

APRIL 2022

DISCUSSION PAPER

Algorithms as a weapon against women:

HOW YOUTUBE LURES BOYS
AND YOUNG MEN INTO THE
'MANOSPHERE'

Reset.
AUSTRALIA

ISD

Powering solutions
to extremism, hate
and disinformation





Foreword

Dr Catriona Wallace, Executive Director, Gradient Institute, Adjunct Professor, AGSM, Founder, Ethical AI Advisory and Executive Chair, Boab AI.

Campaigning for the upcoming Australian Federal election has seen both major parties emphasise the safety and rights of women. Yet, sadly, we do not see this concern reflected in the government or opposition's understanding of, or regulation of, tech giants such as YouTube who continue to design and deploy unacceptable algorithms that promote Manosphere and anti-feminist content, mostly US-generated, to Australia's young men and boys. In this extraordinary study where 10 experimental accounts were set up and tracked, YouTube and YouTube Shorts promoted warped views of masculinity, and encouraged hateful, misogynistic and dehumanising attitudes towards women.

You will find the results of this study shocking, disturbing and alarming.

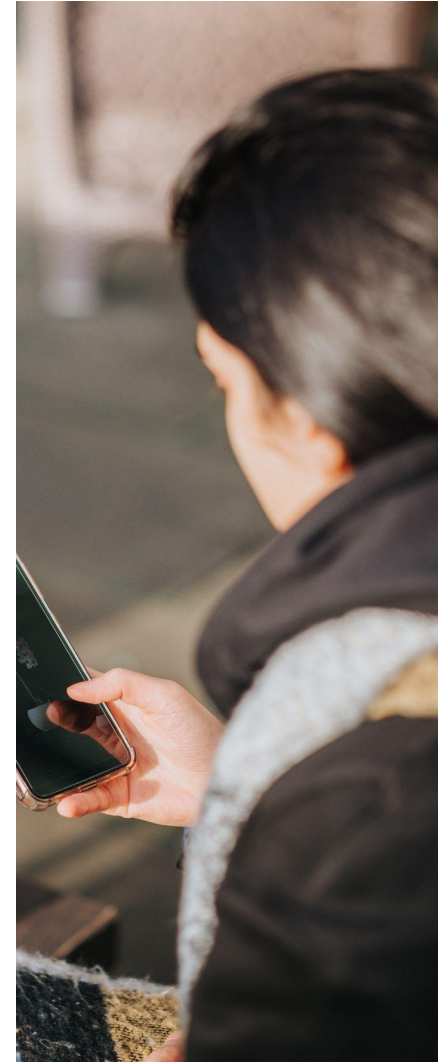
And, with Web3 and platforms such as the Metaverse just around the corner, the threats to women and the vulnerable are only going to get worse, very quickly.

If Australia's major political parties are to effectively improve women's safety and rights, then they must urgently put in place regulation and demand that social media platforms are more transparent about the risks of their algorithms, and are held accountable for the harms they cause. Australia has the opportunity, through this election process, to become both a leader in the safety of women and to better support our young men. Now is the time.



Contents

Executive Summary	1
Introduction	3
Methodology	5
Platform selection	
Avatar accounts	
Seed list and monitoring	
Analysis	7
Main emerging themes	
Main YouTube algorithm responds more rapidly when there are minimal inputs	
The YouTube Shorts algorithm appears to operate more aggressively	
Recommendations on Shorts are the same for under and over 18s	
The majority of the content is from the United States	
Other topics	
Policy recommendations	16
Conclusion	18



Executive Summary

This research documents how YouTube's algorithms contribute to promoting misogynistic, anti-feminist and other extremist content to Australian boys and young men. Using experimental accounts this research tracks the content that YouTube, and their new 'YouTube Shorts' feature, routinely recommends to boys and young men.

This short-term, qualitative study involved analysing algorithmic recommendations and trajectories provided to 10 experimental accounts:

- 4 boys under 18, who followed content at different points along the ideological spectrum, from more mainstream to extreme sources and influencers
- 4 young men over 18, who followed content at different points along the ideological spectrum, from more mainstream to extreme sources and influencers
- 2 blank control accounts that did not deliberately seek out or engage with any particular content, but instead followed the videos offered by Youtube's recommendations.

As the study progressed, each account was recommended videos with messages antagonistic towards women and feminism. Following the recommendations and viewing and liking the suggested content resulted in more overtly misogynist 'Manosphere' and 'incel' content being recommended.

The study found that while the general Youtube interface recommended broadly similar content to topics the accounts originally engaged with, the new shorter video feature, called YouTube Shorts, appears to operate quite differently. Shorts seems to optimise more aggressively in response to user behaviour and show more extreme videos within a relatively brief timeframe. On Shorts, all accounts were shown vastly similar and sometimes even the same specific content from right-wing and self-described 'alt-right' content creators. The algorithm did not make any distinction between the underage and adult accounts in terms of the content served.

This type of content promotes warped views of masculinity, and encourages hateful, misogynistic and dehumanising attitudes toward women. Such content can also serve as a gateway into more extreme ideologies and online communities, and in some cases has led to violent attacks.¹ There are growing calls in a number of countries, including Australia,² to categorise so-called 'incel' attacks, which are motivated by extremist misogyny, as a form of terrorism.

¹ Griffin, J. (2021). *Incels: Inside a dark world of online hate*. [online] Available at: <https://www.bbc.com/news/blogs-trending-44053828>.

² Duriesmith, D., Ryan, L. and Zimmerman, S. (2018). *Misogyny as Violent Extremism*. [online] Available at: <https://www.internationalaffairs.org.au/australianoutlook/misogyny-as-violent-extremism/>.

As Australia seeks to confront the challenges of violence against women and pursue gender equality, it is concerning that a major social media platform appears to be actively promoting such content to its young male viewers without any deliberate prompting.

Ahead of the May 2022 Federal election, both major parties have made women's safety and wellbeing a key component of their policy platforms. The April 2022 Federal Budget, which was passed by both parties, also made significant commitments to supporting women's safety, equality and wellbeing. In particular, the Budget noted the need to address the drivers of violence against women.³

The drivers of violence are complex, but dehumanising and disrespectful attitudes towards women clearly play a central role. The barrage of social media content which promotes misogynistic views and unhealthy perceptions of dating and relationships risks undermining the government's efforts to prevent violence and to educate the community, particularly young men.

Based on these findings, we believe Australia needs to comprehensively reconsider the regulatory framework governing digital platforms to ensure that systemic, community risks such as those posed by YouTube's algorithm, are adequately addressed. This includes:

1. **Focusing on community and societal risks, not only individual risks:** Expand the definition of 'online harms' to address gender-based violence, and violence against women, girls, trans and gender diverse people. The current definitions of online safety – underpinning the Online Safety

Act and other digital regulation - focus on individual harms, and fail to recognise societal or community threats.

2. **Regulation of systems and processes, not only content moderation:** Content-based approaches to regulation, such as those in Australia, have had limited impact in other contexts on the proliferation of harmful content online. In particular, a major problem with these approaches is that although they can help with the removal of specific pieces of harmful content, they do not address the algorithmic amplification of extremist content. Mitigating systems and process risks requires introducing duties of care across the digital regulatory landscape, including the Online Safety Act and Privacy Act.
3. **Platform accountability and transparency:** To address these risks systemically, we recommend that regulation is designed in a way that requires transparency from online platforms, and can compel them to demonstrate that their policies, processes and systems are designed and implemented with respect to the online harms. This might include requirements for algorithmic auditing, or data access for researchers and regulators to assess the effects of platform systems on harmful content and outcomes.
4. **Strong regulators and enforced regulation:** Ensure regulation is strong and enforced, by moving away from self- and co-regulation and resourcing and joining up regulators. Given the limited impact that self-regulation by social media platforms has had on this activity, it is increasingly evident that government regulation of platforms such as YouTube is necessary.

³ Women's Budget Statement 2022-23, The Australian Government, pp. 3 and pp.22, https://budget.gov.au/2022-23/content/womens-statement/download/womens_budget_statement_2022-23.pdf

Introduction

The upcoming federal election has reignited discussions about the role of social media platforms in influencing our public square and information ecosystem. This is particularly true for first time voters who have grown up alongside social media. The role social media platforms can play in recommending political content to young people needs to be well understood and regulated.

There is growing concern around the online radicalisation of young people in Australia. Research examining the radicalisation of Australians aged 19 and below found that active engagement with online social media was a contributing factor in the subject's path to violent extremism.⁴ Similarly, the Australian Security Intelligence Organisation's (ASIO) Director-General Mike Burgess raised awareness in his annual threat assessment of the growing risk of online radicalisation of young Australians, arguing that:

"Two years ago, in my first threat assessment, I noted that ASIO was seeing an increase in the radicalisation of young Australians. Unfortunately and alarmingly, this trend is continuing. The number of minors being radicalised is getting higher and the age of the minors being radicalised is getting lower. Most of the radicalisation occurs online, reflecting the dynamic I raised earlier, but some of it also happens in person, face to face. Children as young as 13 are

*now embracing extremism, and this is happening with religiously motivated violent extremism and ideologically motivated violent extremism."*⁵

A growing body of research has highlighted the link between the 'Manosphere' and escalation to violence and other forms of 'ideologically motivated' extremism. Misogynist spree killers like Elliot Rodgers in the US and Jake Davison in the UK have sparked debate over whether such actions should be defined as terrorism, in particular where they are linked to a sub-set of the Manosphere who identify as 'involuntary celibates' or 'incels'. In 2020, Canada became the first country to charge an attacker with carrying out an "incel-inspired terror attack."⁶

A striking feature of these attacks is the age of the perpetrators. The Canadian perpetrator was just 17 years old. Elliot Rodgers and Jake Davison were both 22 when they carried out their attacks. While it is difficult to collect data on the average age of participants in the Manosphere, it is apparent that it is serving as a pipeline for very young men and boys into extreme forms of misogyny which, in some cases, escalates to violence.

⁴ Cherney, A., Belton, E., Norham, S.A.B. and Milts, J. (2020). Understanding youth radicalisation: an analysis of Australian data. *Behavioral Sciences of Terrorism and Political Aggression*, pp.1–23. [online] Available at: <https://www.tandfonline.com/doi/epub/10.1080/19434472.2020.1819372?needAccess=true>.

⁵ Australian Security Intelligence Organisation (2022). *Director-General's Annual Threat Assessment*. [online] Available at: <https://www.asio.gov.au/publications/speeches-and-statements/director-generals-annual-threat-assessment-2022.html>.

⁶ Teen charged in Canada's first 'incel' terror case. (2020). *BBC News*. [online]. Available at: <https://www.bbc.com/news/world-us-canada-52733060>.

WHAT IS THE 'MANOSPHERE'?

The Manosphere is a loose collection of movements marked by their overt and extreme misogyny. Movements include 'incels', Men Going Their Own Way (MGTOW) and men's rights activists (MRAs).

These groups are largely situated online, and are marked by a unique lexicon which they use to describe women, sex, and other men. They are explicit in their rejection of feminism which they believe has come to dominate society at the expense of men. Although incels do not primarily focus on racially and culturally defined outgroups, there is notable overlap more broadly between the terminology and digital tactics deployed by incels and other far-right extremists.

Recent research has found that the Manosphere itself is also evolving to become more extreme. In a 2021 study, researchers found that “over the past decade, the Manosphere evolved into a diverse ecosystem of Web communities with various interests/ideologies. The growth of communities that espouse more extreme positions against women and women's rights such as MGTOW and Incels overshadows older communities such as PUA [Pick Up Artists] and MRAs.”⁷

The Manosphere could not exist in its current form without social media, and the way in which social media platforms operate inevitably impacts how online Manosphere content and communities evolve. This is particularly true of social media platforms which rely on algorithmic recommendations such as YouTube. Recommendations can serve as a form of cross-propagation between the Manosphere and other fringe or

extreme communities, as was observed in the research conducted for this report.

In a large-scale study of comments in 526 YouTube channels between 2006 and 2018, researchers [observed](#) a strong link between the Manosphere and the 'alt-right' in which “users [on YouTube and Reddit] in the Manosphere systematically go on to consume alt-right content.”⁸

This study set out to contribute to the growing body of research on the online radicalisation of young people towards extreme ideologies. The aim of this investigation is to understand the balance - or lack thereof - of the content that YouTube's algorithms are serving to young Australian men. It explores the nature, volume and extremity of misogynistic content they are served over a short period of time.

⁷ Ribeiro, M.H., Blackburn, J., Bradlyn, B., De Cristofaro, E., Stringhini, G., Long, S., Greenberg, S. and Zannettou, S. (2020). The evolution of the Manosphere across the Web. *arXiv preprint arXiv:2001.07600*. [online]. Available at: <https://arxiv.org/pdf/2001.07600.pdf>.

⁸ Mamié, R., Horta Ribeiro, M. and West, R. (2021). Are Anti-Feminist Communities Gateways to the Far Right? Evidence from Reddit and YouTube. *13th ACM Web Science Conference 2021*, pp. 139-147. [online]. Available at: <https://arxiv.org/pdf/2102.12837.pdf>.

Methodology

PLATFORM SELECTION

YouTube is one of the most popular social media platforms in Australia;⁹ a 2018 study by the eSafety Commissioner found that 86% of teens aged between 13 and 17 have used Youtube over the course of a year.¹⁰ For this reason, our investigation focussed on YouTube and additionally, on YouTube Shorts, a new feature introduced in 2021, that mirrors the short video format of TikTok.¹¹ The recommendation algorithm is of particular interest as it drives 70 percent of user watch time.

⁹ Ramshaw, A. (2020). *Social Media Statistics for Australia*. [online] Genroe. Available at: <https://www.genroe.com/blog/social-media-statistics-australia/13492>.

¹⁰ Office of the eSafety Commissioner (2018). *State of Play - Youth, Kids and Digital Dangers*. [online] Available at: <https://www.esafety.gov.au/sites/default/files/2019-10/State%20of%20Play%20-%20Youth%20kids%20and%20digital%20dangers.pdf>.

¹¹ Youtube Australia team (2021). *YouTube Shorts arrives in Australia*. [online] Google. Available at: <https://blog.google/intl/en-au/products/connect-communicate/youtube-shorts-arrives-australia/>.



AVATAR ACCOUNTS

To be able to understand how extremist messages are being pushed to young Australian men, ISD analysts set up 10 avatar accounts on YouTube aged between 13 and 20 and located in Australia. In order to monitor how recommended content may vary between younger and older audiences, 4 of the avatars were set up to be under 18 years old and 4 over 18. Two out of the ten accounts were created as research controls that did not engage with content by seed list actors, but instead followed the recommendations of YouTube, focusing on political content.

SEED LIST AND MONITORING

Eight of the accounts were initially "seeded" by following some known right-wing actors from across the ideological spectrum; some could be characterised as more mainstream, while others push extremist messages, and with a varying topical focus. Crucially, the seed lists did not include Manosphere figures.

On eight of the avatar accounts, analysts engaged with content from these actors, watched and liked their videos and subscribed to their channels over the course of two weeks between the 15th and 31st of March, 2022. To be able to track the difference between the recommended videos, some of the avatar accounts started out with more mainstream content, while other avatar accounts watched, liked and subscribed to extremist actors. On the other two accounts, analysts only watched and engaged with YouTube's own recommendations.

The avatar accounts initially interacted with videos solely on the general YouTube interface and subsequently moved to YouTube Shorts. However, throughout the investigation, the accounts' activity alternated between the two features. Analysts then monitored the content being recommended to the avatar accounts.

Analysis

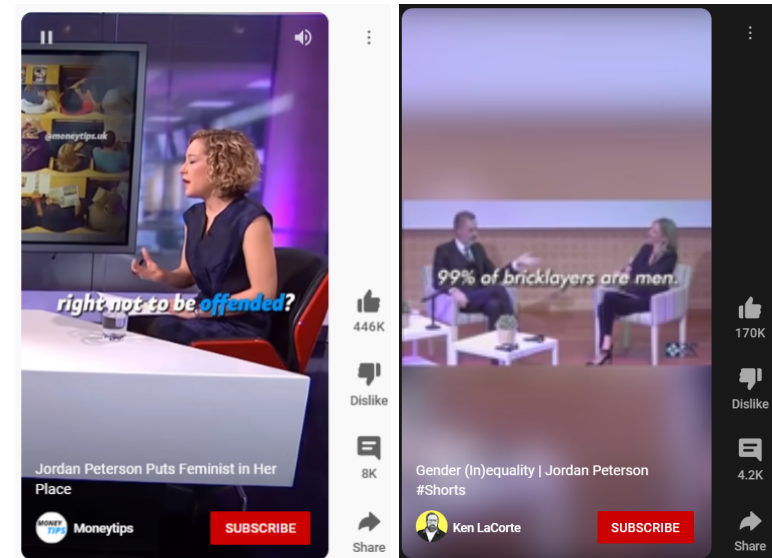
MAIN EMERGING THEMES: ANTI-FEMINIST, MANOSPHERE AND JORDAN PETERSON CONTENT

All of the accounts were recommended content criticising feminist, gender messages, and occasionally abortion. These were the most prominent topics recommended by the algorithm, including for the accounts that were not seeded with right-wing actors.

All accounts were recommended this type of content within hours of creation, without any prompting from the researchers, and such messages continued to be consistently shown to the accounts during the research period. Many of these videos were clips of various media personalities' "takedowns" of feminist arguments.

Notably, such content overwhelmingly featured clips of monologues by or interviews with Jordan Peterson. Peterson is a Canadian author and motivational speaker who focuses on questions of masculinity and gender relations, in particular opposition to feminism. Peterson was not included on the seed list, but content featuring Peterson from a range of channels was aggressively promoted to the accounts in the study. This occurred both on regular YouTube videos and on Shorts.

In one of the below example videos, Peterson echoes a common argument of his that there are certain occupation types where men are strongly overrepresented, for example bricklaying, but these occupations are never a concern for feminist activists, as these are gruelling, labourous occupations.



Videos featuring Jordan Peterson “Put[ting a] Feminist in Her Place” and on “Gender (In)equality” / Source: [Moneytips](#), [Ken Lacorte](#)

The prominence of content featuring Peterson aligns with findings from a recent study of YouTube recommendations by researchers at the Queensland University of Technology. The researchers found that on keyword searches for the word ‘feminism’: “...looking at the top recommended videos across steps, an anti-feminist trend was clear. Videos featuring the controversial public intellectual Jordan Peterson emerge as “winners.”¹²

¹² Matamoros-Fernández, A., Gray, J.E., Bartolo, L., Burgess, J. and Suzor, N. (2021). What's 'Up Next'? Investigating Algorithmic Recommendations on YouTube Across Issues and Over Time. *Media and Communication*, [online] 9(4), pp.234–249. Available at: <https://www.cogitatiopress.com/mediaandcommunication/article/viewFile/4184/4184>.

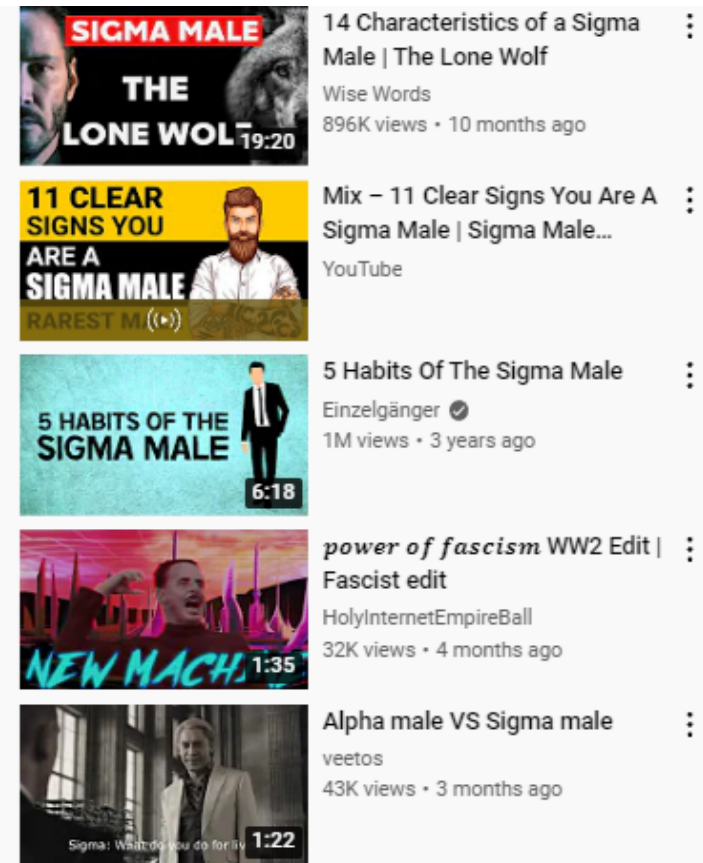
What this study adds to that finding is that content featuring Peterson is recommended to young men on YouTube even when they do not search for 'feminism' or related terms. Liking and engaging with Peterson's content appeared to serve as a gateway for recommendations for more overt Manosphere content, for example videos about 'Sigma males' or anti-feminist content creators like Carl Benjamin (AKA 'Sargon of Akkad'), a former UKIP candidate who made a rape threat to an MP during the UK's 2019 general election campaign.¹³

WHO IS MOST LIKELY TO BE TARGETED WITH MANOSPHERE CONTENT?

Recent survey of 506 16 and 17 year olds conducted by Reset Australia and YouGov found that:

- Respondents who identified as White were more likely to see narratives about the theme 'feminists are lying, men have it worse'. 46% of those identifying as White saw these narratives online compared to 29% of those identifying as Black, Brown, Asian or coming from other Non-White backgrounds, and 19% of those identifying as Aboriginal or Torres Strait Islander background.
- 25% and 44% of respondents identifying as men had seen content from Jordan Peterson and Ben Shapiro respectively, compared with 6% and 24% of respondents identifying as women.

Race and gender clearly influences algorithmic recommendations with young white men being fed particular content themes. There is an urgent need to better understand this algorithmic disparity and its consequences.



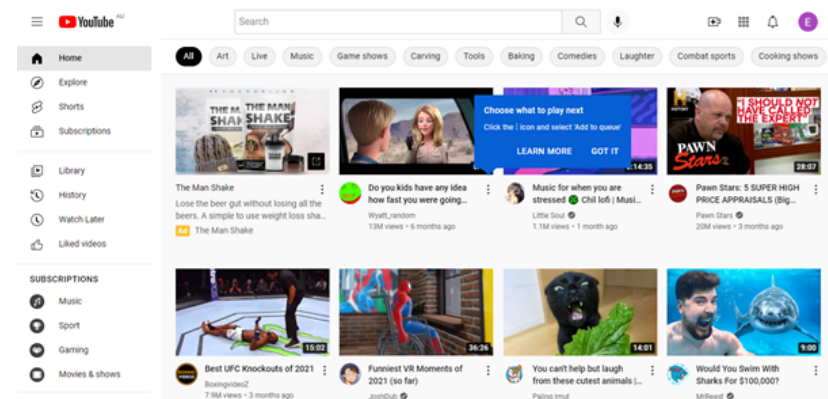
Screenshot of side-bar recommendations to the blank account for 'Sigma male' content (as well as fashwave content. See below for a further discussion of the evolution from Manosphere to fashwave content for this account)

¹³ Phillips, J. (2019). *Police investigate Ukip candidate over Jess Phillips rape comments*. [online] The Guardian. Available at: <https://www.theguardian.com/politics/2019/may/07/police-investigating-ukip-candidate-youtube-carl-benjamin-jess-phillips-comments>.

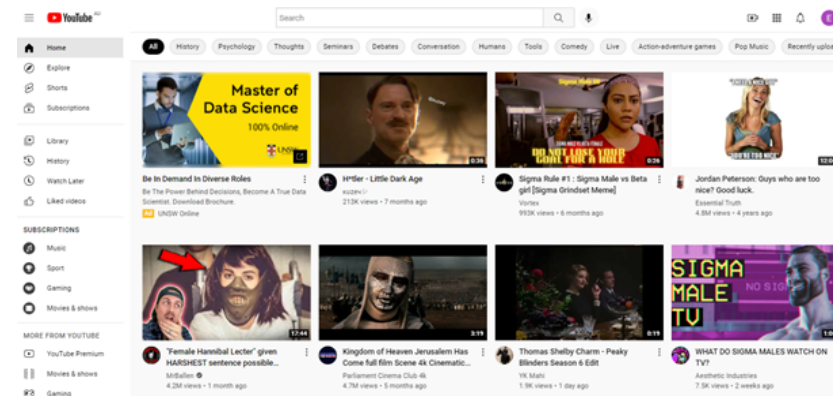
MAIN YOUTUBE ALGORITHM RESPONDS MORE RAPIDLY WHEN THERE ARE MINIMAL INPUTS

When it comes to recommended content, analysts observed that the algorithm on the general YouTube interface was mostly recommending videos from already subscribed channels or other non-relevant videos. For the accounts that were seeded, the algorithm did not appear to recommend content that was more extreme, but it also did not appear to recommend content that was less extreme. The content YouTube provided appeared to stay within the same alt-right bubble.

However, in the case of the 'blank' control accounts which did not engage with seed list content creators and instead purely followed YouTube's recommendations, the algorithmic recommendations did appear to direct the accounts down an increasingly concerning path. For example, one of the blank accounts (which was created as a 19-year-old man) evolved in just over an hour from the home page recommendations on the left, when the account was first created, to the home page recommendations on the right. This happened purely through viewing and liking recommended content (and without, for example, searching for specific terms or subscribing to channels).



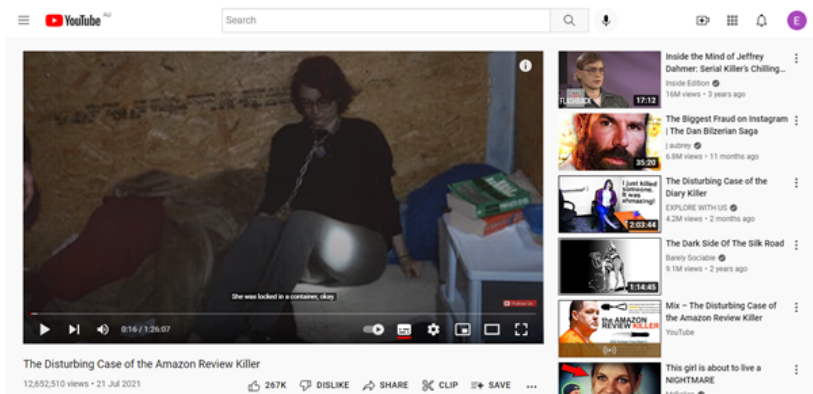
Initial recommendations for blank account when first created



Recommendations after approximately an hour of engaging with recommended content. Recommendations include Manosphere content and a fan-made Hitler music video.

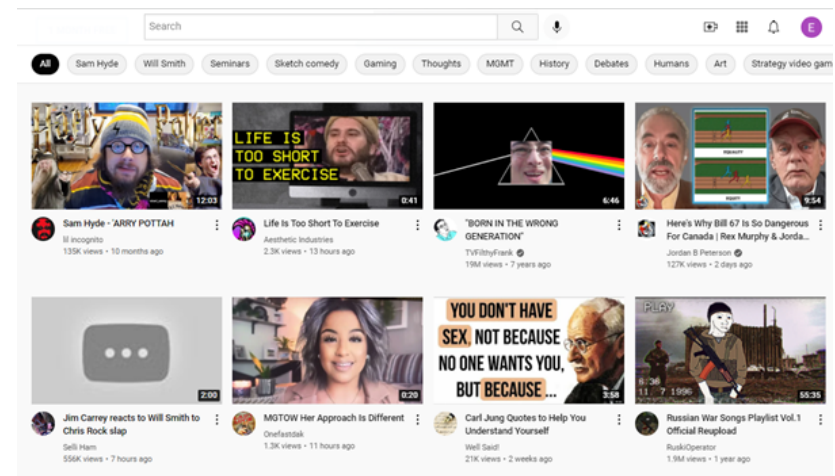
For this blank account, viewing and liking recommended 'Sigma Male' content appears to have served as a bridge to recommendations for [fashwave aesthetic](#) content, and from there to implicit white nationalist content, as we outline below.

Disturbingly, at the same time the YouTube algorithm also began recommending documentaries about notorious male murderers with female victims. This occurred within the first 90 minutes of the account's creation. Content consumed by broad audiences (such as true crime videos) can be problematic when included amongst a deluge of anti-feminist content that could encourage violence against women. Algorithmic amplification of harmful content is not about singular videos, but how videos are linked together to produce a broader mosaic of promoted content.

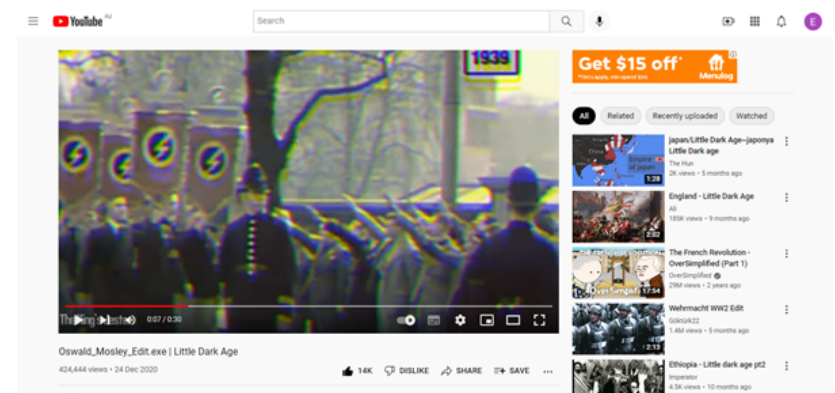


Screenshot of documentary about 'Amazon Review Killer' Todd Kohlhepp, which was algorithmically recommended to the blank account / Source: [EXPLORE WITH US](#)

By the second day of actively using the account, the top recommendations on first opening the YouTube homepage included Men Going Their Own Way (MGTOW) content, which as outlined above is closely related to other forms of Manosphere content.



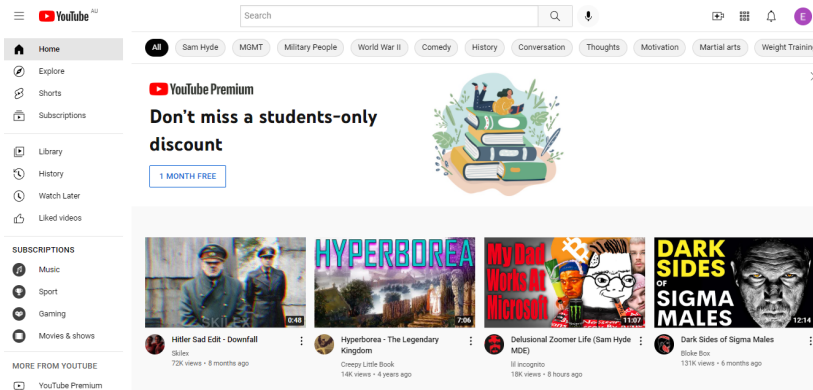
The account also received recommendations for content glorifying the Nazi military in WWII, and then for explicitly pro-fascist content.



Screenshot of music edit glorifying British fascist leader Oswald Mosley, which was recommended to the blank account on the second day

By the fourth day of active use, when opening YouTube the top home page recommendations were a video glorifying Hitler; a video

referencing Hyperborea, which plays a key role in neo-Nazi and Esoteric Nazi content; a video by controversial 'alt-right' comedian Sam Hyde; and a video about “the dark sides of Sigma Males” - as well as an advertisement for a students-only discount for Youtube Premium.’



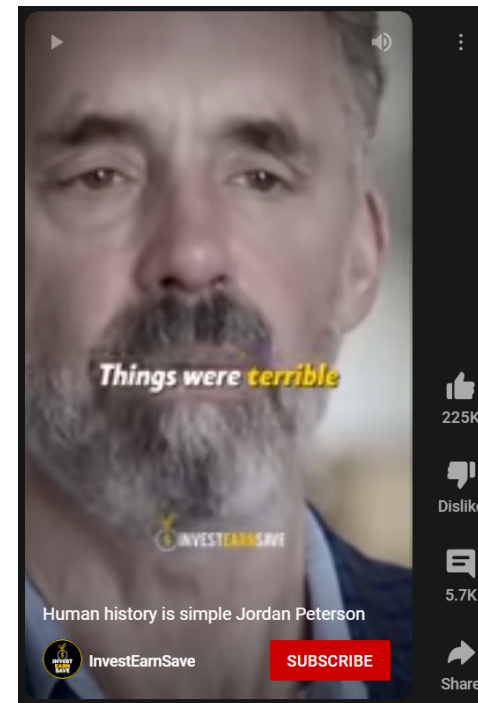
Screenshot of home page on fourth day of active use of the blank account

The behaviour of the YouTube algorithm in the case of this account was notably different from the other accounts in this study. The YouTube algorithm appeared to be more responsive to the account's behaviour. This may have been to compensate for the lack of other inputs beyond the basic account and demographic information provided when researchers set up the account.

THE YOUTUBE SHORTS ALGORITHM APPEARS TO OPERATE MORE AGGRESSIVELY

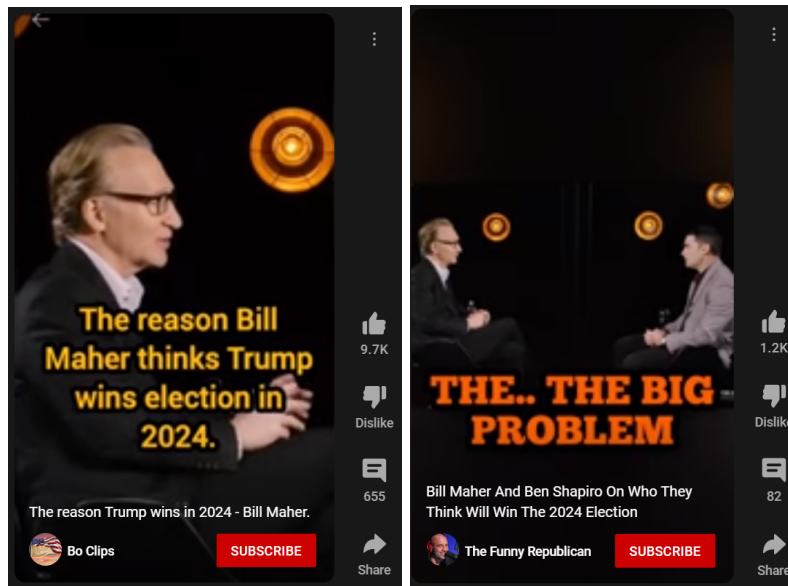
YouTube Shorts recommendations appeared to be faster to catch on and could be trained to push more extreme content faster and more aggressively. In fact, regardless of which seed list actors were fed

into the avatar accounts' activity (and thus how extreme the content they initially engaged with), all accounts were recommended similar content fairly quickly. Notably, not only was similar content recommended to all accounts, but we observed that certain videos were shown to nearly all accounts.



A video of Jordan Peterson describing the hardships people faced in the past shown to nearly all accounts / Source: [InvestEarnSave](#)

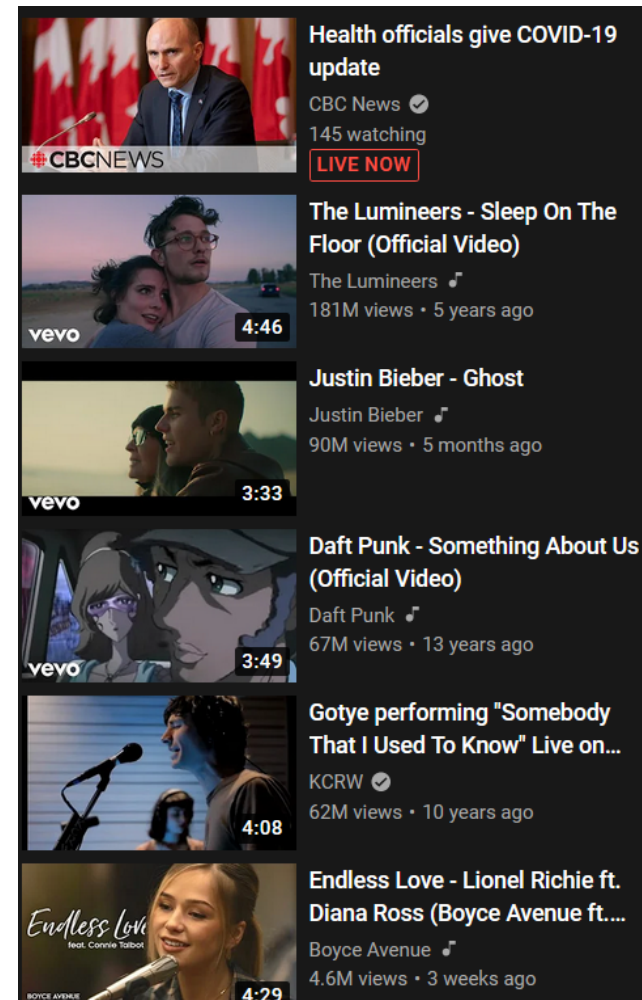
In certain instances, videos depicting the same event, speech or interview, but with a slightly different cut and posted by different YouTube channels, were also shown to the accounts. This signifies that particular messages may be especially popular among audiences and thus, are spread by a number of different actors.



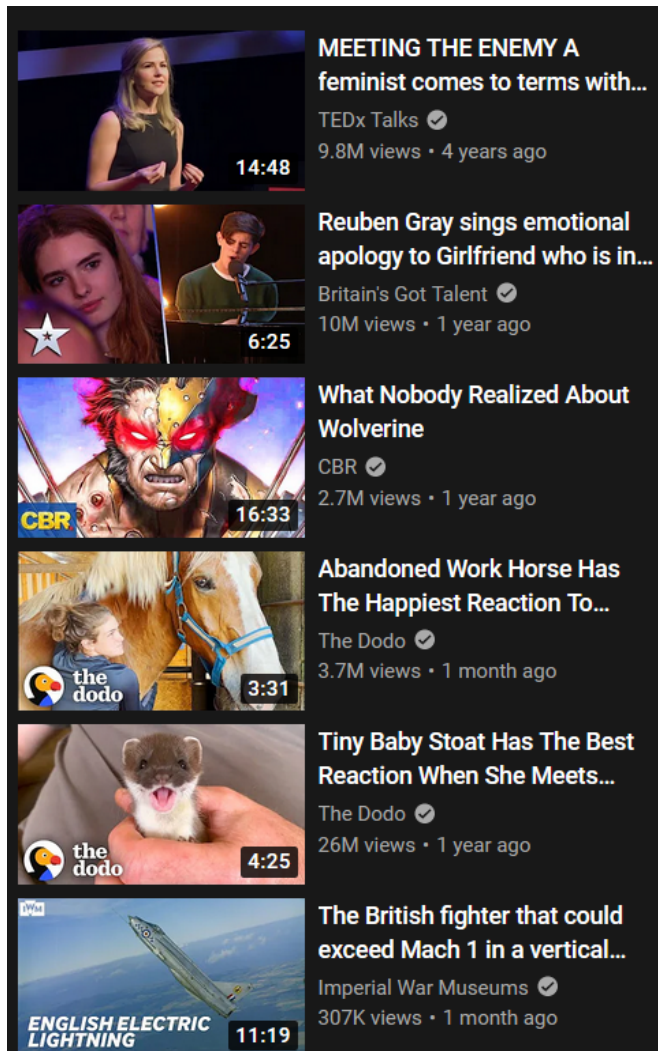
Bill Maher and Ben Shapiro discussing the possibility of Donald Trump winning the 2024 Presidential election - an example of the same event being pushed by different accounts / Source: [Bo Clips](#), [The Funny Republican](#)

RECOMMENDATIONS ON SHORTS ARE THE SAME FOR UNDER AND OVER 18S

Furthermore, analysts observed that on YouTube, recommended videos for accounts aged under 18 were often age-appropriate: funny videos, music videos, videos about animals or interesting facts, as shown below. However, there was no difference in recommendations on Shorts between the underage and adult accounts. The variance of content between accounts overall was minimal on Shorts. This means that viewers below 18 are subjected to the same more ideological messages as those over 18.



Examples of age-appropriate recommended videos with mild content for accounts aged 15 and 13, who engaged with extreme content



Further examples of age-appropriate recommended videos with mild content for accounts aged 15 and 13, who engaged with extreme content

THE MAJORITY OF CONTENT IS FROM THE UNITED STATES

As the examples above show, the majority of videos feature right-wing politicians, political commentators and other actors from the United States and, thus, it appears that young male audiences in Australia are overwhelmingly shown content from the US. Analysts came across videos from Canada and occasionally from Australia among the recommendations, but these were overshadowed by the volume of US-based content.

As described above, Jordan Peterson is by far the most often featured individual, but other prominent right-wing and ‘alt-right’ actors from the United States also often appeared in content shown to the avatar accounts, such as:

- Ben Shapiro, controversial right-wing commentator and founder of The Daily Wire;
- Joe Rogan, controversial podcaster;
- Andrew Klavan, American writer and podcast host at The Daily Wire;
- Benny Johnson, the conservative Turning Point USA’s chief creative officer;
- Candace Owens, pro-Trump activist and right-wing political commentator.

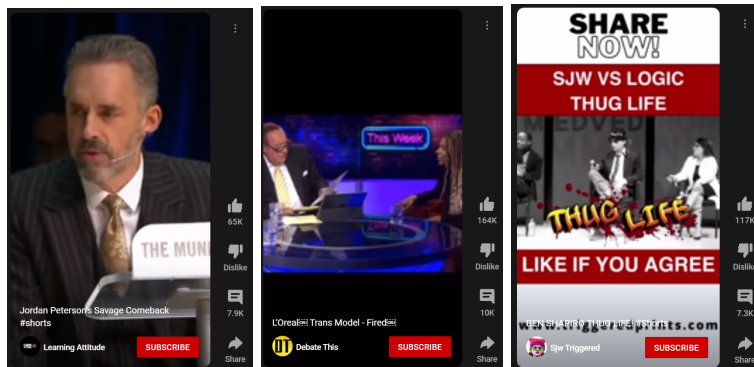
It’s worth noting that content featuring these actors are rarely posted by their own YouTube accounts. Instead, as examples above showed, channels with generic names, such as ‘Learning Attitude’ or ‘Moneytips’, publish these clips.

OTHER TOPICS

There were a number of other prominent and often overlapping topics and themes that emerged in the case of nearly all accounts.

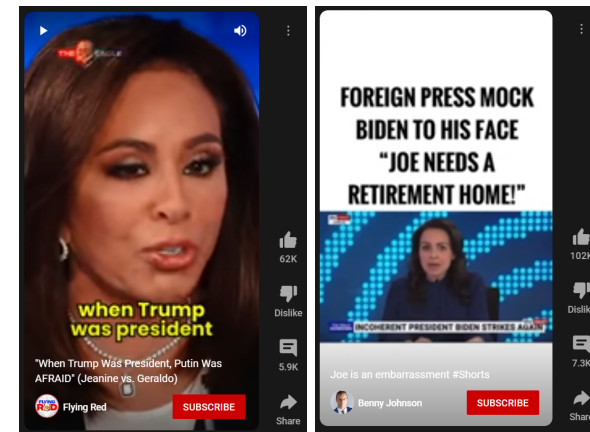
While the public debate around these topics varies considerably, they are almost exclusively right-leaning in ideology, further illustrating that YouTube's algorithm does not present young men with balanced or opposing points of view.

1. Anti-trans content discussing preferred pronouns with right-wing and 'alt-right' commentators arguing that this limits their freedom of speech. In more extreme cases, commentators questioned the existence of transgender people, not acknowledging their preferred gender and mocking LGBTQ+ activists.
2. Challenging the concept of white privilege.



Examples of various commentators challenging the concept of white privilege / Source: [Learning Attitude](#), [Debate This](#), [Sjw Triggered](#)

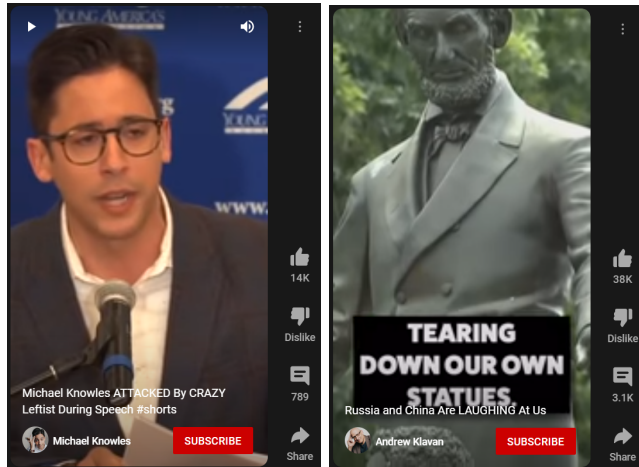
Support for Donald Trump featuring clips of his speeches and interviews, or other political commentators advocating for him. At the same time, content smearing President Joe Biden (and occasionally Vice-President Kamala Harris) also appears frequently, often in combination with pro-Trump content.



Examples of content supporting Donald Trump in which Jeanine Pirro of Fox News claims that Ukraine wouldn't have been invaded if Trump was President and mocking Biden on Sky News Australia /

Source: [Flying Red](#), [Benny Johnson](#)

3. Criticising liberals, progressives, leftists and generally 'woke' culture.



Examples of videos in which progressive narratives are ridiculed /
Source: [Michael Knowles](#), [Andrew Klavan](#)

Additional topics that were recommended, but in lower volumes were:

- Content in support of Vladimir Putin (often in relation to pro-Trump content);
- Covid-sceptic, anti-vaxx, anti-mask content;
- Libertarian content;
- Content advocating for gun ownership;
- Military content;
- Content opposing police (often in relation to Covid-19 restrictions);
- Content smearing journalists and mainstream media; as well as
- Motivational videos featuring aspirational content and guides for life.

Policy recommendations

Australia needs to comprehensively reconsider the regulatory framework governing digital platforms to ensure that systemic, community risks such as those posed by YouTube's algorithm, are adequately addressed. This includes:

- **Focus on community and societal risks, not only individual risks:** Expand the definition of 'online harms' to address gender-based violence, and violence against women, girls, and non-binary people. The current definitions of online safety – underpinning the Online Safety Act and other digital regulation - focus on individual harms, and fail to recognise societal or community threats.
- **Regulation of systems and processes, not only content moderation:** Move beyond content-based approaches to platform regulation and start thinking about how platforms' systems and processes affect harmful activity and content online. Content-based approaches to regulation, such as those in Australia, have had limited impact in other contexts on the proliferation of harmful content online. In particular, a major problem with these approaches is that although they can help with the removal of specific pieces of harmful content, they do not attempt to address the systemic issues with platform design which might underpin the spread and visibility of harmful content. Content-based approaches that focus on individual takedowns do not address the

algorithmic amplification of extremist content. Mitigating systems and process risks requires introducing duties of care across the digital regulatory landscape, including the Online Safety Act and Privacy Act.

- **Platform accountability and transparency:** To address these risks systemically, we recommend that regulation is designed in a way that requires transparency from online platforms, and can compel them to demonstrate that their policies, processes and systems are designed and implemented with respect to the online harms. This might include requirements for algorithmic auditing, or data access for researchers and regulators to assess the effects of platform systems on harmful content and outcomes. The recent report by the House Select Committee on Social Media and Online Safety made strong recommendations about regulation of algorithms, that would harmonise Australia's regulation with international developments in the EU and UK, including directives to:
 - conduct a review of the use of algorithms in digital platforms (recommendation 13),
 - require social media and other digital platforms to report on the use of algorithms through the eSafety Commissioner (recommendation 14) and;

- conduct a Digital Safety Review that examines the possible models for a single regulatory framework under the Online Safety Act (recommendation 19).¹⁴

Strong regulators and enforced regulation: Ensure regulation is strong and enforced, by moving away from self- and co-regulation and resourcing and joining up regulators. Evidence suggests that although platform-driven attempts to limit extremist use of social media through the enforcement of terms of service can impact on the presence of extremist content in the short term, this enforcement does not always have a sustained impact on the presence of extremist communities.¹⁵ Given the limited impact that self-regulation by social media platforms has had on this activity, it is increasingly evident that government regulation of these platforms such as YouTube is necessary.

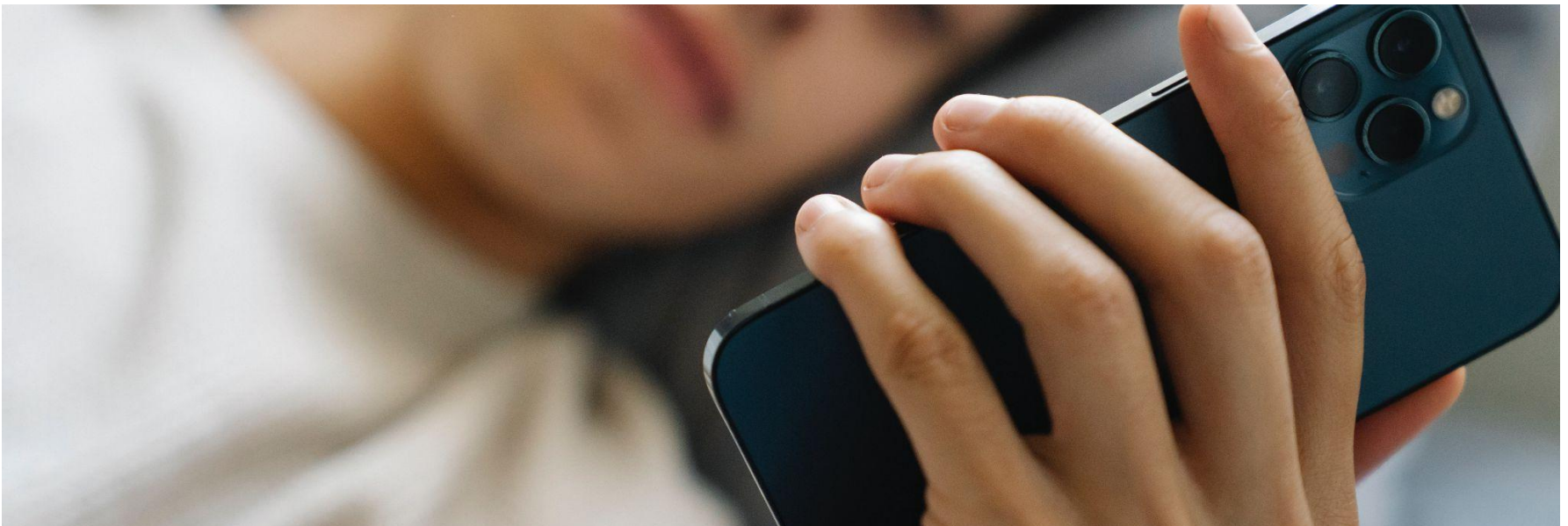
¹⁴ House of Representatives Select Committee on Social Media and Online Safety (2022). Social Media and Online Safety. Available at: https://parlinfo.aph.gov.au/parlInfo/download/committees/reportrep/024877/toc_pdf/SocialMediaandOnlineSafety.pdf;fileType=application%2Fpdf

¹⁵ Hart, M., Davey, J., Maharasingam-Shah, E., O'Connor, C., and Gallagher A. (2021). An Online Environmental Scan of Right-wing Extremism in Canada. Available at: <https://www.isdglobal.org/wp-content/uploads/2021/07/ISDs-An-Online-Environmental-Scan-of-Right-wing-Extremism-in-Canada.pdf>.

Conclusion

In conclusion, our findings show that the avatar accounts of young Australian boys and men were algorithmically recommended content with overtly misogynist and 'Manosphere' messages. Videos promoting unfavourable views on women's rights and feminists, were recommended to all avatar accounts and was the most salient topic of the investigation. Controversial commentator Jordan Peterson, and other similar online and media influencers were remarkably prominent during our investigation. Overwhelmingly the content recommended was by North American right-wing and 'alt-right' actors.

Furthermore, our research has shown that the YouTube Shorts algorithm fails to account for the age of the viewers. It showed extreme and hateful content to all avatar accounts within hours of engaging with videos on the platform. While the general YouTube interface did not escalate the type of content shown to the avatar accounts, it continued to recommend videos from extremist content creators, regardless of the age of the viewers, when the algorithm was primed to do so.



Whilst this study has not considered how the consumption of Manosphere content leads to committing acts of violence, the fact that Manosphere content appears to be being promoted to young men en masse, completely unprompted by the YouTube algorithm, raises serious concerns. Even if the content does not lead to radicalisation, it encourages a dehumanising, disrespectful attitude towards women, and is delivered in deluge without opposing perspectives or any semblance of balance that could represent a dialogue or debate.

As the 2022 Federal election looms, both major parties have promised to do more to address women's safety and rights. As part of the national conversation about the safety of women and girls, and the institutions and systems that cause harm, the role of social

media platforms must be considered. It is unacceptable that algorithms – human designed pieces of code – recommend misogynistic content to young men and boys.

Regulation is urgently needed to demand that social media platforms are both more transparent about the risks of their algorithms, and accountable for the harms they cause.



Reset Australia is an independent
organisation raising awareness and
advocating for better policy
to address digital threats
to Australian democracy.

hello@au.reset.tech | au.reset.tech

