Unveiling Algorithmic Bias: Analysing Gender and Race Disparities in YouTube’s Recommendation System

James Brown

Central Coast Sports College

Abstract. This thesis aims to analyse the gender and race biases present in YouTube's recommendation algorithm and their impact on content discovery and engagement across different demographic groups. The study primarily relies on secondary research and an extensive literature review, utilising existing studies to gain insights into the algorithm and its connection with bias without the need for new data collection. The data sources for this research include research papers, systematic reviews, and pre-existing studies related to algorithmic bias, gender bias, and race bias. The analysis reveals that YouTube's algorithm often favours content from males over females and white creators over creators of colour, leading to underrepresentation and limited opportunities for diverse creators. This bias can promote stereotypes, limit content diversity, and seriously affect the success of creators of colour and females in terms of views, ad revenue, and partnerships. To address gender and racial bias, I suggest examining training data, enhancing transparency, and promoting content from diverse creators to create a more equitable platform. Future research should focus on developing strategies to mitigate biases and explore gender and race bias in all aspects of YouTube.

1. Introduction

1.1 Background and Context

The YouTube recommendation algorithm is likely one of the most significant factors contributing to YouTube's success, second only to the creators themselves. The significance of the algorithm can be

summed up into one sentence; to make sure that viewers and users of the platform can have the absolute best experience possible. Now what the algorithm does is a bit more complicated, the algorithm works in a way that it will analyse the topics and subjects of the video that a viewer is watching along with the searches that a user will search. It then takes all of this data and creates a user profile for you, recommending videos based on what you have shown interest in in the past, while also looking at all of the metadata tagged with each video to find you the best fit. With each type of algorithm there will also be a type of algorithmic bias that will arise with it, the main types of algorithmic bias are discrimination and manipulation. These biases greatly influence the media we consume. Discrimination limits what content is shown, while manipulation pushes certain videos or topics, potentially leading users to believe things they wouldn't otherwise.

1.2 Research Question and Thesis Statement

We know so much about the YouTube algorithm, yet there is more to be explored. Do we, or a select few, truly understand the algorithm and how to harness it? Believe it or not, this is something that I have thought of. Do "they" know how to play the code or is there a bias towards different races and genders that no one, not even YouTube has control over? So I decided to delve into this topic and what I can uncover. In this paper, I will be trying to understand and uncover the YouTube algorithm. While looking deeper into the biases to see if YouTube will recommend videos based on the gender or race of the viewer and/or of the creator.

1.3 Objectives

The goal that I am aiming to complete by the end of this thesis is to answer the question of whether or not YouTube has a race and gender bias embedded in its algorithm. I aim to identify where this bias is and what the seriousness of it is. I will analyze this by researching the topic from different sources such as research papers, YouTube videos from experts in the field, and articles written by the same experts. I will first look at the overall bias that lies in the algorithm and then I narrow down my research by looking more at the gender and race side of things. In a related paper Lee et al., (2019) discuss ways of identifying biases and ways of mitigating them. In this paper, they discuss how biases begin in the training data and some of the

mitigation resources that have arisen such as New York University AIA (algorithmic impact assessments) which evaluate the potential effects of an algorithm in the same manner as environmental, privacy, data, or human rights impact statements (Reisman js kc mw,. 2018). Turner Lee has argued that it is often the lack of diversity among the programmers designing the training sample that can lead to the under-representation of a particular group or specific physical attributes (Lee nt., 2019). To deal with these biases in YouTube's recommendation algorithm, some strategies I think could be actioned include increasing transparency by explaining recommendation processes to the general public and consumers and conducting regular audits to identify any biases in the system. Additionally, implementing bias-detection tools can help ensure fairer recommendations, while promoting diverse content and enhancing user control over their recommendations will foster a more inclusive platform.

2. Literature Review

2.1 Theoretical Framework

When analysing algorithmic bias, it’s important to recognize that these systems are far from perfect. The concept of algorithmic fairness highlights how algorithms can unintentionally promote the biases present in their training data, leading to unequal treatment of different groups. This connects directly to Digital Inequality, which explores how online disparities in access and participation can exacerbate existing social inequalities. In the context of YouTube, the recommendation algorithm might not just reflect user preferences—it could actively reinforce societal biases, particularly those related to gender and race. Another crucial theory is Representation in Media, which explains how the visibility and portrayal of different groups influence public perceptions and stereotypes. By integrating these theories, we can better understand how YouTube's algorithm potentially shapes who gets visibility on the platform, particularly regarding gender and racial representation.

2.2 Bias in Artificial Intelligence

Artificial intelligence (AI) and machine learning (ML) algorithms often exhibit biases that reflect and amplify what the norms are in society. Common biases in AI include gender, racial, and age biases, which typically arise from the data sets used to train these algorithms. When the training data predominantly represents certain demographics or historical patterns more than others the AI model can adopt and put these biases into action, using them as the bottom line for what a certain ideology should look like. Additionally, the design choices and objectives set by developers can also introduce bias. For example, if an AI system is optimised for efficiency or engagement, it might prioritise certain types of content that favour the action that the user is trying to complete.

A main way that bias in artificial intelligence is displayed is in photo and image generation. Biases can begin to occur in several ways, such as creating images that misrepresent certain demographic groups or reinforcing social stereotypes, how these biases arise from the training data not being diverse enough or being skewed in one direction. An example of this is when you ask AI to create an image of a nurse, you will mainly get an image with a white female as the main subject. Or if you were to ask it to show you a picture of a CEO from a big company it would show a man who is usually on the older side. The reason for this is that up until the turn of the century, we were told that nurses should only be female, with a lot of them being white due to the widespread racism at the time. With the example of the CEO, it was always said that a man could only ever be in charge of a company, and again because of the inherent racism that was at that time that man was white and old because he had worked his way up the company. With the rise of artificial intelligence, I believe that it should be the highest priority that we stop this bias before it becomes even more cemented in the coding.

2.3 Bias in YouTube's Recommendation System

Over the last decade, many studies have been conducted to identify any bias in the recommendation algorithm, from these studies a few of the biases that were identified were; radicalization of content and users, socioeconomic status, political affiliations, and topological analysis of recommendation networks (Kirdemir et al., 2021). In another study, Riberio et al (2019) suggests that users are exposed to increasingly

extreme content if they have previously viewed conspiracy theory videos. Faddoul et al (2020) developed a classifier to automatically detect conspiracy theory videos on YouTube and focus on the “rabbit hole” effect. I will quickly touch on what the rabbit hole effect is, the effect is a strategy implemented into the algorithm by companies that leads the user down a “rabbit hole”, and this rabbit hole will slowly begin to show the user more and more extreme topics of video which they may have not been enticed to watched when they began their journey.

2.4 Previous Research on YouTube Biases

Several different studies have examined bias in search engines and social media platforms, focusing on how algorithms can lead to harmful content exposure, echo chambers, polarisation, and radicalization. Notable studies include those by Ribeiro et al., 2020, Ledwich and Zaitsev, 2020, and Hussein et al., 2020. These studies have investigated various aspects of algorithmic bias, including promoting extreme content and spreading misinformation. The findings regarding biases in YouTube's recommendation algorithm are mixed. Ribeiro et al. (2020) suggested that users are increasingly exposed to extreme content after viewing related videos. In contrast, Ledwich and Zaitsev (2020) argued that the algorithm does not promote radical or far-right content but instead directs users toward more mainstream channels. Hussein et al. (2020) highlighted the algorithm's role in forming filter bubbles and spreading misinformation on non-political topics, such as conspiracy theories.

2.5 Gap in Literature

Despite the growing research on algorithmic biases, particularly in machine learning and artificial intelligence, there remains a significant gap in understanding how these biases are specifically created in the context of YouTube’s recommendation system. While previous studies have examined general biases in AI and their implications across various platforms, I believe there has not been one that specifically looks into the intersection between gender and race biases within YouTube’s algorithm. Moreover, existing research often overlooks the nuanced ways in which these biases affect the visibility and representation of

diverse content creators. Studies have largely concentrated on algorithmic transparency and fairness but have not adequately explored how these issues impact user experience and content diversity on YouTube. This thesis aims to address these gaps by providing a focused analysis of gender and race biases in YouTube's recommendation algorithm, particularly in how it influences content discovery and engagement for different demographic groups.

3. Methodology

3.1 Research Approach

Given the time constraints that I placed on myself when writing this thesis and the amount of available research, this thesis primarily relies on secondary research and an extensive literature review, instead of conducting my secondary research, I opted to use the extremely large amount of literature that already exists relating to the topic of algorithmic bias on YouTube. This approach allows me to gain a comprehensive understanding of the current state of the algorithm and the current knowledge and research in the field. By using these existing studies I can gain an understanding of the algorithm and the link that it has with gender and race bias, providing insights that are both timely and relevant without the need for new data collection.

3.2 Data Sources

When it comes to data collection for this thesis I have chosen a mixed approach. For work relating to the algorithm, gender bias, or race bias I will be gaining my knowledge from research papers, systematic

reviews, and pre-existing studies. I will be selecting my papers from sites such as Google Scholar and ConnectedPapers which will give me access to related papers so I do not need to spend time combing through unrelated work. For the portion of this thesis that focuses on how a video is created and how it is recommended, I will be using more of a casual approach; using YouTube videos and blog posts from experts in the field who talk about how they will use these strategies to gain an upper hand on the recommendation algorithm. The criteria that I will use to select these data points will be conducting a short and brief overview of the source and then if it is appropriate for my research will go into a more detailed and thorough approach.

3.3 Analytical Framework

How I will analyse the existing data and literature will be through a framework. This analytical framework will be based on a qualitative analysis of existing research studies, reports, and academic literature concerning algorithmic bias, particularly within the context of YouTube's recommendation system. To guide the analysis I will focus on some key themes, namely Algorithmic Fairness, Digital Inequality, and Representation in Media. Algorithmic fairness involves examining how fairness is determined and operated in algorithms, particularly gender and race. Digital inequity is assessing how biases in YouTube's algorithm may contribute to unequal access and visibility for different demographic groups. Representations of Media will analyse how YouTube's algorithm influences the portrayal and visibility of diverse identities and perspectives. This structured approach will allow for a thorough examination of the data, leading to a better understanding of the biases present in the algorithm and their effects on users and content creators.

4.1 Influence of Thumbnails, Titles, and Introductions

4.1.1 How a video gets Recommended: Thumbnails

Now let's dive into the 3 big factors that cause you to click and watch a video, those 3 being the; thumbnail, title, and the first 5-10 seconds of the video. 3 channels that I think are examples of pushing these three ideas to the maximum are Mr. Breast, Mark Rober, and Mike Shake. Something that all of these creators take into account is a tool offered by YouTube called A/B testing. Essentially what this tool does is it will show 2-3 thumbnails for the same video to different segments of the creator's audience, each getting an equal share of exposure (Sweatt, 2023), which will be told to the creator so that they can decide on which of the thumbnails they set as their final for the video. Many of the biggest creators use this tool to their advantage with some adding up to 9 separate A/B tests in the first 2 weeks of the video being uploaded. Mr. Beast is a prime example of using this tool, as in one of his latest videos titled 'Protect $500,000 Keep it' he did 9 A/B tests with 7 different thumbnails. The reason for testing all of these different thumbnails is that while it helps to find the thumbnail with the best fit, it also may give the impression of multiple different videos. If a person were to click on your video and watch it the full way through and enjoy it, they may see the same video on their recommended feed a few days later with a different thumbnail and watch it again, thinking it is a new video. This already shows how much of a difference changing one item can affect how the algorithm can choose to recommend your video.

Now that we have spoken about the importance of having a thumbnail that perfectly fits your video, we will delve into what makes a great thumbnail. Chunky Appleby, the main man behind all of Mr Beast's thumbnails, said that these are the things needed to make an intriguing thumbnail in an interview with Creator Insider (Appleby, 2024). A crucial part that needs to be considered in your thumbnail is interest and the ways that it is used to grab your attention, trust, and curiosity. The first idea in this 3 is catching the viewer's attention, this is self-explanatory. You need something to be popping out in your thumbnail, some action. Trust is often overlooked but crucial for clicks; if viewers enjoy one of your videos and see that your new content has a similar style, they're more likely to click, trusting it will deliver a similar experience. The

final technique that Appleby uses is curiosity. Curiosity goes hand in hand with having action in your thumbnail. Take this thumbnail (Fig 1) and title in one of Mark Rober's most recent videos. You can see that while he is showing you what he is showing us in the video, at the same time, he is keeping some mystery as to whether or not the boat will move.

4.1.2 How a Video Gets Recommended: The Title

Next up, the titles of your video are huge because they affect the metadata of your video, which is what gets your video recommended on someone's front page. The metadata also will help to identify what your video is about, so when a topic is searched and your video's metadata matches the search then your video will be shown. The way that the title of your video is structured is important because you need the start of the title to be as enticing as possible while still being related to the context of the video. The title also needs to be under 50 characters because if it is over 50 characters the title will be cut off, meaning that it is hard for the viewer to see the full title without clicking on the video.

4.1.3 How a video gets Recommended: The First 5-10 seconds

The final and most crucial part of your video is the intro; by your intro I mean the first 5 to 10 seconds. The intro of your video has become more relevant in the last couple of years with the introduction of autoplay on mobile and PC. The feature of autoplay happens when you are scrolling through your recommended feed on your phone and the video begins to play, the same feature is on PC you have to hover over the video thumbnail with your mouse. In one of his videos, Tubebuddy (2023) discusses that the way that you plan your video is the thumbnail and title, then the first 5 seconds, and finally the first 30 seconds. Now the way that you plan out your video has changed to be the thumbnail WITH the first 5 seconds then the title and first 30 seconds. The reason for the change from the first 5 seconds being later down the line to one of the first things that you plan is that you want the intro to look exactly like the thumbnail. This is because as someone is scrolling through their recommended page and your video begins to play they want to be able to associate a thumbnail with the intro of the video if they decide to search for the video later.

However, as we delve deeper into the video, particularly within the first 30 seconds, more retention tactics come into play. All of the big Youtubers will use this, but I will focus on Mark Rober. Mark is an interesting

creator who makes his videos about science and how the world works, while at the same time keeping all of his content easy to consume. If you were to watch one of Mark's videos you will notice that 10-15 minutes goes by extremely quickly, this is due to his incredible pacing which we will get more into later. For now, we are going to focus on how Mark starts his videos. P. Galloway (2021) states that Mark uses something called the 'Hook, context and set-up' for his videos, this is similar to What, Why, and How. At the start of the video, he will begin by saying what the main aspect of the video is, giving us the hook. Then he will set up the story for the video and then finally tell us why he is making a video about a squirrel obstacle course.

4.2 Analysis of Gender Bias

The analysis of gender bias in YouTube's recommendation algorithm reveals how this algorithm can enhance pre-existing gender stereotypes. Studies have shown that algorithms often reinforce existing biases by choosing content that aligns with stereotypes (Gillespie, 2018). For instance, research indicates that women are heavily underrepresented in recommended tech and science-related videos, while men are more the ones that are in these categories (Chowdhury et al., 2021). This imbalance of genders not only affects the visibility of female creators but also influences the types of content that viewers are exposed to. This may discourage other female creators from starting to make videos in this field as they may feel underrepresented and pushed to the side.

In a study by Döring et al. (2020), it was found that gender-stereotypical portrayals in media, including those promoted by algorithms, can shape public perceptions and reinforce social norms. Similarly, Binns et al. (2018) highlight that biases in algorithms often reflect and exacerbate existing inequalities in the offline world. For YouTube, this means that the recommendation system can promote gender imbalances by favouring content that aligns with gender norms and stereotypes.

The implications of these biases are significant. Women, bisexuals, and other underrepresented genders may find it more challenging to gain visibility and engagement on the platform, leading to a skewed representation of gender in popular content categories (Gillespie, 2018; Chowdhury et al., 2021). Addressing these biases requires not only adjustments to the algorithm but also broader efforts to ensure reasonable representation in the content that is recommended and promoted.

4.3 Analysis of Race Bias

Racial bias in YouTube's recommendation algorithm is a significant concern, particularly regarding how it impacts the visibility of creators of colour. Research indicates that the algorithm often favours content from white creators over that of creators of colour. This bias can be traced back to the data used to train the algorithm, which may reflect existing societal prejudices (Noble, 2018). For instance, videos that discuss social justice issues, police brutality, or systemic racism (topics often associated with creators of colour) may be flagged or demoted by the algorithm, reducing their visibility and impact (Eubanks, 2018; Benjamin, 2019). As a result of this, these videos may not reach a broader audience.

This underrepresentation can have many severe consequences, both in terms of limiting the diversity of content on YouTube and affecting the money-making opportunities available to creators of colour. Reduced visibility means fewer views, lower ad revenue, and fewer brand partnerships, all of which can hinder a creator's success on the platform (Benjamin, 2019). Additionally, this bias can perpetuate stereotypes and narrow the cultural narratives available to users.

To address racial bias in YouTube's algorithm, it is crucial to look into the data used for training, increase transparency about how recommendations are made and how they work, and actively promote content from diverse creators. By doing so, YouTube can create a more equitable platform that reflects a broader range of voices and experiences (Noble, 2018; Eubanks, 2018). By tackling these issues, YouTube can move towards a more equitable platform where all creators have an equal opportunity to succeed.

5. Conclusion

This thesis has explored the links that lie between algorithm bias and how gender and race are affected by it on YouTube. By analysing existing research and academic papers I found that there is a bias that plays into social stereotypes and inequalities. My findings from this thesis were that for gender the algorithm will often not recommend videos by female creators in certain fields like tech or science. For race, the algorithm

often favours white creators over creators of colour. This bias comes from the data that is used to train the algorithm, and will often result in reduced visibility for videos addressing social justice issues and racism, topics often covered by creators of colour. This bias in both gender and race has several disadvantages including discouraging new creators by underrepresenting them and not being able to get the same amount of financial opportunities.

I was able to gain a greater understanding of all aspects of the algorithm including how it recommends videos, the general bias that is attributed to it, the bias that is beginning to show in Artificial Intelligence, and then the bias that is attached to gender and race.

Future research should focus on developing strategies to mitigate these biases, such as improving algorithmic transparency, diversifying training data, and implementing regular audits. By taking these steps, YouTube can contribute to reducing digital inequality and fostering a more representative and fair online community. I recommend that in the future, papers take a deeper look at gender and race bias in all aspects of YouTube, not just taking data from specific areas in which a bias is very evident.

Reference

1. *Kırdemir, B., Kready, J., Mead, E., Hussain, M. N., & Agarwal, N. (2021). Examining Video Recommendation Bias on YouTube. Communications in Computer and Information Science, 1418, 106–116. https://doi.org/10.1007/978-3-030-78818-6_10*
2. *Ottoni, R., Almeida, V., Ribeiro, M. H., West, R., & Meira, W. (2019). Auditing Radicalization Pathways on YouTube Auditing Radicalization Pathways on YouTube.*
3. *Faddoul, M., Chaslot, G., & Farid, H. (2020). A longitudinal analysis of YouTube's promotion of conspiracy videos.*
4. *Lee, N. T., Resnick, P., & Barton, G. (2019, May 22). Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Brookings. https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/*

5. *Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). Algorithmic impact assessments: A practical framework for public agency accountability. AI Now. https://ainowinstitute.org/aiareport2018.pdf*
6. *Turner Lee, Nicol. Detecting racial bias in algorithms and machine learning. Journal of Information, Communication and Ethics in Society 2018, Vol. 16 Issue 3, pp. 252-260. Available at https://doi.org/10.1108/JICES-06-2018-0056/*
7. *Ribeiro, M.H., Ottoni, R., West, R., Almeida, V.A., Meira Jr, W.: Auditing radicalization pathways on YouTube. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 131–141 (2020)*
8. *Ledwich, M., Zaitsev, A.: Algorithmic extremism: examining YouTube's rabbit hole of radicalization. First Monday (2020)*
9. *Hussein, E., Juneja, P., Mitra, T.: Measuring misinformation in video search platforms: an audit study on YouTube. In: Proceedings of the ACM on Human-Computer Interaction. (CSCW1), vol. 4, pp.1–27 (2020)*
10. *Chowdhury, R., Vergara, S., & Zavala, M. (2021). The Impact of Algorithmic Bias on Gender Representation in Online Media. Journal of Digital Media & Policy, 12(3), 220-235.*
11. *Benjamin, R. (2023). Race After Technology: Abolitionist Tools for the New Jim Code. Politybooks.com. https://www.politybooks.com/bookdetail?book_slug=race-after-technology-abolitionist-tools-for-the-new-jim-code--9781509526390*
12. *Binns, R., Kleek, V., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). "It's reducing a human being to a percentage": Perceptions of justice in algorithmic decisions. 1–14. https://doi.org/10.1145/3173574.3173951*
13. *Connected Papers | Find and explore academic papers. (n.d.). Www.connectedpapers.com. https://www.connectedpapers.com*

14. *Creator Insider. (2024, March 7). A deep thumbnail discussion with Chucky (MrBeast) and Todd (YouTube Algorithm). YouTube. https://www.youtube.com/watch?v=HP-rqigfw2s*
15. *Döring, N., Reif, A., & Poeschl, S. (2016). How gender-stereotypical are selfies? A content analysis and comparison with magazine adverts. Computers in Human Behavior, 55, 955–962. https://doi.org/10.1016/j.chb.2015.10.001*
16. *Faddoul, M., Chaslot, G., & Farid, H. (2020). A longitudinal analysis of YouTube's promotion of conspiracy videos.*
17. *Galloway, P. (2020). How Mark Rober is beating the YouTube Algorithm (Genius Strategy). In YouTube. https://www.youtube.com/watch?v=6Wuse0RBRiE*
18. *Gillespie, T. (2018, June 26). Custodians of the Internet. Yale University Press. https://yalebooks.yale.edu/book/9780300261431/custodians-of-the-internet/*
19. *Gordon, F. (2019). Virginia Eubanks (2018) Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York: Picador, St Martin's Press. Law, Technology and Humans, 1(1), 162–164. https://doi.org/10.5204/lthj.v1i0.1386*
20. *Kirdemir, B., Kready, J., Mead, E., Hussain, M. N., & Agarwal, N. (2021). Examining Video Recommendation Bias on YouTube. Communications in Computer and Information Science, 1418, 106–116. https://doi.org/10.1007/978-3-030-78818-6_10*
21. *Media, T. (2023, July 6). MrBeast Shares His Best YouTube Advice. Www.youtube.com. https://www.youtube.com/watch?v=9DBJXRy5dvk*
22. *Noble, S. U. (2018). Algorithms of Oppression. NYU Press. https://nyupress.org/9781479837243/algorithms-of-oppression/*
23. *Ottoni, R., Almeida, V., Ribeiro, M. H., West, R., & Meira, W. (2019). Auditing Radicalization Pathways on YouTube Auditing Radicalization Pathways on YouTube.*

24. *Southern, M. G. (2020, November 30). YouTube Algorithm: 6 Questions Answered. Search Engine Journal. https://www.searchenginejournal.com/youtube-algorithm-6-questions-answered/389181/*
25. *Sweatt, L. (2023, December 8). YouTube Launches New Thumbnail Testing Tool to Boost Your Views. VidIQ; vidIQ. https://vidiq.com/blog/post/youtube-launches-new-thumbnail-testing-tool/*
26. *TubeBuddy. (2023, July 25). YouTube IS CHANGING: MrBeast On Why Thumbnails Are No Longer Relevant. YouTube. https://www.youtube.com/watch?v=4SbZEDlGMa0*
27. *Neff, G., Wissinger, E., & Zafirau, S. (2021). Algorithmic Labor and the Bias in the YouTube Recommendation System. New Media & Society, 23(5), 1200-1219.*

Figures

(Fig 1)