Proposal for a New Course in Computer Science: Ethical Challenges in Technology
(Technical Paper)

The Struggle over YouTube's Recommendation Algorithm
(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Sarah Snow
Fall, 2020

On my honor as a University Student, I have neither given nor received
unauthorized aid on this assignment as defined by the Honor Guidelines
for Thesis-Related Assignments

Introduction

On October 1st, 2017, the deadliest mass shooting in U.S. history took place as a gunman opened fire on an outdoor concert in Las Vegas, ultimately resulting in the loss of 58 lives. YouTube users immediately began uploading conspiracy theories full of misinformation claiming that the shooting was a "false-flag" act perpetrated by the U.S. government to promote their opposition to the second amendment (Thompson, 2020). Mainstream news outlets with established fact-checking procedures were not able to publish content as quickly as independent creators. As a result, conspiratorial videos dominated the top of YouTube search results for "Las Vegas shooting" and became prominently featured in the algorithmically-powered recommendation sidebar. YouTube's default settings automatically play recommended videos, which led to a situation where many users found themselves unintentionally watching radical and conspiratorial content (Warzel, 2017).

The proposed STS research seeks to understand the area of causation, looking at whether the increasing popularity of recommendation engines has made users more susceptible to extremist content. The implications of this research are of utmost importance in a time where our world is increasingly divided along political and ideological lines. This paper will additionally explore the ways that social pressure and public scrutiny have influenced YouTube's content moderation policies and algorithmic design, evaluating whether these changes have been effective at reducing the spread of misinformation and harmful videos.

This investigation is just one example of the many ethical challenges that have resulted from technologies developed by computer scientists. The following technical topic will propose a new computer science course, *Ethical Challenges in Technology*, that explores the

intersection of ethical and social issues related to modern computing challenges. Computer software has the power to fundamentally change the way that society functions and the curriculum taught to undergraduate computer scientists must empower students to connect with the underlying social issues that frame technical challenges (Karoff, 2019).

Technical Topic

At the University of Virginia, there are no required courses for a degree in computer science that focus specifically on the ethical ramifications of modern-day computing. Searching the keyword "ethics" on an archive of descriptions for nearly 80 computer science courses offered over the years yields just two results: the now-defunct CS Education Practicum and Introduction to Cybersecurity (Bloomfield, 2020). The notable absence of such courses is not for a lack of demand; undergraduate computer science departments at many well-regarded universities are incorporating ethics education into their curriculum in creative ways. In 2018, Harvard University and the Massachusetts Institute of Technology made major headlines when they announced an effort to collaborate on a joint course addressing "the ethics and regulation of artificial intelligence" (Singer). At Stanford, a variety of new courses with an ethical focus have been added to the undergraduate course offering in recent semesters. The optional elective courses offered at these universities include topics such as "Ethics, Public Policy, and Technological Change," "AI Interpretability and Fairness," and "Ethical and Social Issues in Natural Language Processing" (Abate, 2020).

Beyond individual course offerings, a team of philosophers and computer scientists at Harvard have pioneered Embedded EthiCS, a program that "adds short ethics modules to computer science courses in the core computer science curriculum," an approach which attempts to "addresses shortcomings of stand-alone courses" (Grosz et al., 2019). As the scope and impact

of computer science have grown, it is increasingly essential that undergraduate students are made aware of the potential consequences their work could impose on the general public. These consequences are especially relevant for students who intend to pursue a career in technology. The technical topic will propose a new course that introduces students to the ethical challenges presented by computing.

The proposed course, *Ethical Challenges in Technology*, will be a high-level overview of ethical issues within the technology industry. The course will be a combination of current event analysis and case studies of landmark events in the history of computing. Students will analyze specific products and policies that have fundamentally shifted the course of the industry. This material will be specifically designed for undergraduates in their last year of study, preparing students with the necessary tools to evaluate the impact of public-facing industry work. The content will not be limited to students pursuing a career in software development; instead, the curriculum will cover a broad range of potential professional pursuits. Each unit of the course will focus on an existing topic in the University of Virginia computer science offering: Algorithms, Cybersecurity, Artificial Intelligence, Machine Learning, and others. Collaborating with professors that teach these topics will ensure that the material is not overly repetitive or beyond the scope of the original course. This structure will allow students to combine existing knowledge with a broad perspective on how these technologies affect the larger industry.

The final technical deliverable for this capstone will include a detailed syllabus for the course, as well as a sample week's-worth of material. There will be an outline included that describes the delivery method for the course content, and an example take-home assignment will be produced. This project aims to present a practical, thorough course proposal that will allow

undergraduate students to develop an understanding of the ethical implications of work being done by computer scientists.

STS Topic

In early 2012, YouTube was a massively popular platform bringing in over four billion views per day. Despite this success, YouTube's engineering director Cristos Goodrow made an incredibly risky decision to completely rework the existing recommendation system. Instead of prioritizing videos that receive a large number of views, the site would now promote content that increased the watch time of the average user (D'Onfro, 2015). Goodrow's idea was simple: more engagement would lead to increased advertisement revenue. YouTube wanted their users to be addicted to their website, and the easiest way to achieve that goal was to move away from a subscription-based service and towards a system that revolves around recommendations. On paper, it was a massive success. Technology researcher Casey Newton wrote an article describing the quantitative impact of the decision and noted that "more than 70% of the time people spend watching videos on the site [was] now driven by YouTube's algorithmic recommendations" (2017). However, this shift in algorithmic design would play a critical role in the formation of a supposed "Rabbit Hole of Radicalization" on the platform, driving users to consume radical content (Ledwich & Zaitsev, 2019). This STS research paper seeks to understand the role of YouTube's recommendation algorithm in the assimilation of false information and extremist ideology.

The most important implication of this research is the scope: recommendation systems exist across various web platforms and are not unique to YouTube. The machine learning algorithms that drive the YouTube homepage are proprietary to Google and notoriously difficult to audit or replicate. However, even without directly replicating the YouTube algorithm, other

recommendation systems are likely to run into similar issues resulting from two common phenomena: "echo chambers" and "filter bubbles" (Faddoul et al., 2020). Researchers from DeepMind, an artificial intelligence company acquired by Google in 2014, identified these two terms to describe the way that recommendation systems may influence human behavior. The term echo chamber refers to the idea of "user's interest being positively or negatively reinforced by repeated exposure to a certain item or category of items" (Jiang et al., 2019, p. 1). Similarly, the filter bubble effect refers to the idea that such systems "select limited content to serve to users online" (Jian et al., 2019, p. 1). Together, these two phenomena describe the reasons why users interacting with content from recommender systems often find themselves in feedback loops and so-called rabbit holes.

YouTube presents an interesting case study of these phenomena for many reasons. Specifically, YouTube is distinguished from other social networks because users are not required to make an account. The content that users consume on YouTube is not limited to a network of friends or subscriptions. As a result, individuals who publish videos on YouTube have access to a large mainstream audience typically unavailable to independent content creators (Roose, 2019). However, this is not unique to YouTube. The hugely successful short-form video platform TikTok delivers users to a homepage full of carefully curated content that is not linked to their existing social network. The popularity of TikTok and the success of the YouTube recommendation page are not simple flukes. Technology reporters hypothesize that data-driven curated content delivery methods are the future of social applications (Herrman, 2019). The impact of this shift cannot be understated. In a world that is increasingly divided along ideological lines, there are significant social implications to having our virtual lives dictated by artificial intelligence. The user is no longer in charge of the content they are consuming.

Platform-driven content moderation and censorship have become major areas of concern.

Most researchers agree that misinformation is rampant on social networks. However, it is unclear whether these platforms enact change in social behavior or simply mirror existing interactions in a virtual space (Ribeiro et al., 2020). This research project will utilize the framework of technological determinism to evaluate whether recommendation algorithms have been a powerful driving force in social change. Technological determinism refers to the belief that "changes in technology exert a greater influence on societies and their processes than any other factor" (Smith, 1994). Viewing the topic through the lens of determinism will allow for a greater analysis of the causal relationship between the rise in recommendation engines and increased assimilation of fringe content.

One of the most common criticisms of the technological determinist approach is highlighted by Daniel Chandler in his 1995 critique of the determinism framework. Chandler draws attention to the fact that we live in a naturally ambiguous world; any idea that identifies a specific cause or source for change is an oversimplification of the complex circumstances of our society. As a result of this, many theorists have argued that the degree of determinism should be considered as a part of the deterministic approach. The need for a more nuanced approach led to a rise in the popularity of soft determinism, which "claims that the presence of a particular communication technology is an enabling or facilitating factor" rather than the sole cause of social development (Chandler, 2014). Although this approach may result in less explicit interpretations of technological advancements, it is aligned with the reality of our intricate world.

Research Question and Methods

This paper will answer the research question: How has the algorithmically-driven YouTube video recommendation system affected the spread of extremist and conspiratorial content among its users? The media has hypothesized, with largely anecdotal evidence, that

YouTube is a breeding ground for hate speech and radical ideology (Lewis, 2018). The quantitative research on this topic is not as clear-cut, and this paper will utilize documentary research and discourse analysis to review the most relevant articles, studies, and critiques of this topic.

To address this question, a wide variety of academic journals and newspaper articles have been analyzed. Initially, the keywords "Algorithmic Extremism," "YouTube Recommendations," "Recommendation Bias," and "YouTube Radicalization" are used to search for reputable academic research on the subject. In addition to academic research, there have been numerous articles from popular newspapers on the dangers of artificial intelligence and the specific consequences of the YouTube platform. These articles were compiled and sorted into chronological order to enable a clear understanding of the evolution of this topic. The YouTube recommendation algorithm is constantly evolving, as are the company's policies surrounding content moderation and censorship. The chronological approach was the most appropriate method of analyzing a tool that is susceptible to such rapid changes.

After the research is organized chronologically, it is labeled according to the overall findings: sources that support the existence of a radicalization pipeline within YouTube, and sources that did not find quantitative evidence of algorithmic bias. Many of the studies on this topic have been publicly critiqued and subsequently defended in independent blog publications. Peer-revied studies were the main source of information, but the ensuing commentaries were also taken into account to evaluate the validity of the research methods being used. This topic combines a highly technical area of computing with a very complex social response. Gathering sources to explain the technical aspects of the algorithm while also taking the anecdotal experiences of users into account will prove critical in the analysis of this research topic.

Conclusion

The most frightening ethical concern created by modern computing advancements is how little our society considers the true consequences of new technology. Developers are often exclusively concerned with how fast or efficiently they can code a new program. Technology companies are incentivized by profit and increased user engagement. Often, the possible ramifications of the product are not fully recognized. The capstone project seeks to improve the level of attention given to ethical issues involving computing. Starting at the beginning of a computer scientist's career, it is of the utmost importance such impacts are a major consideration. The technical deliverable will propose a new undergraduate course, *Ethical   Challenges in Technology*, that helps students to gain a better understanding of the consequences of technological developments. The STS deliverable will provide research on the YouTube recommendation system and address the causality behind the increased consumption of misinformation and dangerous content on the platform.

References

Abate, T. (2020, August 20). *How the Computer Science Department is teaching ethics to its students.* Stanford Engineering.

    https://engineering.stanford.edu/news/how-computer-science-department-teaching-ethics-its-students

Bloomfield, L. (2020, October 30). *UVa Course Catalog.* Lou's List.

    https://louslist.org/CC/CompSci.html

Chandler, D. (2014, July 3). *Technological or Media Determinism.* Retrieved October 30, 2020,

    from http://visual-memory.co.uk/daniel/Documents/tecdet/

D'Onfro, J. (2015, July 3). *The 'terrifying' moment in 2012 when YouTube changed its entire*

    philosophy. Business Insider.

    https://www.businessinsider.com/youtube-watch-time-vs-views-2015-7

Faddoul, M., Chaslot, G., & Farid, H. (2020). A longitudinal analysis of YouTube's promotion

    of conspiracy videos. *Cornell University*. https:// arxiv.org/abs/2003.03318

Grosz, B. J., Grant, D. G., Vredenburgh, K., Behrends, J., Hu, L., Simmons, A., & Waldo, J.

    (2019). Embedded EthiCS. *Communications of the ACM, 62*(8), 54-61.

    doi:10.1145/3330794

Herrman, J. (2019, March 10). *How TikTok Is Rewriting the World.* The New York Times.

    https://www.nytimes.com/2019/03/10/style/what-is-tik-tok.html

Jiang, R., Chiappa, S., Lattimore, T., György, A., & Kohli, P. (2019). Degenerate Feedback

    Loops in Recommender Systems. *Proceedings of the 2019 AAAI/ACM Conference on AI,*

    Ethics, and Society. doi:10.1145/3306618.3314288

Karoff, P. (2019, January 28). *Harvard works to embed ethics in computer science curriculum.*

    The Harvard Gazette. https://news.harvard.edu/gazette/story/2019/01/harvard-works-to-

    embed-ethics-in-compute r-science-curriculum/

Ledwich, M., & Zaitsev, A. (2020). Algorithmic extremism: Examining YouTube's rabbit hole of

    radicalization. *First Monday.* doi:10.5210/fm.v25i3.10419

Lewis, B. (2018, September 18). *Alternative Influence.* Data & Society.

    https://datasociety.net/library/alternative-influence/

Munger, K., & Phillips, J. (2019). A supply and demand framework for YouTube politics. *Penn*

    State Political Science. https://osf.io/73jys

Newton, C. (2017, August 30). *How YouTube Perfected the Feed.* The Verge.

    https://www.theverge.com/2017/8/30/16222850/youtube-google-brain-algorithm-video-r

    ecommendation-personalized-feed

Nicas, J. (2018, February 7). *How YouTube Drives People to the Internet's Darkest Corners.* The

    Wall Street Journal.

    https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corne

    rs-1518020478

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A., & Meira, W. (2020). Auditing

    radicalization pathways on YouTube. *Proceedings of the 2020 Conference on Fairness,*

    Accountability, and Transparency. doi:10.1145/3351095.3372879

Roose, K. (2019, June 08). *The Making of a YouTube Radical.* The New York Times.

    https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html

Singer, N. (2018, February 12). *Tech's Ethical 'Dark Side': Harvard, Stanford and Others Want*

    to Address It. The New York Times.

    https://www.nytimes.com/2018/02/12/business/computer-science-ethics-courses.html

Smith, M. R. (1994). Technological Determinism in American Culture. *Does Technology Drive*

    History?: The Dilemma of Technological Determinism. (pp. 1-17). Cambridge,

    Massachusetts. London, England. The MIT Press.

Thompson, C. (2020, September 18). *YouTube's Plot to Silence Conspiracy Theories.* Wired.

    https://www.wired.com/story/youtube-algorithm-silence-conspiracy-theories/

Warzel, C. (2017, October 04). *Here's How YouTube Is Spreading Conspiracy Theories About*

    The Vegas Shooting. BuzzFeed News.

    https://www.buzzfeednews.com/article/charliewarzel/heres-how-youtube-is-spreading-con

    spiracy-theories-about