alphabet/google-uncovered-russia-backed-ads-on-youtube-gmail-source-idUSKBN1CE192

Wagner, B. (2016). Algorithmic regulation and the global default: Shifting norms in Internet technology. *Etikk i praksis-Nordic Journal of Applied Ethics*, (1), 5-13.

Wakabayashi, D., & Confessore, N. (2017). Russia's favored outlet is an online news giant. YouTube helped. N. Y. Times. Retrieved from https://www.nytimes.com/2017/10/23/ technology/youtube-russia-rt.html

Walker, K. (2018). An update on state-sponsored activity. Google. Retrieved from https://www.blog.google/technology/safety-security/update-state-sponsored-activity

Wherry, A. (2018). Doug Ford is a populist, but it's not yet clear what kind. *CBC News*. Retrieved from https://www.cbc.ca/news/politics/doug-ford-trump-populist-analysis -wherry-1.4663465

Wilson, J. (2018). A 'political hit job'? Why the alt-right is accusing big tech of censorship. the Guardian. Retrieved from https://www.theguardian.com/the-chain/2018/mar/04/ alt-right-big-tech-censorship-lawsuits

Wilson, T. D., & Gilbert, D. T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science*, 14 (3), 131–134.

Winkler, R., Nicas, J., & Fritz, B. (2017). Disney Severs Ties With YouTube Star PewDiePie After Anti-Semitic Posts. WSJ. Retrieved from https://www.wsj.com/articles/disney -severs-ties-with-youtube-star-pewdiepie-after-anti-semitic-posts-1487034533

Winsa, P. (2017). He says freedom, they say hate. The pronoun fight is back. *The Star*. Retrieved from https://www.thestar.com/news/insight/2017/01/15/he-says-freedom -they-say-hate-the-pronoun-fight-is-back.html

Wojcicki, S. (2018). My five priorities for creators in 2018 [Blog Post]. Retrieved from https://youtube-creators.googleblog.com/2018/02/my-five-priorities-for-creators-in-20 18_1.html

Wojcicki, S. (2019). YouTube in 2019: Looking back and moving forward [Blog Post]. Retrieved from https://youtube-creators.googleblog.com/2019/02/youtube-in-2019-looking- back-and-moving.html

Wong, J. C., & Levin, S. (2019). YouTube vows to recommend fewer conspiracy theory videos. the Guardian. Retrieved from https://www.theguardian.com/technology/2019/jan/25 /youtube-conspiracy-theory-videos-recommendations

Woo, Andrea. (2017). Jagmeet Singh vows to decriminalize petty drug charges at NDP debate. The Globe and Mail. Retrieved from https://beta.theglobeandmail.com/news/politics/ jagmeet-singh-vows-to-decriminalize-petty-drug-charges-at-ndp-debate/article3623441 5

Woodie, A. (2014). Inside Sibyl, Google's Massively Parallel Machine Learning Platform.

Retrieved from https://www.datanami.com/2014/07/17/inside-sibyl-googles-massively-parallel-machine-learning-platform

YouTube. (n.d.). About YouTube. YouTube. Retrieved from https://www.youtube.com/yt/about

YouTube Community Guidelines enforcement (n.d.). Retrieved June 12, 2019, from YouTube Transparency Report website: https://transparencyreport.google.com/youtube-policy/removals

YouTube Creators [Creator Academy]. (2017, August 28). '*The Algorithm' - How YouTube search & discovery works* [Video File]. YouTube. Retrieved from https://www.youtube.com/watch?v=hPxnIix5ExI

YouTube Creators [Creator Academy]. (2017, August 30). *How YouTube's Suggested Videos Work* [Video File]. YouTube. Retrieved from https://www.youtube.com/watch?v=E6pC6iql5xM

YouTube elevates most popular users to partners. (2007). Official YouTube Blog. [Blog Post]. Retrieved from https://youtube.googleblog.com/2007/05/youtube-elevates-most-popular-users-to.html

YouTube for press. (n.d.). YouTube. Retrieved from https://www.youtube.com/yt/about/press/

YouTube now: why we focus on watch time. (2012). YouTube Creator Blog. [Blog Post]. Retrieved from https://youtube-creators.googleblog.com/2012/08/youtube-now-why-we-focus-on-watch-time.html

YouTube search, now optimized for time watched. (2012). YouTube Creator Blog. [Blog Post]. Retrieved from https://youtube-creators.googleblog.com/2012/10/youtube-search-now-optimized-for-time.html

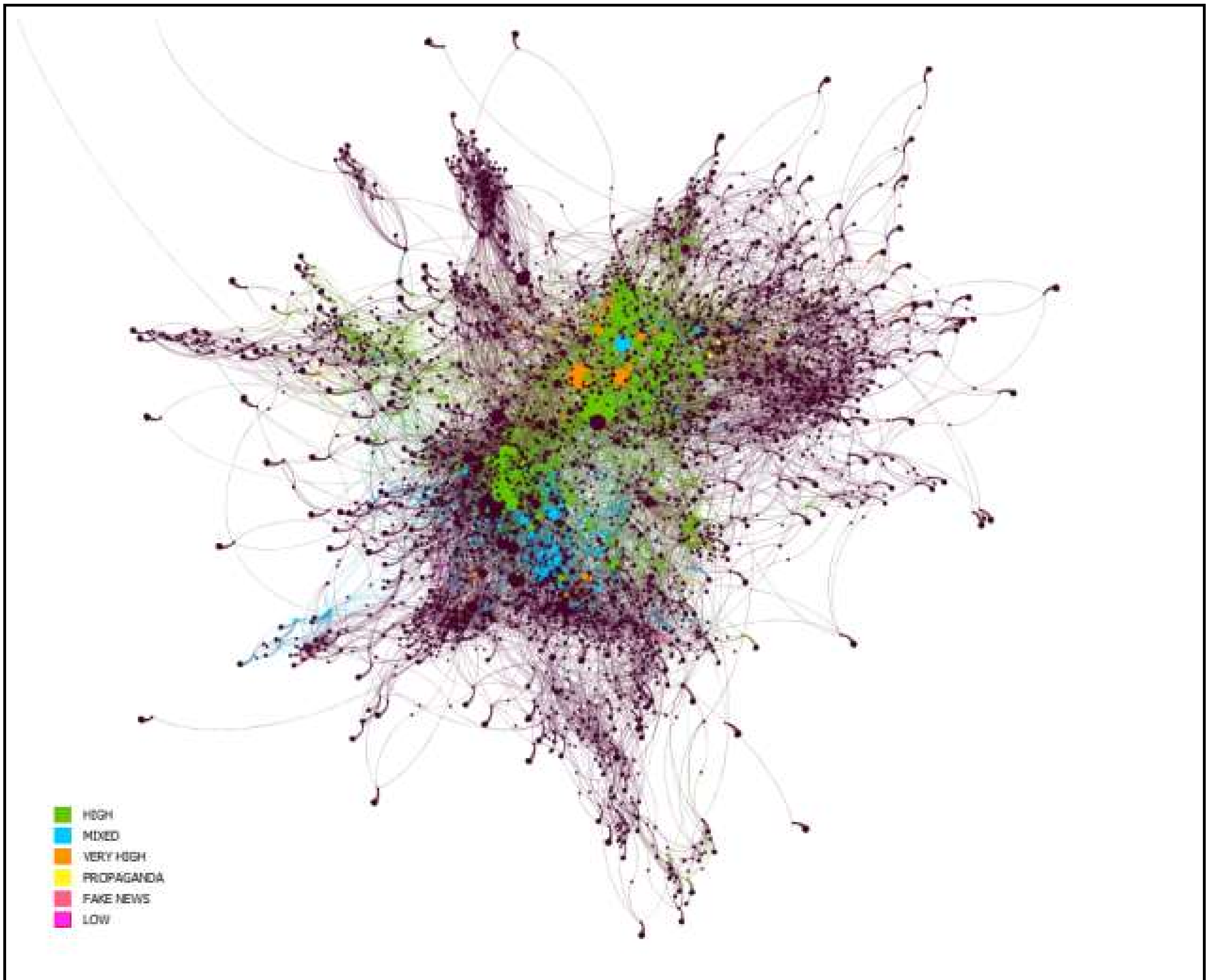
Zhou, R., Khemmarat, S., Gao, L., Wan, J., & Zhang, J. (2016). How YouTube videos are discovered and its impact on video views. *Multimedia Tools and Applications*, 75(10), 6035-6058.

Zhou, R., Khemmarat, S., & Gao, L. (2010). The impact of YouTube recommendation system on video views. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement* (pp. 404-410). ACM.

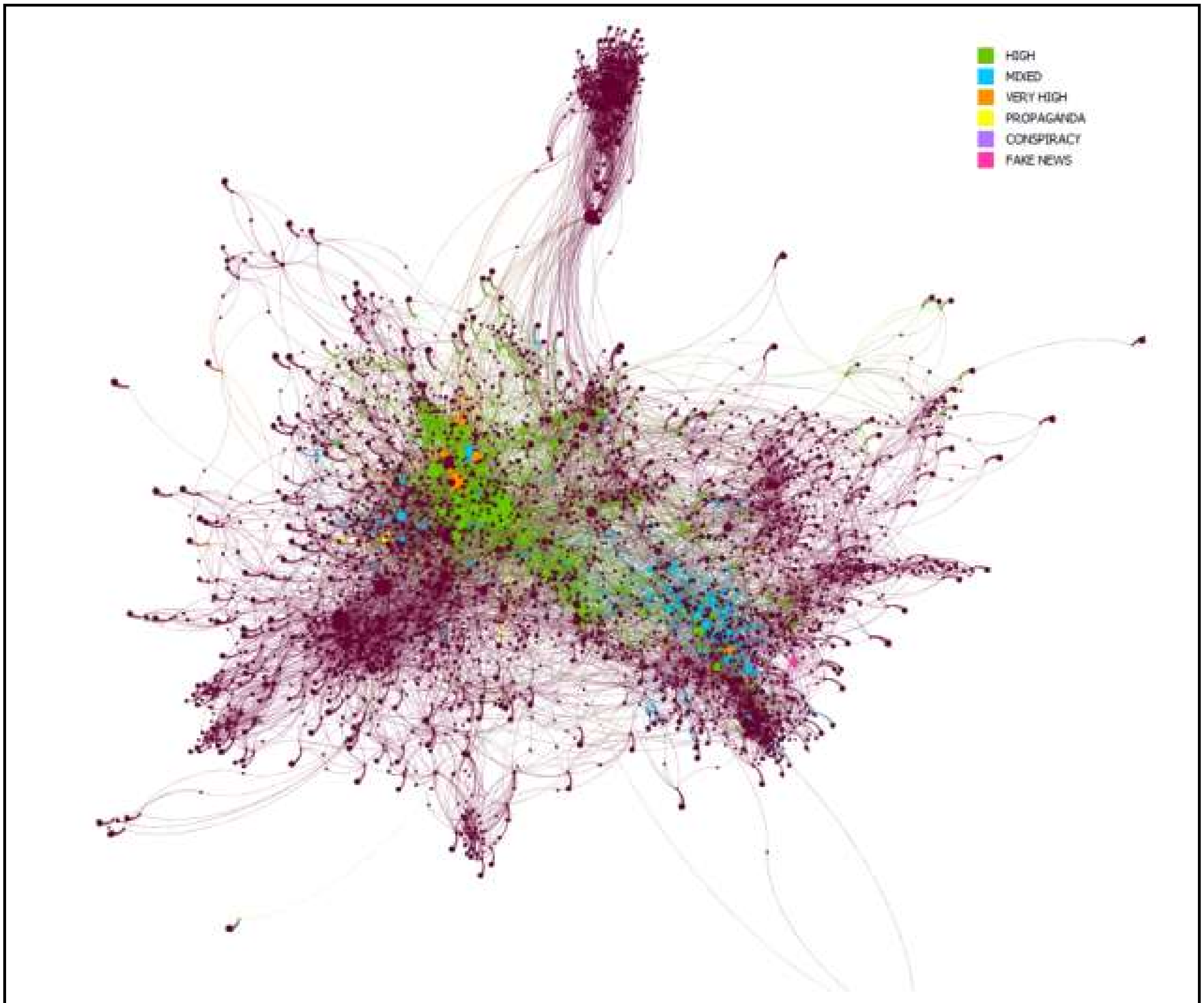
Ziegler, C., McNee, S., Konstan, J., and Lausen, G. (2005). Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on the World Wide Web*. 22–32.

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist — it's time to make it fair. *Nature,* 559(7714), 324.

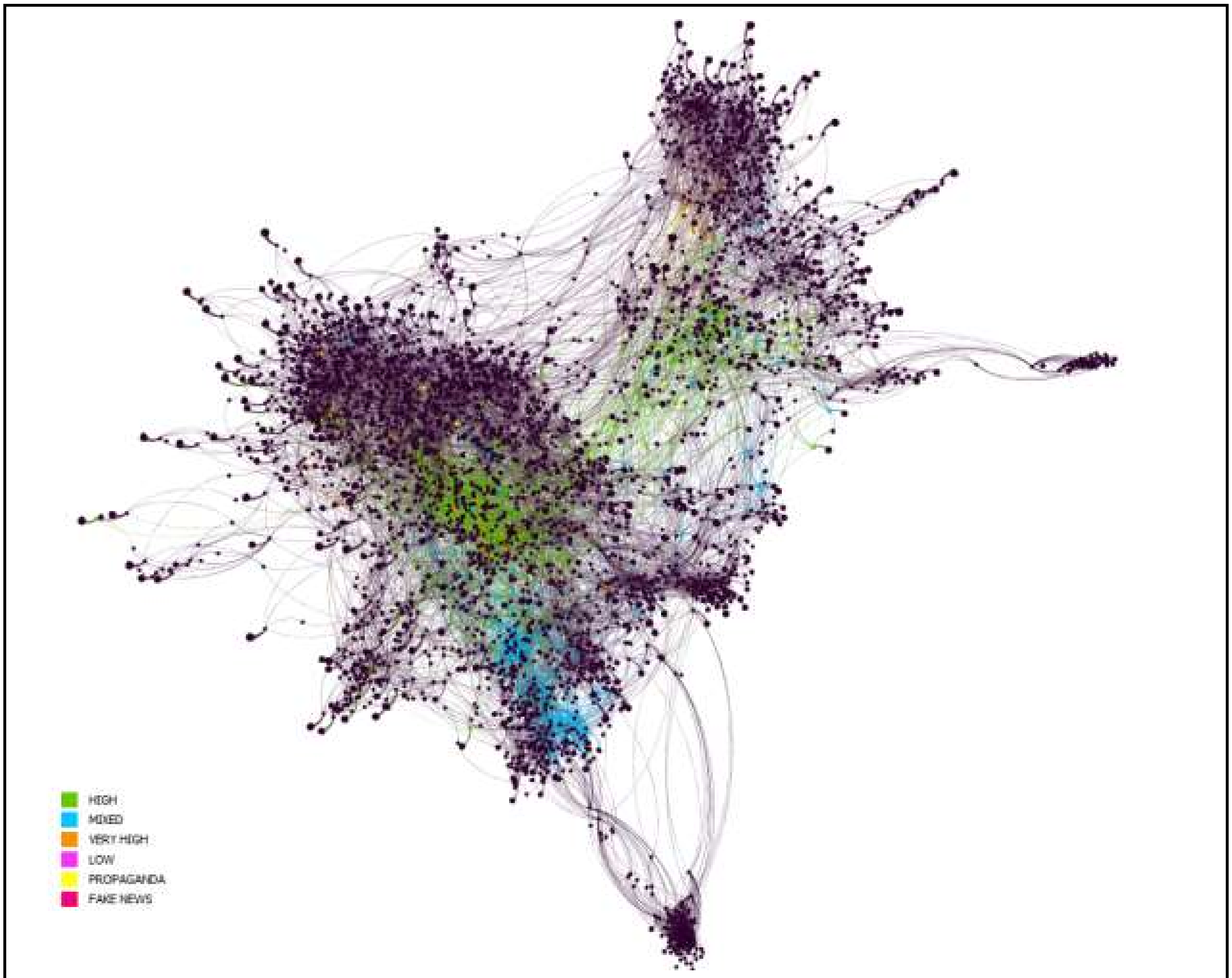
APPENDIX A: NETWORK VISUALIZATIONS



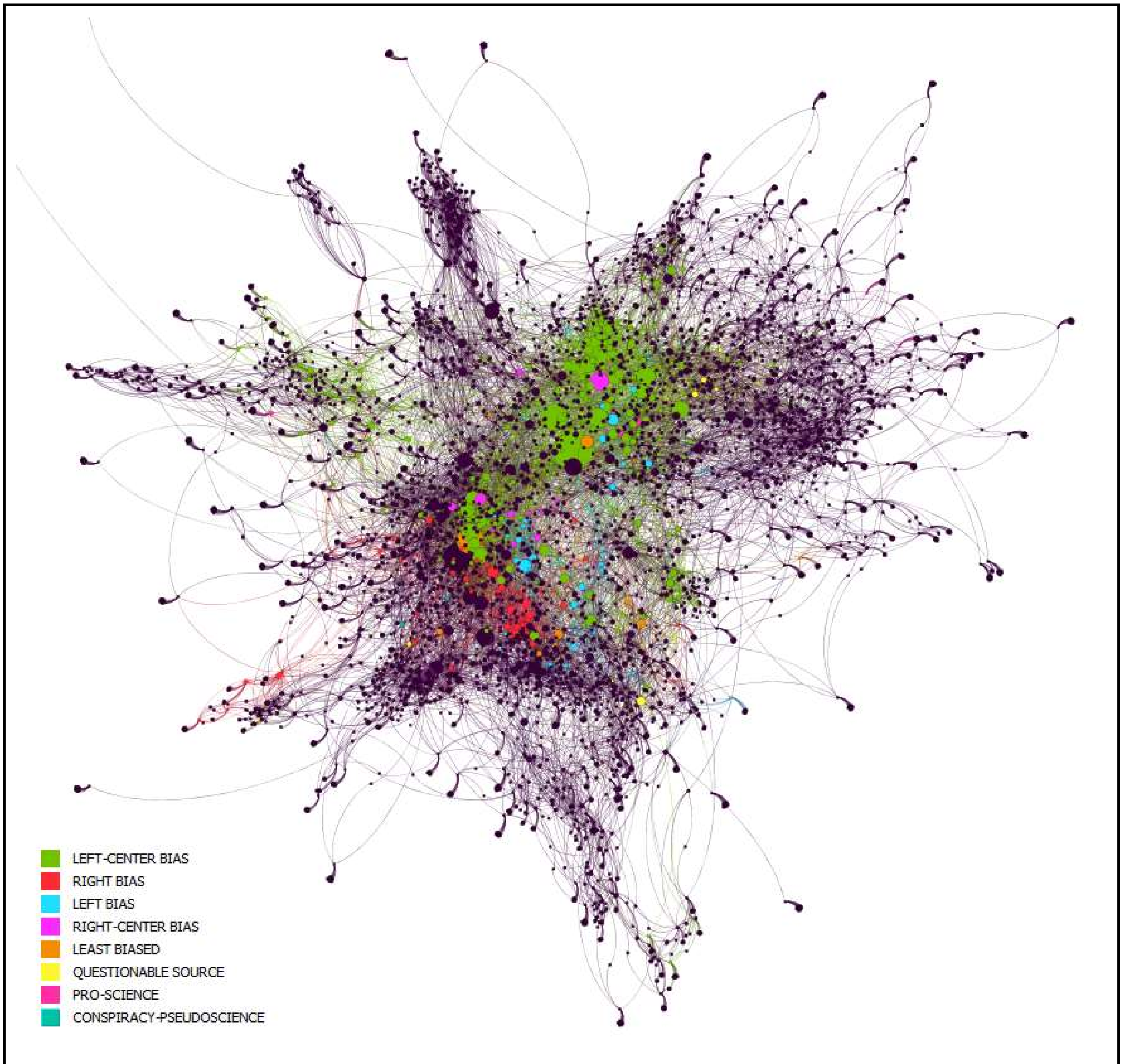

*Factual Accuracy of Sources in Scheer Dataset*



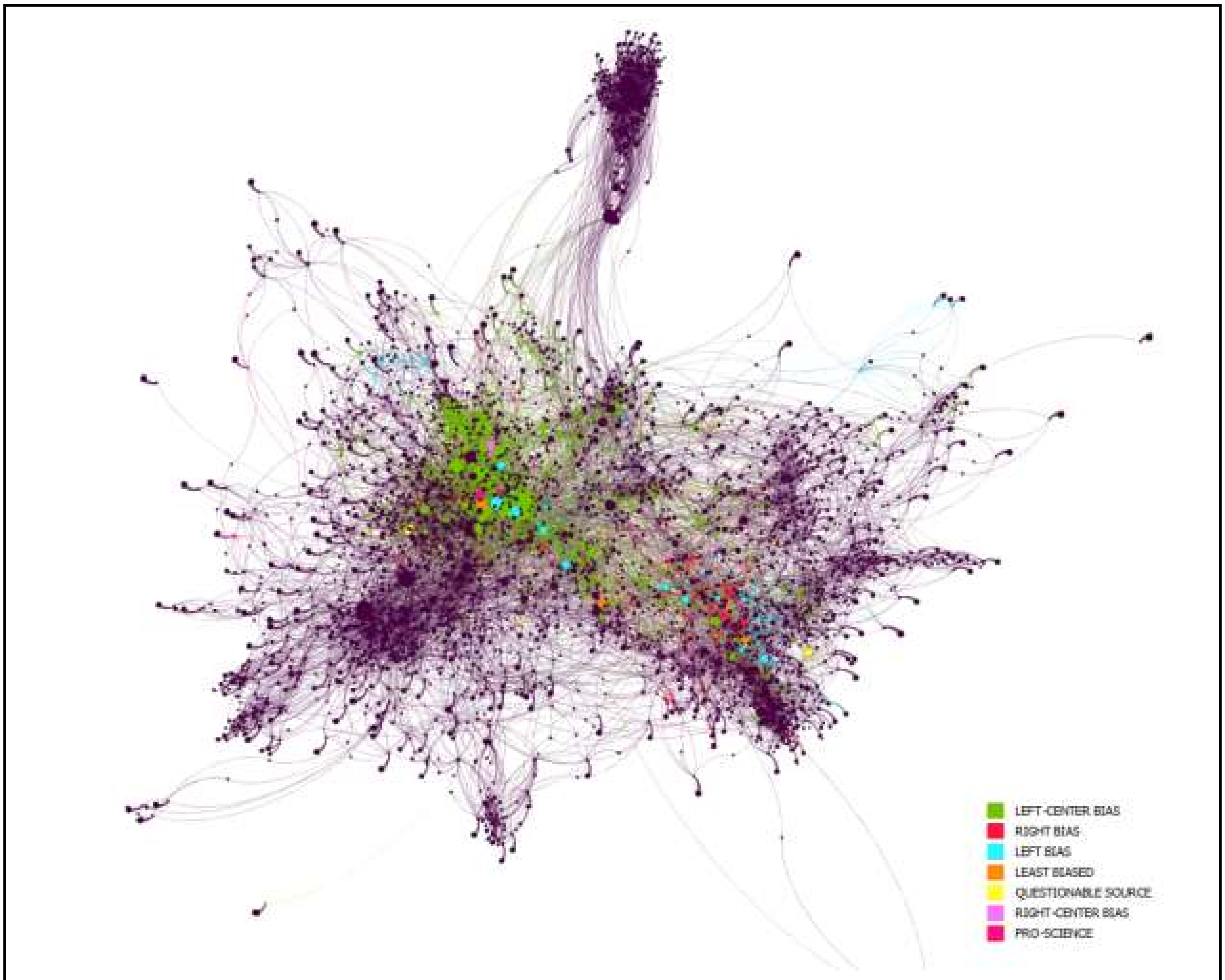

*Factual Accuracy of Sources in Singh Dataset*



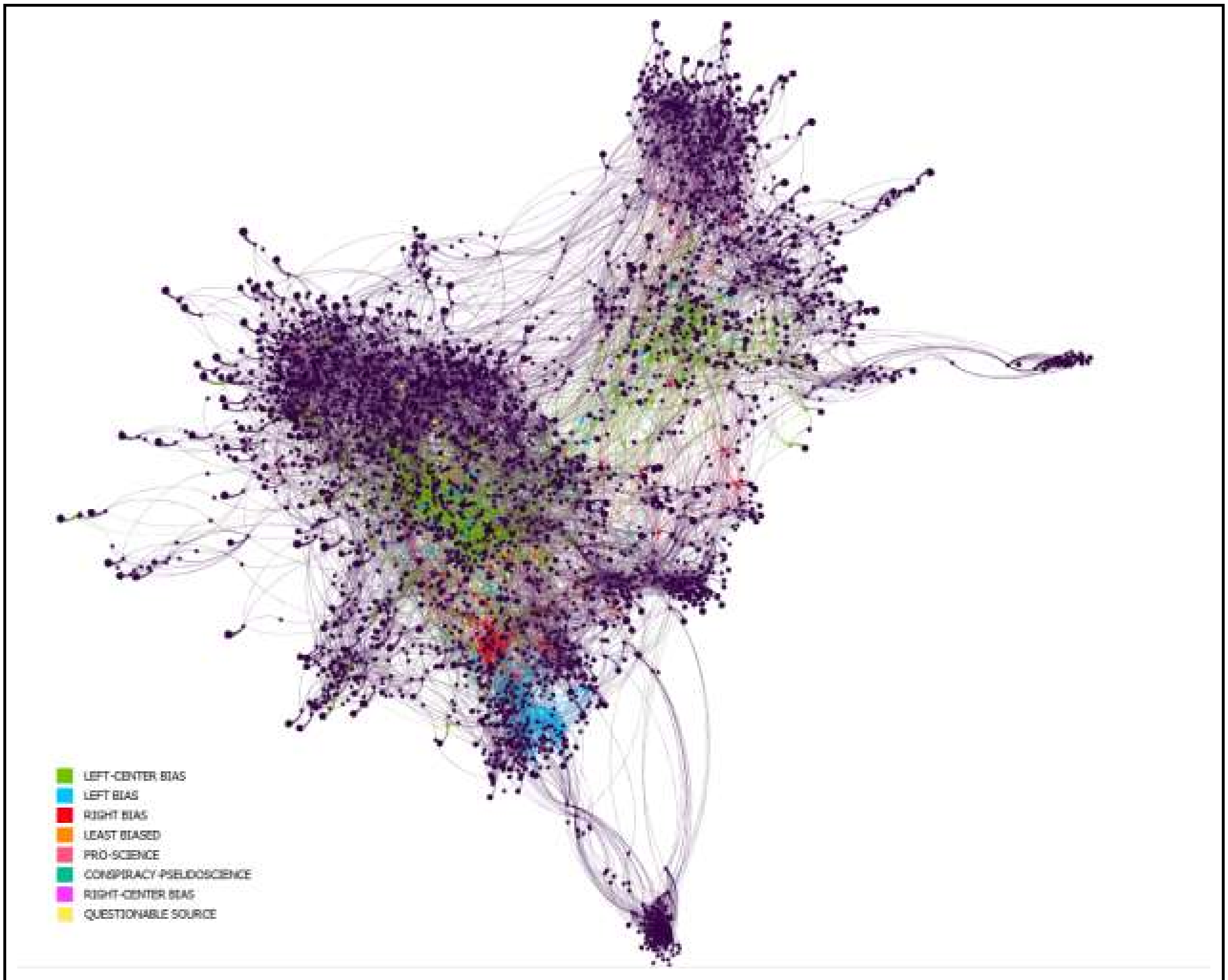

*Factual Accuracy of Sources in Trudeau Dataset*




*Media Bias of Sources in Scheer Dataset*




*Media Bias of Sources in Singh Dataset*




*Media Bias of Sources in Trudeau Dataset*

# APPENDIX B: TIMELINE OF RELEVANT EVENTS

| | |
|---|---|
| 2005 | • February: YouTube is launched (Graham, 2005) |
| 2006 | • October: YouTube is acquired by Google for $1.65 billion (La Monica, 2006) |
| 2007 | • May: YouTube adds Partners Program, allowing users to monetize their content ("YouTube elevates", 2007) |
| 2010 | • February: Autoplay feature added ("Latest changes", 2010)<br>• July: YouTube Leanback is launched (McCracken, 2010) |
| 2012 | • March: YouTube begins to prioritize watch-time over viewcounts ("Changes to Related", 2012) |
| 2014 | • ~June: To better rank recommendations, YouTube implements Sybil, a parallel machine learning platform (Woodie, 2014) |
| 2016 | • ~September: Google Brain implemented at YouTube to recommend videos (Covington et al., 2016) |
| 2017 | • July: YouTube begins redirecting searches for violent extremist content to debunking playlists ("Bringing new Redirect", 2017) |
| 2018 | • July: YouTube launches Top News shelf to better surface trusted journalistic organizations (Schindler, 2018)<br>• July: Fact-checking links and information panels are added (Glaser, 2018) |
| 2019 | • ~January: Reinforce algorithm implemented to recommend videos (Chen, 2019)<br>• January: YouTube announces efforts to reduce recommendations of borderline content ("Continuing our work", 2019)<br>• February: Wikipedia links to anti-vaccination videos are added (O'Donovan, 2019)<br>• June: YouTube redoubles efforts to reduce the amount of borderline content recommended ("Ongoing work to tackle hate") |

## APPENDIX C: DATASET SAMPLE

| COLUMN HEADER | DESCRIPTION |
|---|---|
| Title | Title of Video |
| Channel | YouTube channel the video was posted to. |
| Stance | Relevant videos are coded according to their depiction of the candidate.<br>SSch: Supportive of Andrew Scheer<br>CSch: Critical of Andrew Scheer<br>SSin: Supportive of Jagmeet Singh<br>CSin: Critical of Jagmeet Singh<br>STru: Supportive of Justin Trudeau<br>CTru: Critical of Justin Trudeau<br>O: Neutral or objective depiction |
| Date | Date the video was posted to YouTube. Manually added. |
| Alternative Influence | Coded as being one of the 65 Alternative Influence Network members as described by Rebecca Lewis. |
| Nb_recommendations | Number of times each title was recommended in the dataset. |
| Bias | Media bias as designated by mediabiasfactcheck.com. |
| Factual Reporting | Level of factual reporting as designated by mediabiasfactcheck.com. |
| ID | YouTube URL ID. |
| Views | Number of views the title has received. |
| Dislikes | Number of dislikes the title has received. |
| Likes | Number of likes the title has received. |
| Mult | The average amount (in percentage) title is recommended. |
| Depth | Level of recommendations distant from the seed video that the title was retrieved from. |

*Legend*

**Notes:** This sample was taken from the Scheer dataset. Omitted are 19 additional columns containing the YouTube URL IDs for 19 associated recommendations for each title.

| title | channel | stance | date | alternative influence | nb_ recommendations | bias | factual reporting | id | views | dislikes | likes | mult | depth |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| The Realities Of Trump's Trade War: VICE on HBO Special Report | VICE News | | | | 89 | LEFT-CENTER BIAS | HIGH | 9wjjQ55S4Nc | 1226058 | 2304 | 11186 | 17.13664596 | 1 |
| Has Trudeau Destroyed Canada's Resource Future? - Rex Murphy | Cambridge House International Inc. | CTru | Jan 25, 2019 | | 63 | | | _toqPJpyeGw | 134642 | 172 | 3595 | 12.13043478 | 1 |
| Former Ontario attorney general weighs in on SNC-Lavalin affair | cpac | | | | 58 | LEAST BIASED | HIGH | CkWd_3ShjQA | 21470 | 15 | 489 | 11.16770186 | 1 |
| Why so many Americans in the middle class have no savings | PBS NewsHour | | | | 54 | LEFT-CENTER BIAS | VERY HIGH | tamC-M8TxtY | 1571767 | 1774 | 12448 | 10.39751553 | 1 |
| Who’s behind the Chinese takeover of a U.S. pork producer? | PBS NewsHour | | | | 43 | LEFT-CENTER BIAS | VERY HIGH | b_atdCfEoOE | 551574 | 664 | 3384 | 8.279503106 | 1 |
| The mathematics of weight loss \| Ruben Meerman \| TEDxQUT (edited version) | TEDx Talks | | | | 36 | | | vuIlsN32WaE | 6117198 | 4606 | 91749 | 6.931677019 | 1 |
| Expert warns Australia could turn into slums in 20 years \| 60 Minutes Australia | 60 Minutes Australia | | | | 35 | | | vRSdiq3sOTc | 2072332 | 2937 | 14579 | 6.739130435 | 1 |
| Stockman: The Crashing Economy Is Going to Bring Down Trump | David Stockman | | | | 35 | | | sdyWCuSbcEs | 87913 | 139 | 1086 | 6.739130435 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Andrew Scheer leaves Justin Trudeau speechless, he didn't see this coming | True Liberty | SSch, CTru | UNAVAILABLE | | 34 | | | _CX_HT67ECk | 90845 | 128 | 1622 | 6.546583851 | 1 |
| Jordan Peterson Calmly EDUCATES Baiting Host On Her Own Show | Crysta | | | Jordan Peterson | 34 | | | m58YoCGpPpE | 503587 | 329 | 8305 | 6.546583851 | 1 |
| D'Souza absolutely DESTROYS leftist college student | Dinesh D'Souza | | | | 33 | | | tN9bu6CP318 | 4656508 | 6589 | 53282 | 6.354037267 | 1 |