WWG1WGA
#QANON

ENTERING
AREA
51

Donald J. Trump
@realDonaldTrump

I WON THIS ELECTION, BY A LOT!

Official sources may not have ca... the race when this was

10:36 PM · Nov 7, 2020

Q

OUCH!

5G
IS
TOXIC
RADIATION

kun

THE GAME

**UP NEXT:**
**THE MANOSPHERE**

YouTube's Anti-Feminist
Communities' Power to
Radicalise and
How to Prevent It

Justice Monir Mooghen

Master's Thesis
June 2021

AALBORG UNIVERSITET

Techno-
Anthropology

# Master's Thesis

**Title:**

Up Next: The Manosphere


 **Sub-title:**

YouTube's Anti-Feminist Communities' Power to Radicalise and How to Prevent It


**Project period:**

10th Semester,

February-June 2021


**ECTS:**

30 ECTS


**Supervisor:**

Lars Botin


**Name:**

Justice Monir Mooghen


**Report Pages:**

69 pages (165,552 characters)


**Appendix pages:**

32

# Abstract

The online *manosphere* contains communities where misogyny, anti-feminism, and other harmful ideologies are spread, and has been linked to what is called male-supremacist terrorism which studies point to being on the rise in many Western states. Less research has been made in Denmark on online radicalisation of users in the manosphere. This thesis aims to determine if YouTube's recommendation algorithm radicalise in anti-feminist ways and examine preventative solutions which can be made in Denmark. This thesis problem field was analysed based on reviews of literature of YouTube's recommendation algorithm, desk research of preventative measures for online radicalisation, and two conducted expert interviews from Denmark. The thesis finds that YouTube's recommendation algorithm has been found to guide users towards content with extremist tendencies, and that the algorithm (especially) seems to recommend channels in the manosphere to users watching mainstream political content like CNN, Barack Obama, and Bernie Sanders. On this basis, it is recommended that legislation and governmental bodies in Denmark work to ensure ability to monitor, test, and evaluate machine learning algorithms, so that changes can be ensured when needed. The thesis also recommends several other suggestions, where further research is needed post-implementation to monitor their effects.

# Table of Content

# 1 Introduction: Welcome to the Manosphere

YouTube is one of the most widespread social media platforms on the internet and contains billions of user-created contents (Roose and Conger 2019), music videos, educational DIYs and much more. And whilst the video platform has visitors from all over the world and from all ages, a certain corner of YouTube known as the the *manosphere* has been known in recent years to draw men[1] into rabbit holes, slowly radicalising them and introducing them to extremist communities. The 'manosphere' is an unofficial but frequently used umbrella term by scholars, journalists, and experts researching anti-feminist, misogynist and masculinity promoting online movements (Lumsden and Harmer 2019; DareGender and Cybernauterne 2020; Karacan and Crone 2020). The communities act from a point of view where feminism and equality "have gone too far" and that it works in oppressive ways against men (DareGender and Cybernauterne 2020:9). Manosphere communities like *incels*[2] direct their rage of not having sex or girlfriends towards women in general (DareGender and Cybernauterne 2020:9) and often use rhetoric around women not choosing to be with "good guys" (which they see themselves as). They became known after several terror attacks were committed by men connected to incel-communities. Though radicalisation and terrorism have many faces and broad definitions, terms like male supremacist terrorism or incel-related violence have been coined in media and research due to the rising terror attacks which have been conducted by men connected to the online manosphere. These communities exist on both mainstream platforms, and on alternative platforms like the 4chan and 8chan where all users are anonymous, and formal moderation does not exist. Some mainstream platforms like Reddit, YouTube, Facebook, and Twitter have in recent years started *deplatforming*[3] some popular individuals and communities tied to the manosphere, forcing users to find other platforms to engage in. But what are the consequences with these choices, and which other suggested measures exist to prevent online radicalisation?

This thesis will analyse YouTube's machine learning algorithmic infrastructure called the *Deep Neural Network* which recommends videos and is designed to maximise the user's viewing time.

---

[1] This thesis focusses on primarily (cis gendered) white and/or Western men
[2] Involuntary celebrates
[3] Deplatforming is shutting down or removing individuals, groups or content creators

Studies point this to cause users to fall into *filter bubbles* or *rabbit holes* which continuously recommends radicalising content, and few counter-narratives, why it is an important aspect to research if we want to prevent online radicalisation. Currently, there is a lack of governmental authority and legislation to control and monitor the machine learning algorithms affecting the interface of social media platforms, which makes the responsibility and control of algorithms lie in the hands of the tech companies creating them. Through looking within YouTube's algorithm, this thesis will shine light upon whether its side effect can play a part of radicalising users.

The *manosphere* will be introduced and which harmful cultures and consequences, like 'edgy' humour and terror attacks is associated with it. The thesis will also theorise radicalisation, terrorism, and *anti-feminism*. Lastly, the report will explore which preventative measures can be suggested and done in a Danish setting to possibly prevent future terror attacks from being carried out, and more men from becoming radicalised through YouTube.

This thesis will be explored from a techno-anthropological framework, which seems to have been lacking in existing research, since most Danish research focus on pedagogical or legislative suggestions to preventative measures (Stender Petersen and Albert Pers 2020; TechDK Kommissionen 2020; Udenrigsministeriet 2020). Therefore, this thesis seeks to first and foremost study the YouTube algorithm and radicalisation in the online manosphere, to thereafter provide suggestions for preventative measures. I aim to build a bridge between research areas and combine existing findings in a techno-anthropologic scope to focus on social, governmental, and technological suggestions to preventative measures.

## 1.1 What is the Manosphere?

The manosphere is most known for online forums and groups like incels (involuntary celibates), infamous anonymous users of 4chan and 'Red Pillers', using an analogy from The Matrix franchise where men become convinced that they have taken the 'Red Pill', and now see reality as it really is; a society infested by feminism which only favours women (Karacan and Crone 2020:33). The manosphere contains many other different groups of men and male movements like Pick-Up Artists (PUA), Men's Rights Activists (MRA), Men Going Their Own Way (MGTOW) and more. The manosphere is vast and growing on different platforms like YouTube, Facebook, Reddit, 4chan, 8chan, and Telegram. Common within the manosphere are ideologies based on misogynist, anti-feminist, anti-LGBT, and racist values and ideologies seen in for example

ethnonationalism, anarcho-capitalists, neo-Nazis, MRAs, and traditionalists (DareGender and Cybernauterne 2020). Their discourses often focus on the damaged state of our current society and a wish to "go back" to traditional gender norms and family constellations, and the communities frequently have masculinity as their focus. Often the conversations on anti-feminism go together with racist ideas like pure ethno-states and global segregation of cultures and "races", and race realism which argues that different biological human races exist, as seen in several of the radical YouTubers. Feminism is seen as a threat to the traditional masculinity and family constellations and often blamed for "promoting" Western immigration of Muslims and Jewish people "threatening" the white people living in the US and Europe (as seen in Anders Breikvik's manifesto introduced later). The manosphere has many subcommunities, but all exist in the same "universe", many of whom share the same ideas. Because of the strong proximity to each other and shared ideologies between the different sub-communities in the radical Right-leaning political online universe, the term *manosphere* is used as a term that connects these throughout this thesis. The manosphere will be expanded on further throughout the thesis.

## 1.2 Problem field

"[The YouTube algorithm] suggests that being a conservative on YouTube means that you're only one or two clicks away from extreme far-right channels, conspiracy theories, and radicalizing content." (Kaiser and Rauchfleisch 2018)

In recent years, research has started exploring the issues behind the rising male-supremacist terrorism[4] and these have roots in the online manosphere. In a Danish setting, Cybernauterne, a tech-focused organisation specialising in e.g., digital cultures, online hate communities and online harassment, and certain other bodies and commissions are exploring the issue of online radicalisation. Research in publications, reports, and articles on misogynist online spaces and online radicalising still lacks clear preventative solutions and theoretical frameworks to analyse these issues (Stender Petersen and Albert Pers 2020; TechDK Kommissionen 2020; Mogensen and

---

[4] Male supremacy is described as a misrepresentation of women as being genetically inferior, manipulative, and stupid, and reduces them to their reproductive or sexual function with sex being something that they owe men. It is also driven by a biological analysis of women as fundamentally inferior to men, and male supremacists' malign women specifically for their gender. It also comes with a conviction that the current system oppresses men in favour of women are the unifying tenets of the male supremacist worldview. From https://www.splcenter.org/fighting-hate/extremist-files/ideology/male-supremacy (Accessed June 3, 2021)

Helding Rand 2020). Often these focus on either preventative solutions for online radicalisation from a pedagogical point of view, legislative point of view, or analyse machine learning algorithms on social media platforms, but seemingly none which combine all of these scopes. Former projects of mine have focused on discriminatory algorithms, mainly on Instagram, Twitter and YouTube, the lack of transparency with these algorithms and how minorities use social media platforms and (are forced to) enact in defensive ways to avoid both silencing and deplatforming (flagging, removing of content and banning) from social media and being subject to hate and harassment on the platforms.

A lack of algorithmic transparency on social media and the still-new technology of machine learning algorithm has unforeseen consequences, as is written in Lauren Valentino Bryant's paper *The YouTube Algorithm and the Alt-Right Filter Bubble* (2020): "[e]xactly how the algorithm works is a bit of a black box, some of its internal logic is opaque even to its engineers.". The current technological age around knowledge and laws of machine learning algorithms has been described as an era without proper legislative management by Danish expert in digital cultures, Katrine K. Pedersen in a podcast interview (Høst 2020). Even if it is not intentional, YouTube's machine learning algorithm seems to have a bias towards Right-leaning and content with extremist tendencies, making moderately conservative videos only a few clicks away from the manosphere and 'far-Right' content (Bryant 2020, Kaiser and Rauchfleisch 2018, Roose et al. 2020, DareGender and Cybernauterne 2020, Rogers 2020). In late years, YouTube has especially become known for harbouring these extremist content creators and is claimed to serve as a *pipeline* to other, more radical communities and content (Bryant 2020; Kaiser and Rauchfleisch 2018; Roose et al. 2020).

Granted, many movements can be dangerous and incite terrorism, but why this thesis focuses on the manosphere and 'far-Right' communities is because of male supremacy's underrepresentation in correlation to Islamic terrorism (Parker et al. 2018) in research as well as with implemented preventative methods although the threats of online 'far-Right' and communities in the manosphere are rising (Beckett 2021). In two interviews I conducted with prominent Danish researchers in this field, Maia Kahlke Lorentzen from Cybernauterne and Telli Betül Karacan, a researcher in Islamic and Right-wing terrorism, my interviewees expressed their frustration with the lack of focus, research grants, and (acknowledgement of) expert knowledge on this area. Even though there are research centres in Denmark, some also focusing on online and male supremacy

extremism, like NCPE, National Centre for Prevention of Extremism, which recently released a report on online juvenile extremism (Stender Petersen and Alberg Peters 2020), there is still a need for continuous research, and on preventative measures from online radicalisation:

"Research on extremism is far more expanded than research about prevention of extremism. Where there in recent years has been built a solid research tradition on extremism in general, and partially on prevention, the field on online extremism and the ways into extremism through the internet is still young. The research on the prevention of online extremism continues to lack a theoretical clarity and empirical data, and this is despite online extremism having gradually been incorporated as an important element in both international (EU), national, regional and local prevention initiatives and action plans"[5] (Stender Petersen and Alberg Perts 2020:5)

With this still being a new field, especially in a Danish setting, this report aims to focus on and collect solution-based ideas and strategies to make social media platforms safer. The study will combine the existing research and knowledge, and exploratively analyse and discuss solutions and preventative methods to deal with the rising problem of radicalisation and terrorism with roots to the online manosphere.

---

[5] Translated from Danish

## 1.3 Problem statement

Based on the problem field addressed above, the focus of this research ascribes to finding out:

*Does the YouTube algorithm radicalize in anti-feminist ways,*

*and which preventative solutions for online radicalisation can be made in Denmark?*

To answer this problem statement, further research questions will be used:

*What is the 'manosphere'?*

*How can radicalisation, terrorism and anti-feminism be theorised?*

*How does the YouTube recommendation algorithm work?*

# 2 Literature review

For this thesis, I have used both technical papers to learn about the YouTube algorithm, and academic research reports to collect knowledge on online radicalisation and possible solutions and tools to prevent it. This section will outline the most important papers used as a basis for this report. I have gained a lot of information on the topic from podcasts, news articles, personal experiences of algorithmic bias from both myself, people I know and other users of social media platforms, along with governmental and institutional reports, peer-reviewed articles, and academic books. Early in the process, I conducted two expert interviews, and through these included explorative interview questions (Lynggaard 2015:158), leading me to relevant literature, and "monumental" documents (Andersen 1999) also called a "mother document", which are often cross-referenced by other sources in my literature search. Mother documents used in this thesis are amongst others Google's paper about the YouTube algorithm (Covington, Adams and Sargin 2016), and a Danish desk research on juvenile online radicalisation (Stender Petersen and Albert Pers 2020). Asking my interviewees exploratory questions 'mapped' the field out and introduced several important actors within the problem field. To research literature, a method known as the "snowball method" has also been used, which is described as following internal references between documents (Lynggaard 2015:157). At the end of the literature review, papers from the explorative interview questions, the snowball method, and search inquiries gave few new insights correlating to my problem field. In total, 74 sources have been included in this thesis which have given insights about the YouTube algorithm, the theories and knowledge on radicalisation, extremism and terrorism, the research on preventative methods and current governmental strategies, and other important aspects correlated to the problem field.

## 2.1 The YouTube Algorithm

Learning about the YouTube algorithm has not been an accessible task as a researcher. I have used a paper provided by Google itself, written by Paul Covington, Jay Adams and Emre Sargin (2016), which is a mother document on the technicalities of the YouTube algorithm. The paper is technical, but not very explicit or reflexive about the effects of the algorithm, nor transparent enough to provide knowledge on possible changes in the machine learning algorithm. Since then, it is also to be expected that the algorithm used by YouTube has changed both by machine

learning and through human interaction to optimise it further. The Google research paper though creates a basis for how the algorithmic model works or has been created to work originally. *The YouTube Algorithm and the Alt-Right Filter Bubble* (Bryant 2020) is a paper that analyses the Google paper and criticises its lack of transparency. Bryant also analyses and uses quanti-qualitative research which has been done on the YouTube algorithm. Data from the website AlgoTransparency.com, made by a former YouTube algorithm coder Guillaume Chaslot, runs daily tests on YouTube, scraping to see which videos are recommended from 800+ top information channels, and provides knowledge on radical content and users of YouTube, and how the recommendation algorithm works, and which videos and channels it recommends from popular channels. A research project by Kaiser and Rauchfleisch (2018) for The Media Manipulation Initiative at Data & Society have made a paper on the 'far-Right' communities within German and US channels on YouTube and have visually mapped these out with the software program Louvaine. A recently published Danish report sponsored by PET, the Danish Security and Intelligence Service titled *Under Influence: Ways into Extremist Digital Communities through Gender and Masculinity*[6] (DareGender and Cybernauterne 2020) has mapped out communities in the online manosphere, as well as collected scrapings of 12 million comments from YouTube channels and other platforms, and interviewed several people affected by the *rabbit holes* of YouTube and felt radicalised by its content.

## 2.2 Online Oppression

To learn about online oppression, podcasts, news articles, academic journals, and books, like *Online Othering* (Lumsden and Harmer 2019) and *Algorithms of Oppression* (Noble 2018) focus on unfairness on social online platforms and provide knowledge how algorithms, community guidelines and (a lack of) moderation unintentionally can cause oppression and discrimination on social media platforms. Podcasts and research reports have provided qualitative knowledge through interviews with people who have been radicalised by the manosphere and have given insights into how YouTube works as a platform, and how users are affected by the content on the platform. Feminist theory has also been used to find frameworks for this project, shaping the methodology and theory of this project, as well as being used to define anti-feminism, and other

---

[6] Translated from Danish "Under Indflydelse: Veje ind i ekstreme digitale fællesskaber gennem køn og maskulinitet". Since the publication is in Danish, quotes used throughout my report are translated

oppressive structures happening in the online manosphere, mainstream media, and general society.



| Type | Name | Author | Status | Tags | Publisher | Score /5 |
|---|---|---|---|---|---|---|
| Podcast | 📞 Big Sister Hotline: S2E1 Feat. Laura Bates | Clementine Ford | Finished | MRA | | ⭐⭐⭐ |
| Article | 🇺🇸 'It let white supremacists organize': the toxic legacy of Facebook's Groups | Kari Paul | Finished | Facebook  Online cultures | The Guardian | ⭐⭐⭐ |
| Report | 🖥 Strategi for Danmarks Teknologiske Diplomati 2021-2023 | Udenrisministeriet | Finished | Technology  Strategy  Politic | Udenrigsminis | ⭐⭐ |
| Report | 🪁 Reframing Islamic State - Trends and themes in contemporary messaging | Telli Betül Karacan | Finished | ISIS  Radicalisation | DIIS | ⭐⭐⭐ |
| Report | 🐞 Under indflydelse: Ny rapport om ekstreme digitale fællesskaber | Cybernauterne | Finished | Radicalisation  Gender  YouT | | ⭐⭐⭐⭐⭐ |
| Podcast | 👥 Under Indflydelse 1: Radikalisering på YouTube | Cybernauterne | Finished | MRA  YouTube  Algorithms | | ⭐⭐⭐ |
| Academic Jou | ♀ Rethinking Standport Epistemology: What is "Strong Objectivity"? | Sandra Harding | Finished | Feminism  Theory  Objectivi | Michigan Stat | ⭐⭐⭐ |
| Article | 📍 The Vignette: A bad ethnographic category | eli thorkelson | Finished | Theory  Objectivity | | ⭐⭐⭐ |
| Podcast | 🐰 Rabbit Hole series | Kevin Roose | Finished | Algorithms  YouTube | New York Tim | ⭐⭐⭐ |
| Academic Jou | 🖊 Incels, kvindehad og den nye antifeminisme | Telli Betül Karacan | Finished | Radicalisation  Online cultures | | ⭐⭐⭐⭐⭐ |
| Article | ⛔ Debat: Deplatforming virker, men udfordrer demokratiet og ytringsfrihede | Telli Betül Karacan | Finished | Online cultures  Facebook  S | Jyllands Poste | ⭐⭐⭐⭐⭐ |
| Academic Jou | 🖥 The YouTube Algorithm and the Alt-Right Filter Bubble | Lauren Valentino B | Finished | Algorithms  YouTube  Radica | | ⭐⭐⭐⭐⭐ |
| Event | 📣 TalkTown - Anti-feminisme | Mikkel Thorup og | Listening | Anti-feminism  MRA | TalkTown | |
| Podcast | 👥 Under Indflydelse 2: Ned i Kaninhullet | Cybernauterne | Listening | MRA | | |
| Book | 🍎 Sad By Design | Geert Lovink | Reading | Design  Online cultures | | |
| Book | 👥 Kill All Normies - Online Culture Wars from 4chan and Tumblr to Trump ar | Angela Nagle | Reading | Online cultures | Zero Books | ( ? ) |

Screenshot of my study list in the process of reviewing the literature in the program Notion.

## 2.3 National Knowledge and Measures

To provide knowledge for the existing material on national research, plans and prevention strategies, I have used a strategic report plan for Denmark's technological diplomacy between 2021-2023 from the Office of Denmark's Tech Ambassador (Udenrigsministeriet 2021) and a "mother document" *Mapping Knowledge on Prevention of Online Juvenile Extremism*[7] by the Danish National Centre for Prevention of Extremism[8] (hereafter NCPE), along with a report on democracy and digital technologies, made by a technological commission (TechDK Kommissionen) set up by Djøf. These have given insights into the current strategies and knowledge within a Danish setting and government. The NCPE report provides extensive desk research on different issues within online extremist cultures and currently known and researched prevention strategies, whilst the TechDK report suggests law changes and other preventative measures to provide democratic values to machine learning algorithms and digital culture. The strategy report from Udenrigsministeriet (The Ministry of Foreign Affairs) provides a plan and

---

[7] Translated from Danish: Desk Research: Kortlægning af viden om forebyggelse af ekstremisme online blandt børn og unge

[8] Translated from Danish: Nationalt Center for Forebyggelse af Ekstremisme

maps current concerns on how to handle different elements and issues in online spaces and on a legislative basis. Here, the report describes wanting to "hold tech companies accountable for their societal responsibility" (Udenrigsministeriet 2021:2) and problematizes that the infrastructure of algorithms is largely unknown, nontransparent and enforces online "echo chambers" (Udenrigsministeriet 2021:4). The report also mentions several objectives planned to be worked towards within Danish context (Udenrigsministeriet 2021:7-9). Some of these are working to ensure that human rights and minorities are protected online, that Denmark and the EU have a voice in the international collaboration about digital rules and regulations, that international network that monitors violence and potential genocide encouraged in digital spaces, and a better handling of cyberthreats through strengthened collaboration between governments and tech companies

## 2.4 Radicalisation, Extremism and Terrorism

To define the terms "radicalisation" and "extremism" both of my expert interviews and academic journals point to there being an overrepresentation and focus on Islamic radicalisation and terror, and that white, misogynist, and anti-feminist violence is strongly overlooked. Research on non-Islamic terror and online terrorism in a Danish setting has only recently been initiated over the past few years. A study within a Danish and UK context has analysed how the press covers the so-called *lone-actor terrorism* between 2010-2015 (Parker et al. 2018). The terminology of lone-actors or lone-wolves is admittedly problematic as it potentially glamourises attackers and inaccurately frames them as acting on their own without connections to others and being independent (Parker et al. 2018:111). The Danish data set in the study are drawn from the national daily tabloid newspapers and broadsheet papers from, amongst others, BT, Berlingske, Jyllands-Posten, Kristeligt Dagblad and Politiken (Parker et al. 2018:14). This study concludes that the emerging "lone-actor" terrorism is a primary, if not *the* primary, security threat facing many Western states, but also that this type of terrorists is often framed as mentally ill or evil in comparison to Islamist extremists being criminal and act as a violent sub-set of a community (Parker et al. 2018:125). A similar study conducted today would be important as the past six years have passed Still, it gives a picture of how *terrorism* is normally seen and understood. In this report's section *Defining Radicalisation and Terrorism*, two NCPE reports will be used to outline radicalisation and extremism theoretically.

## 2.5 Reflections on Literature Review

In conclusion, existing literature extends to many different disciplines of analysis, some technical, some sociological, and some strategic. All existing knowledge adds to this problem field, but from the research found through this literature review, there seems still to lack an interdisciplinary approach, combining a multi-faceted scope considering how the technological factors must be dealt with to prevent radicalisation in online spaces, as well Danish research on this. There is still need for further research on the machine learning algorithm used on YouTube, which does not seem to be accessible for researchers as of now. Existing literature also point to there being a need to monitor and research the effects of preventative measures, but this depends on measures being implemented first. Especially on the legislative and governmental level, literature and knowledge on specific strategies and plans are difficult to gain insights in, although these may exist outside the public eye.

# 3 Methodology and Theory

In this section, the methodology and theoretical frameworks used in this thesis will be presented and discussed. First, the methods used to conduct two expert interviews will be described, as well as presenting the two interviewees used in this thesis. Afterwards, *feminist standpoint epistemology* and other feminist theories used as a framework will be presented. This section will also describe 'objectivity' which in *Discussions* at the end of the report will also be reflected upon. Furthermore, radicalisation, terrorism, and anti-feminism will be theorised as of how it is used throughout this report.

## 3.1 Interviews

For the interviews, semi-structured interview guides have been used, which helps focusing on specific questions which I wanted to talk with my interviewees about. At the same time, this method lets the conversations flow lead to other aspects which were not planned as it leaves 'room' to the interviewee and conversation. The interviews both lasted approximately 50 minutes and were recorded with the consent of both interviewees. Both interviews were held in Danish, meaning that the quotes used in the report have been translated. Original transcriptions can be found in the appendix. Afterwards, I have manually transcribed both interviews, and during transcription thematically coded them in brackets. The codes are used to "map" the transcriptions so I could more easily navigate them afterwards. The codes can be used to quickly find the different answers which have been important to the research question and focus. Codes are words such as "deplatforming", "responsibility", "legislation", and "masculinity". Both transcriptions were sent to be approved by the interviewees afterwards, to make sure that all data could be used in my report.

### 3.1.1 Feminist Narrative Interview Method

Adrianna Kezar's (2003) paper on feminist narrative interviews works as inspiration for my interview method. Although the approach in Kezar's paper is developed to address power structures in interviews with non-feminist or non-'critical' elites to transform and create change, and while those people I interview are not in power positions (per se) or responsible for big

companies, the paper still provides relevant reflection and methodology. Kezar describes how research and interviews without the focus on transformation can be a lost opportunity to create change and to "break down oppressive systems within our society" (Kezar 2003:396). Kezar writes, with reflection of many scholars on narrative inquiries, and say the method aims to

"(...)view stories as representing ideology and meaning making, see fieldwork and interviewing as inherently collaborative and relational, identify human experience as constructed through subjective and intersubjective interpretation and stories as a primary way to understand these constructions, deny a universality of experience or reality, believe that multiple narratives exist and can both be in conflict and reflect truth, and place narratives within a sociological context illustrating that stories reflect history as well as create history" (Kezar 2003:400).

The stories of the two interviewees that I have chosen to include in my research are, naturally, not representative of everyone, but both interviews represent unique stories of being a feminised researcher, like myself, and working in a field that is heavily on target for not being objective because of feminist or political ideologies. Kezar also describes that a researcher should strive to create and have (a) commitment and engagement, (b) mutual trust, (c) reflexivity, (d) mutuality, (e) egalitarianism, (f) empathy and ethic of care, and (g) transformation through consciousness-raising, advocacy, and demystification (Kezar 2003:400). In practice this means that I make sure that the narratives presented in this report are checked by my interviewees, and that my goals, thoughts, criticism, and ideas are presented and shared throughout the interviews. At the same time, this allows reflective and analytical interviews where both interviewer and interviewee share experiences and thoughts.

## 3.1.2 Interviewees

I wanted to talk to two researchers in a Danish setting to map out the national research being done in this field to gain knowledge on the subject, and I knew both these researchers from my literature review and my prior interest in the topic. The first person I interviewed is Telli Betül Karacan who works at the Danish Institute for International Studies (hereafter DIIS) as a research assistant. She has exceeding knowledge in radicalisation and terrorism and is currently working on a five-year research project called "Explaining Transnational Jihad - Patterns of Escalation and Containment". Furthermore, she also researches Right-wing cultures and how they unfold online

and is part of a three-year research project called "World of the Right", and has written academic journals like *Incels, Misogyny, and the new Anti-feminism*[9] (Karacan and Crone 2020), and a recent article in the Danish news website Jyllands Posten titled *Deplatforming works, but challenges democracy and freedom of speech*[10] (Karacan 2021). With her knowledge of the digital 'far-Right' and anti-feminism and research of terrorism and radicalisation, she fits well into my research.

The second interviewee is Maia Kahlke Lorentzen, a specialist, trainer, and consultant working with digital literacy, online activism, and online right-wing cultures. She has written a book and done a TEDx Talk on online trolling cultures, is part of Cybernauterne, and provides research, podcast, education, and information on the manosphere and violent anti-feminist online communities. Recently, she has facilitated online talks on anti-feminism and the manosphere with the Danish feminist debating event TalkTown. She has also been part of the research resulting in the report from DareGender and Cybernauterne (2020), used throughout this thesis.

Both Karacan and Kahlke Lorentzen have done interviews of radicalised individuals, and therefore also act as sources of qualitative knowledge in my thesis.

## 3.2 Objectivity and Standpoint

Both in and outside academia, conversations on objectivity and feminism have been charged and are oftentimes showing the hypocrisies around objectivity and good science and how this is perceived differently with different researchers, most recently in the Danish Parliament where the topic of activism in certain academic circles and research are discussed[11]. Sandra Harding, philosopher of feminist theory, epistemology, and research methodology, writes in her paper *Rethinking Standpoint Epistemology: What Is "Strong Objectivity"?* (1992) what she believes good research to be. Due to the notion from positivism about "objective" research and researchers, white men often have had and have the privilege of being assumed to be in a "neutral" position when doing research. As Donna Haraway writes (1988), these researchers are perceived *unmarked* and "claim the power to see and not be seen, to represent while escaping representation"

---

[9] Translated from Danish: "Incels, kvindehad og den nye antifeminisme"

[10] Translated from Danish: "Deplatforming virker, men udfordrer demokratiet og ytringsfriheden"

[11] Proposition F49 from June 2021

(Haraway 1988:581). In comparison, *marked bodies* can be defined as those who are not invisible; the women, queers, femmes, BIPOC[12] and others deemed *marked* by the colonial and patriarchal institution that is academia. Harding criticizes the notion of unmarked bodies being able to make better research and instead argues that dominant groups fail to "critically and systematically interrogate their advantaged social situation and the effect of such advantages on their beliefs [which] leaves their social situation a scientifically and epistemologically disadvantaged one for generating knowledge" (Harding 1992:442). This thesis therefore aims to speak from my own perspective, and others', often targeted on social media and from ideologies, rhetoric, and opinions existing in the social media sphere.

## 3.2.1 Feminist Standpoint Epistemology and Minority Taxation

Feminist standpoint epistemology holds that knowledge is socially situated, and that researchers must be self-reflective and interrogative around their own privileges and biases in order to produce better knowledge. Sandra Harding argues that knowledge production should "start from marginalised lives" when scientific "problems" are identified and hypotheses conjured, as a method to maximise value-neutral objectivity (Harding 1992:462). Still, Harding argues that the definition of objectivity should change to properly reflect bias in research production. In her paper, she also criticises academia's practices in terms of excluding marginalised groups, voices, and researchers. Haraway (1988) puts a perspective on Harding's paper with a critique of "the politics of positioning, science situating itself at the hierarchy of all knowledge-claims, claiming a God's eye-view, an unaccountable, disembodied gazing." (Yadav 2018). So, I would argue that marginalised voices and researchers can provide insights otherwise overlooked or hidden, although these are still *situational*, meaning not universal or absolute.

Indian-Danish-US-American researcher Tess Skadegård Thorsen, a PhD in Social Sciences, specialises in representation, oppression, discrimination, and unfair structures. She coined the term *minority taxation* inspired by the US term Cultural Taxation by Padilla (1994). Minority taxation is a tool that visualises "both the concrete and affective extra-work minoritized academics are 'taxed with' performing due to structural inequalities" (Skadegård Thorsen 2019:31).

---

[12] Black, Indigenous and People of Colour

**Minority taxation**
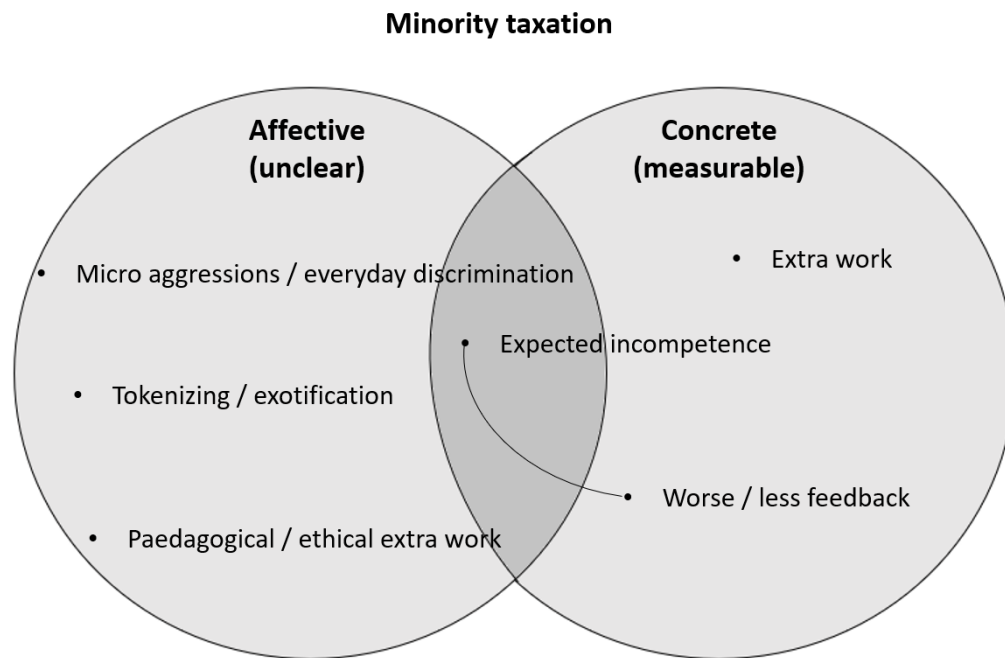


Figure 1: Illustration by Tess S. Skadegård Thorsen (2018) translated by me.

In this figure some of the experiences that a minority student or worker in academia has been listed. She stresses that this list is incomplete as there can be several other experiences as well (Skadegård Thorsen 2019:36). Affective experiences can be hard to "measure" as they often cost emotional labour. I am both Iranian-Danish, queer and non-binary which marginalises and alienates me in certain settings, and which then also *taxes* me. I often experience the world from a different position than my peers because my voice, experiences, and the issues I raise and see are not familiar to others. For example, being non-binary, meaning that I identify neither as man or woman and use they/them pronouns instead of he/him and she/her, I frequently experience micro-aggressions and discrimination. This includes the risk of getting misgendered in university settings or not being presented gender diverse sources beyond the binary men/women (sometimes not even beyond men) in lectures. These experiences tax in unclear ways as it presents choices of either saying something, causing annoyance, disagreement, or dismissal or I can "complicity" choose to stay silent and accept the cissexist[13] norm of the world, but in any case, cause me emotional unrest and disruption of my work and wellbeing. As for being Iranian-Danish, I hear, see, and feel racism in ways that my white peers and professors do not. This can mean having to "debate" racism with the white peers at the university. As I research from a feminist

---

[13] The normative idea that everyone is cis gendered (identifying with their assigned gender) and binary

standpoint, and because I am "othered" by the dominant groups in academia, Skadegård Thorsen gives a tool to show how these experiences can cause *taxations* in measurable ways, as shown in Figure 1. Often it is necessary for me to find literature that has not been presented to me during my education and, as said, I have not had the luxury to specialise myself in areas I would have wished to. Oftentimes, this has also led group members or supervisors to be sceptical or critical of my work, theories, and position beforehand, which then causes me to get worse and/or less valuable feedback. All these taxations make my position and work in academia more exhausting and intensive. But, as Harding argues, my position also gives me experiences to identify different issues and problems needing to be examined and researched, which is why I have chosen to write about this for my master's thesis. Knowing first-hand from friends, myself and my community being harassed, at risk of getting banned or having to invent code words to communicate and exist on social media, whilst seeing sexism, racism, transphobia, violence, and the likes exist freely on the same platforms is, for me, a reason to be invested. These tools also give a framework for understanding how certain problems are overlooked or subsided within an academic or societal setting. In this thesis, I do not seek to define whether the online manosphere is dangerous, or a problem, but navigate this problem field from a point of view, where it *is* an issue needing a solution. I therefore map out this thesis how and why I see this being an issue, to find appropriate preventative suggestions which can be implemented.

## 3.3 Defining Radicalisation and Terrorism

The terms 'radicalisation' and 'terror' can have many different meanings and definitions. In this section, I will outline how different actors define extremism, radicalisation, and terror.

In a Danish setting, the Danish research centre NCPE, which works under the Immigration and Integration Ministry defines radicalisation as a "shorter or longer process where a person aligns with extremist views or legitimises their actions by extremist ideologies" (Nationalt Center for Forebyggelse af Ekstremisme 2020). NCPE defines extremism as often characterised with (Nationalt Center for Forebyggelse af Ekstremisme 2020):

- Having a lack of respect for other people's freedom and rights
- Having a lack of respect for institutions and decision processes in the representative democracy

- Having simplified world views and demonisations where certain groups or societal circumstances are seen as threatening
- Isolation and focus on a (perceived) opposition between 'us' and 'them'

In the interview with Kahlke Lorentzen, she says that Maja Touzari Greenwood, a researcher in DIIS, counselled their project (DareGender and Cybernauterne 2020) and describes extremism as having two frameworks: using violence as a method, and being anti-democratic. Kahlke Lorentzen underlines (Kahlke Lorentzen in interview with author, March 2021:5):

| Kahlke Lorentzen: | And then it's about what you define as anti-democratic. That's where the question on the definition is right? |
|---|---|

The NCPE underlines that individuals with extremist opinions and ideologies do not necessarily participate in violence, and vice versa: individuals without extremist opinions and ideologies can participate in violence (Stender Petersen and Albert Pers 2020:19), but normalising violence, hatred, and dehumanisation are themselves also a way to condone violence, even without doing it personally, something which is frequently found in the online manosphere, which will be described later.

### 3.3.1 Theorising Radicalisation and Extremism

Within feminist communities there is a common saying around inclusivity and exclusion that if someone's presence excludes people based on their identity, they are not welcome, one example being if someone allows harassers in spaces, other people will be unsafe being there. Dangerous anti-democratic values can be some that oppress and disallow certain (groups of) people from having safe, healthy, and fair opportunities and conditions within society. Racist discourse then is dangerously anti-democratic as it is unsafe, unhealthy, and unfair and worsens pre-existing racist societal structures which hinder and endanger people of colour.

Radicalisation in other groups can also stem from fighting oppression, like opposing institutional decision-making processes because of feeling unheard, as well as isolation from majority groups to escape oppression (for example LGBT spaces). Karacan says in our interview, not all radicalisation is bad (Karacan in interview with author, February 2021:2):

| Karacan: | Radicalisation isn't necessarily a bad thing - for example, there is a difference when we talk radicalisation correlating to Islamic State [ISIS eds.], or radicalisation in correlation to the Women's Movement, for example. Women wouldn't be able to vote, if it wasn't for the women being radical in their time and questioning some of the structures there are. |
|---|---|

It is worth noting though, that there within both 'far-Left' and feminist communities are individuals and groups, who do have dangerous anti-democratic values such as TERFs (Trans Exclusive Radical Feminists), who are transphobic, and pick up on anti-trans rhetoric and ideologies from the manosphere, but still consider themselves as feminists (and/or Leftists).

### 3.3.2 Who is Prone to Radicalisation?

The NCPE report, which has desk-researched 611 publications, and included 39 in its own report, shows that both well-functioning and dysfunctional children and youths can get involved in extremist online cultures (Stender Petersen and Albert Pers 2020:14-15). Within academic research on who is more prone to get radicalised, there are no absolute answers. Karacan expresses in her interview (Karacan in interview with author, February 2021:2):

| Karacan: | There has been a lot of research about who's prone to get radicalised, and I don't think it's something you can say because it's very situational. It's not like a box you can put people into and then tick off what it is they live up to in requirements for becoming radicalised |
|---|---|

There are some factors that are more frequently found in individuals who have extremist opinions or ideologies, and who have found ways into radical communities. 'Confusion and humiliation' seem to be a driving force for some individuals (Richards 2019:13). Feelings of humiliation can stem from different and complex personal experiences such as marginalisation, oppression, microaggressions[14] (as explained by Karacan in interview with author, February 2021:9), or, as seen frequently in the manosphere, feelings of emasculation (DareGender and Cybernauterne 2020).

---

[14] Common, everyday experiences where marginalised individuals get othered or discriminated against, although these actions might not be explicit, obvious, or done with a bad intent. Examples can be being stared at in public, asked questions like "Where are you really from?" or told "You speak Danish so well!" because of having a different skin colour than the dominant population

Feelings of confusion can stem from being fearful about the state and futures of certain societies (Richards 2019:14). Other examples can be being presented with simple narratives, especially those which feel convincing and trustworthy (as with disinformation[15]), and especially if these live up to already inherent biases in individuals (Richards 2019:14). Finding online communities to engage in where feelings are mirrored, can then be a problem, if these communities at the same time inhabit values which are dangerous anti-democratic.

### 3.3.3 Online Radicalisation

Online radicalisation can unfold in different ways than offline extremism, but the NCPE report says that it often intertwines with offline radical communities and individuals (Stender Petersen and Albert Pers 2020). Whilst deradicalisation and preventative initiatives in Denmark like what is known as the "Aarhus-model" have been built with a focus on Islamic radicalisation, it is unsure to which extent knowledge of specifically online 'far-Right' and extreme communities in the manosphere exist as of now[16]. 'Edgy' meme culture formed in online social media can seem insignificant or like regular meme culture if the audience does not know otherwise. An example is the famous "Trad Wife" character, which got famous in mainstream social media around November 2020 following a viral tweet[17].
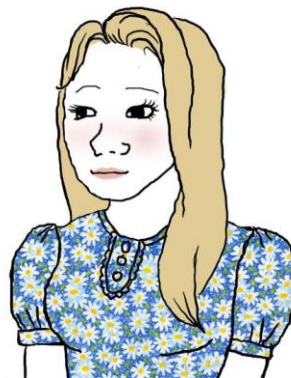


"Trad wife" meme format

---

[15] Misinformation which is intentionally spread (Mehlsen & Hendricks 2019 in Stender Petersen and Albert Pers 2020:15)

[16] In the duration of this thesis, I contacted Aarhus Municipality's office for efforts against radicalisation, the Office of Denmark's Tech Ambassador, and the Ministry of Justice of Denmark to ask for further insights, but unfortunately with no replies

[17] From "Know Your Meme". Accessed May 21, 2021: https://knowyourmeme.com/memes/thank-you-for-changing-my-life#fn1.

Seemingly, the drawing of the girl in the blue dress is merely a meme-format like so many others, but its roots originated from 4chan's imageboard in 2019. Here this blue-eyed, blonde-haired girl represented a traditional wife: a real conservative woman who celebrates traditionally feminine roles, following their husbands, and being white (Masilis 2020). The idea of the "traditional" wife, is often circled around in incel-communities and the manosphere, claiming that there does not exist any "good women" anymore, making it impossible for men in incel-communities to get proper girlfriends.

The "trad wife" is also often used in the manosphere as a comparison for the modern woman, the feminist, where the feminist is mocked, and the traditional wife celebrated (See picture below). Here, the "trad wife" holds values such as staying at home and raising kids, while her husband works, being "knowledgeable about her European roots", and loving her "family, race, and country in that order". The comparison with the "liberated" feminist uphold both sexist, biphobic[18], fatphobic and racist ideas, which is an example of how incel-culture intertwines with racism and anti-feminism in online communities. In an Alma article about the "trad wife" meme, Vi Maislis writes: "According to a report done by the Data & Society Research Institute (Marwick and Lewis 2017), "Far-right movements exploit young men's rebellion and dislike of 'political correctness' to spread white supremacist thought, Islamophobia, and misogyny through irony and knowledge of internet culture." Boards like "politically incorrect," better known as /pol/, are filled with memes like "trad wife" — images that mask violent Nazi ideas, but are passed around as ironic and edgy" (Maislis 2020).

[18] Discrimination against bisexual individuals, often in denial of their sexuality being real

The "Liberated" Feminist
The Tradwife

Wears tons of makeup because of her low self esteem

Damaged hair from over dying

Loves her natural face and only wears light makeup

Natural hair with highlights from the sun

Got a crappy flower tatoo when she was 17

Uses fake tanner so dark she looks mexican

Natural tan from her outdoor lifestyle

Sleeps around to improve her self esteem but it only makes her feel worse

Her husband works to support her staying at home and raising the kids

Wears modest and feminine clothing

Wears clothing that barely fits her

Slim figure from her healthy homemade meals and active lifestyle

Chubby from her diet of fast food and microwave meals

In her 20s but is already married with kids

Aborted her black baby last year

Is "only attracted to black men"

Homeschools her kids so they aren't taught progress bs in school

Loves her family, race, and country in that order

Claims she's bisexual but has only dated men

Works a minimum wage job because her gender studies degree didn't get her any where.

Knowledgable about her European roots

The "Liberated" Feminist vs. The Tradwife[19]

The meme format of the "trad wife" combined with other characters in a similar drawing style spread out on mainstream platforms and got popularised outside the manosphere and 'alt-Right' circles. Lindsay Schubiner, program director at the Western States Center, a non-profit focused on social, economic, racial, and environmental justice, told The Washington Post that "White-nationalist and alt-right groups use jokes and memes as a way to normalize bigotry while still maintaining plausible deniability, and it works very well as a recruitment strategy for young people" (Gibson 2019). Using these images makes it easy to puff them off as "just a joke", allowing ideologies and harmful rhetoric to be safely hidden with plausible deniability (Maislis 2020). This strategy is also called a *dog whistle*.

---

[19] From Reddit on /r/tradwife. Accessed May 21, 2021.:
https://www.reddit.com/r/tradwife/comments/djhs2q/liberated_feminist_vs_tradwife/

A dog whistle is coded or suggestive language, which appears normal to most people, and is made to not attract negative attention, but still, provoke controversies and may be hard to prove false. One example is the "war on terror", which is not racially coded per se, but in practice has meant the stigmatising and warfare of Muslims and Muslim majority countries. One must critically assess or have the knowledge of the meanings behind dog whistles to understand their propaganda or intention. It can be opaque to 'outsiders' when rhetoric, memes, pictures, or coded language are used online and contain harmful messages, and such knowledge has to be used to fight online radicalisation and the communities with extremist tendencies. Most moderation on social media fails to flag, or recognise harmful messages when reported on the sites, allowing discrimination to exist (Lumsden and Harmer 2019; Mooghen and Bau Larsen 2019)

## 3.4 Defining Anti-feminism

Feminism is broad and vast in size, which can both be deeply personal and individual, as well as some defined ground values and political goals. At times, the feminist movement had political fights in Europe and the US for rights such as voting, owning property, equal pay, and more. Though feminism is much more than that, and feminism which includes a critical perspective and inclusion of issues on behalf of various identity markers is often referred to as *intersectional*. Kimberlé Crenshaw coined the term *intersectionality* in her paper *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics* (1989) as a tool to address the intersection of being Black and being a woman. From her background as a law professor, she saw a gap in antidiscrimination law either insisting Black women sue discrimination based either on their gender or their race (Crenshaw 2016). Therefore, when feminism is used in this report, it speaks of the many different oppressions and political fights for justice and rights. Some of these are struggles based on class, sexuality, gender, ability, fatness, immigration, sex work, mental health, houselessness and much more, and how these intersect and affect people differently.

In recent years, the anti-feminist discourse has been found in online communities describing feminists as "Social Justice Warriors" (SJW), ridiculing feminists as having unreasonable demands, and even diminishing the reality of these issues, claiming "SJW" are seeking oppression and victimising, and in Danish context wishing to get offended (krænkelsesparat, krænkelseskultur). On the current online anti-feminist wave, Angela Nagle writes:

"Unlike the anti-feminist tendency in organised men's movements of the 1990s, the anti-feminist online counterculture today does not identify with pro-market or conservative ideologies regarding women and at least on an aesthetic level, appears closer to libertarian left wing tendencies. It is not typically interested in political struggles around rights or equality, but primarily around culture. It is more a style of cultural politics than a political movement and many of its battles are not linked to demands or articulated political goals. In this sense, one could argue it is the mirror image of post-feminism's retreat into culture." (Nagle 2015:28-29).

A Swedish study (Sager and Mulinari 2018) also shows how anti-feminism and racism intertwines, and how both traditional gender roles, and family constellations are used as anti-feminist arguments. In politics, xenophobic, right-wing, anti-feminist party, the Sweden Democrats (SD) uses "an ideology of hate (against migrants, against feminists, against homosexuals, against left activists in solidarity with refugees) through representing the party as providing *trygghet* (safety)", leaving feminist values painted as a threat to a safe nation (Sager and Mulinari 2018:152). Anti-feminism also has links with Right-wing extremism and terrorism, an example seen in Anders Breivik's manifesto with sections as "The name of the devil: cultural Marxism, multiculturalism, globalism, feminism (...) - a recipe for disaster" (Berwick (Breivik) 2011:35) and "How the Feminists' "War against Boys" Paved the Way for Islam" (Berwick (Breivik) 2011:351). Although distinguishable, anti-feminist ideologies and misogyny lie at bay with a lot of the Right-wing communities we know of today, something which seems commonly overlooked.

Anti-feminist ideologies in Anders Breivik's Manifesto's table of content (Berwick (Breivik) 2011:4)

### 3.4.1 Anti-Feminism in Media

Danish media discourse also contains anti-feminist discourse, and have run media articles such as "We are offended like never before, but why have we gotten so thin-skinned? Here is a possible explanation"[20] (Lindberg 2019), and "High School students defy offence culture: Are you "krænkelsesparat"? Then you probably shouldn't attend"[21] (Lund 2019). Similar to "krænkelsesparat" or "SJW" is "special snowflakes", which stems from online trolling culture (Nagle 2017:45) claiming that feminists and "PC"-culture (politically correct) are easily offended and sensitive. Why this is interesting for this thesis, is the one-dimensional and ridiculing idea of feminism, has seeped from online trolling culture to mainstream media and public discourse speaking into an "us-versus-them" mindset. This rhetoric also works uses dog whistles to disguise anti-feminist agendas.

Feminists are seen as a monolith, and, through anti-feminist rhetoric gets ridiculed, and feminist and social justice ideas, and marginalised experiences are diminished. Anti-feminism, in the way defined in this thesis, speaks of communities and individuals who in one way or another frame feminism as something threatening, oppressive, or unimportant. As DareGender and Cybernauterne state in their report:

> "Traditional gender roles, a resistance against feminism, and the feeling of being oppressed or marginalised are central elements in these communities. (...) Feminists and equality advocates are demonised as a threat against the freedom of speech and of men generally."[22] (DareGender and Cybernauterne 2020:2)

Sometimes anti-feminist propaganda also claims that feminists have power and form a secret world order (DareGender and Cybernauterne 2020:16), or generally have societal power, are dangerous, and want to dissolve gender, something which is an important aspect of the communities in the manosphere (DareGender and Cybernauterne 2020:15-16).

---

[20] Translated from Danish: "Vi krænkes som aldrig før, men hvorfor blev vi så sarte? Her er et bud på en forklaring"

[21] Translated from Danish: "Gymnasielever trodser krænkelseskultur: - Er du krænkelsesparat? Så skal du nok ikke møde op"

[22] Translated from Danish

# 4 How YouTube can be a Breeding Ground for Radicalisation

In order to find out how anti-feminist communities exist and whether YouTube's algorithm can radicalise, how the YouTube algorithm was developed and changed into the strong, intelligent algorithmic system that it is today will be explored. I will analyse studies which have researched the functionality of the YouTube algorithm, and whether a bias can be seen in the machine learning algorithmic recommendation sidebar on the site. I will also go over YouTube's community guidelines, rules and how YouTube deals with content moderation on its side to find out how content is moderated.

## 4.1 Defining Algorithms

"The measure of success for the YouTube algorithm is convincing the user to watch an additional video after the end of the first video has finished." (Bryant 2020)

An algorithm is "a mathematical or logical procedure for solving a problem [and] is a recipe for finding the right answer to a difficult problem by breaking down the problem into simple steps" (Domingos 2015:1), which is embedded in every input a computer is given. A new branch of algorithms has been developed which improve on their own, often using pattern recognition to determine a better or faster way to attain the goal(s) originally set by the human engineers which created the coding (Bryant 2020:87). Domingos writes in *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* that "[t]hese algorithms are not programmed to solve particular problems. Instead, they are programmed to learn to solve problems." (Domingos 2015:85). This type of algorithms is called *machine learning*, and the process of how they work can be described as such:

"Every algorithm has an input and an output: the data goes into the computer, the algorithm does what it will with it, and out comes the result. Machine learning turns this around: in goes the data and the desired result and out comes the algorithm that turns one into the other. Learning algorithms— also known as learners—are algorithms that make other algorithms." (Domingos 2015:6)

Therefore, these algorithms can change without human interaction and operate in unintended ways, which is what educated guesses supported by data conclude that the YouTube algorithm may do (Bryant 2020:87; O'Donovan et al. 2019).

## 4.2 Explaining the YouTube Algorithm

In the early days of YouTube, the recommendation algorithm of the platform was primarily designed to recommend users videos that they would be interested in based on their currently playing videos. This meant that if a user went to YouTube and watched cat videos, they would only get more cat videos recommended which created 'thematic loops' for the users (Roose et al. 2020: Episode 1). During political conflicts and terrorist attacks like the Arab Spring in late 2010, and the terror attack in Nice, France 2016 it became evident that these loops restricted the viewers to one-sided parts of reality (Roose et al. 2020: Episode 1 and 4), and from seeing well-informed news media.

> "You would see a video from the side of the protesters [during the Arab Spring], and then it will recommend another video from the side of protesters. So, you would only see the side of protesters. If you start with the side of the police, you would only see the side of the police." (Roose et al. 2020)

YouTube started redesigning the algorithm to do the exact opposite of the then recommendation algorithm, getting the users out of the loops (Roose et al. 2020). Around 2015, YouTube teamed up with Google Brain, a state-of-the-art artificial intelligence (AI) team (Newton 2017) where they started developing a complex self-learning algorithm. The recommendation system was developed with an architecture referred to as a *Deep Neural Network* (Covington, Adams, and Sargin 2016). This type of AI is designed to imitate the human brain, finding connections and patterns designed to get users interested in topics and videos and which learns "approximately one billion parameters and (...) [is] trained on hundreds of billions of examples" (Covington, Adams, and Sargin 2016:1). This powerful algorithm also improves itself through machine learning, meaning that "every time it has a successful interaction, and a user allows one of the suggested videos to be played, the algorithm learns that there is a relationship between the video watched and the video suggested" (Bryant 2020:86). This self-learning process is automatic and happens without any human interaction. Because of this, exactly how the algorithm works is a bit

27

of a black box, having some of its internal logic opaque even to its engineers (Bryant 2020:86), which can make it hard to control or study.

Getting in-depth information about big tech companies' algorithms, whether it be the news feed algorithm on Facebook or YouTube's recommendation algorithm, is unattainable to the public eye or researchers. As Lewis and McCormick write in an article in The Guardian about bias in the video recommendation on YouTube:

"Disclosing that data would enable academic institutions, fact-checkers and regulators (as well as journalists) to assess the type of content YouTube is most likely to promote. By keeping the algorithm and its results under wraps, YouTube ensures that any patterns that indicate unintended biases or distortions associated with its algorithm are concealed from public view." (Lewis and McCormick 2018)

In 2016, shortly after the enrolment of Google Brain's AI technology on YouTube, Google employees wrote an academic journal *Deep Neural Networks for YouTube Recommendations* (Covington, Adams, and Sargin 2016), seemingly the only released paper from Google itself where the YouTube algorithm is explained from the early phases of the new machine learning algorithm. Here it is said to be composed of two neural networks: one for candidate generation and one for ranking (Covington, Adams, and Sargin 2016:2). The candidate generation network "provides broad personalization via collaborative filtering" and analyses the similarity between users in terms of their IDs of video watches, search query tokens and demographics (Covington, Adams, and Sargin 2016:2). This means that the algorithm collects all its information about a user and compares this information to other users' finding similarities and then selecting videos which it thinks will be interesting. Through this process, millions (if not billions now) of videos on YouTube get condensed down to some hundreds before the next phase (see "video corpus" in Figure 2). The candidate phase judges users' watch history and since the algorithm is designed to maximise watch time on the site it does not judge what users have clicked on or how they have given feedback (thumbs up/down, in-product surveys, etc.), but by whether they have finished watching the video, which the algorithm sees as a positive example (Covington, Adams, and Sargin 2016:2).
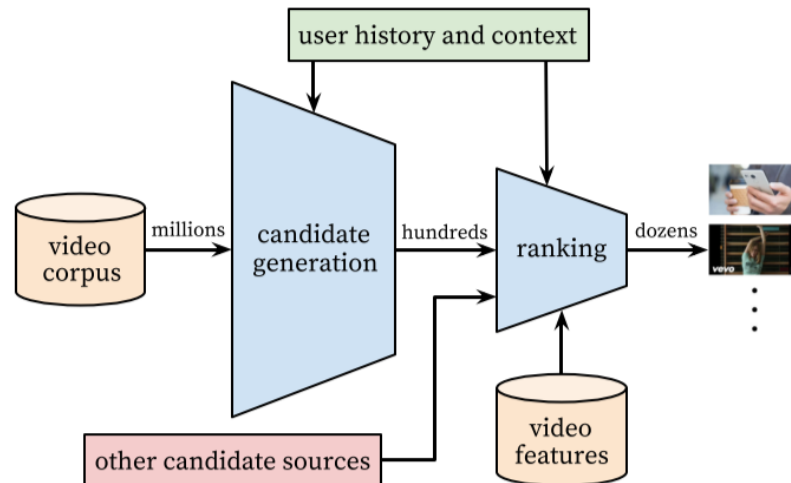
Figure 2: Recommendation system architecture demonstrating the "funnel" where candidate videos are retrieved and ranked before presenting only a few to the user (Covington, Adams, and Sargin 2016:2)

The ranking network phase assigns scores to each video "according to [the] desired objective function using a rich set of features describing the video and user" (Covington, Adams, and Sargin 2016:2). Here, it carefully monitors the user's behaviour on the site to make the best-estimated guess on what will keep users watching:

"As an example, [the algorithm considers] the user's past history with the channel that uploaded the video being scored - how many videos has the user watched from this channel? When was the last time the user watched a video on this topic?" (Covington, Adams, and Sargin 2016:6)

If a user does not watch a recently recommended video, then the algorithm naturally demotes it (Covington, Adams, and Sargin 2016:6), meaning that the algorithm is always "at work", constantly adapting to the users' behaviour. The ranking phase also considers how users rate feedback on videos. The ranking phase, therefore, works to sort through the hundreds of videos that have made it through the candidate sorting phase to determine which videos to recommend that have the highest chance of making users watch videos on the site for the longest time. From here some dozens of videos to feature as recommendations are chosen. Several sources point to around 70% of the videos seen on YouTube currently are from the video recommendations (Høst 2019; Newton 2019; Roose and Conger 2019).

## 4.3 Recommendation Biases on YouTube

When speaking about biases in algorithms, I define it as an (unintentional) disproportionate weight towards certain videos or clusters of channels on YouTube. In *Algorithms of Oppression* (2018), Noble describes machine learning biases in for example the Google search engine as the way they work "to its own economic interests— for its profitability and to bolster its market dominance at any expense" and says that "what is missing from the extant work on Google is an intersectional power analysis that accounts for the ways in which marginalized people are exponentially harmed by Google." (Noble 2018:28). The algorithmic bias on YouTube also works in favour of popular channels, musicians, and viral videos, since they are followed and watched by a lot of users. This thesis focuses on the algorithmic bias which leads users on YouTube to channels in the manosphere with extreme tendencies. Here, the goal is to examine if the algorithm may lead users, who are not originally interested in these communities, to get radicalised, and possibly draw them into rabbit holes.

### 4.3.1 Testing the Algorithm: Case of U.S. Presidential Election

A former YouTube employee Guillaume Chaslot founded a company called AlgoTransparency (hereafter AT) after leaving YouTube which "provides transparency on YouTube's algorithms" (www.algotransparency.org). On the website, they post daily updates on the most recommended videos and channels of YouTube for each day. AT uses an algorithm that starts "watching" YouTube videos from a list of around 800+ US channels without being logged in and without cookies, making the algorithm unable to detect anything but the location of the IP address which is set to New York City, USA. The channels have been selected to be the most recommended political channels ("YouTube: Candidates Favored by the Algorithm" n.d.). From here, AT gathers all videos recommended by the last videos uploaded by the channels. Then they are counted and displayed by how many recommendations they get for each video observed ("YouTube: Candidates Favored by the Algorithm" n.d.). Though this method of extracting data from YouTube does not provide a comprehensive or perfect representative sample of the recommended videos algorithm, it gives a snapshot, Chaslot said to The Guardian (Lewis and McCormick 2018). During the 2016 U.S. presidential election, the AT software was programmed to imitate users interested in either Donald Trump or Hillary Clinton and then followed "a chain of YouTube–recommended titles appearing "Up Next" after initially having searched for

"Trump" or "Clinton" (Lewis and McCormick 2018). The algorithm made sure that the two candidates were searched equally 50% of the time, and then clicked on several of the (usually top 5) search results and noted which videos YouTube recommended as "Up Next". The algorithm ran on different dates throughout the election period and collected data from up to four "layers[23]" of recommended videos and ended up including 8,052 videos. Out of these, 1,000 were qualitatively analysed, 500 which were recommended after "Clinton" and 500 which were recommended after "Trump". The research team deemed that about a third of the videos were either unrelated to the election, politically neutral or insufficiently biased to getting categorised as favouring either campaign but that 643 out of the 1,000 were deemed to have an obvious bias. 86% of these favoured the Republican nominee and 14% favoured Clinton (Lewis and McCormick 2018).

As seen in this case, the study found the recommendation algorithm on YouTube to be republican leaning, and there could be many explanations to this, which does not necessarily mean that users watching said videos are to be radicalised or even identify with the political views expressed in the videos, especially since Trump has been somewhat of an internet sensation and gone viral on several occasions. It is noteworthy though, that media culture and memes may also promote, normalise, or popularise content, and make videos that contain these topics or talk into these cultures rank higher in the recommendation algorithm. But studies do show that conservative videos lie closely to other videos, which are more radical and have more extremist tendencies, as will also be shown in the next section *Mapping the Misogynist YouTube*:

"[The YouTube algorithm] suggests that being a conservative on YouTube means that you're only one or two clicks away from extreme far-right channels, conspiracy theories, and radicalizing content." (Kaiser and Rauchfleisch 2018)

So even if it is not intentional, the algorithm seems to have this bias to Right-leaning and Right extremist content (Bryant 2020:87), making moderately conservative videos only a few clicks away from 'far-Right' filter bubbles (Bryant 2020, Kaiser and Rauchfleisch 2018, Roose et al. 2020, DareGender and Cybernauterne 2020, Rogers 2020). As will be explored in *Mapping the*

---

[23] Meaning following the recommended videos four times from the original video which was clicked on in the search results

*Misogynist YouTube* there are "bound relationships between the Right-leaning content on YouTube, specifically "Fox News," "Alex Jones[24]," and "white nationalists," along with some conspiracy theories, anti-feminist, anti-political correctness channels, and a channel called the Manosphere" which creates an algorithmic bias to recommend racist or white supremacist videos more often to users (Bryant 2020:87; Kaiser and Rauchfleisch 2018). Because these are clustered in the recommendation algorithm, these intertwine and lead users to rabbit holes of content. As seen in DareGender and Cybernauterne (2020), Danish users also follow American channels, content, and groups, why this is relevant for this thesis as well.

## 4.4 Moderation and Community Guidelines

YouTube, naturally, has community guidelines and standards to especially prevent illegal content on the platform. This includes areas such as sexual content, violent and graphic content, hateful content and hate speech. But as the platform grows, getting more than 500 hours of new videos uploaded every minute (Roose and Conger 2019), far from all content that violate the community guidelines is removed. When the corona pandemic broke out, Susan Wojcicki, CEO of YouTube, explained in an interview with Kevin Roose that YouTube was reviewing all videos about the coronavirus, acting very proactively to prevent and remove any mis- and disinformation on the site about the virus, and promoting content from reliable sources like the World Health Organisation (WHO) (Roose et al. 2020: Episode 4). At the same time, every video about the coronavirus had a link, which directed the viewer to health authorities' websites. Kevin Roose explains:

"They created this feature that, basically when you search for information about the coronavirus, the first thing you see is this section called top news that contains essential information from authoritative, trusted sources." (Roose et al. 2020: Episode 4)

So, whilst YouTube knows how to monitor and control the content on its site so that harmful or false information gets removed or viewers directed to other, more reliable sources, the platform has received heavy criticism for not doing enough to prevent racist and extreme content, and for having unclear, and inconsistent rules about moderation (Roose and Conger 2019). The

---

[24] Famous YouTube conspiracy theorist

moderators working at YouTube under subcontractor Accenture, are reported as being at high risk for getting PTSD through reviewing the extreme and violent content uploaded on the site, and they must sign a statement of acknowledgement before starting their jobs, stating that they know of the severe mental health issues which can be a side effect of the job (Newton 2020). These workers are also often low-paid immigrants working for 18.50 USD an hour (Newton 2020), making the jobs both stressful and traumatising. In an earlier research project of mine[25] where empirical data was collected through focus group design sessions to explore how to optimise social media platforms, making them safer and less discriminatory and oppressive, a suggestion was to optimise moderation. This could be through more resource location, optimised moderation bots, and independent moderation units, which (preferably) could moderate on behalf of experiences of vulnerable groups (Mooghen and Bau Larsen 2019:39). Naturally, moderation bots would ultimately be able to moderate and remove the most harmful, violent, and traumatising content to protect the well-being of human moderators.



**Spam & deceptive practices**
The YouTube Community is one that's built on trust. Content that intends to scam, mislead, spam, or defraud other users isn't allowed on YouTube.
- Spam, deceptive practices, & scams policies
- Impersonation policy
- External links policy
- Fake engagement policy
- Additional policies

**Sensitive content**
We hope to protect viewers, creators, and especially minors. That's why we've got rules around keeping children safe, sex & nudity, and self harm. Learn what's allowed on YouTube and what to do if you see content that doesn't follow these policies.
- Nudity & sexual content policies
- Thumbnails policy
- Child safety policy
- Suicide & self-injury policy

**Violent or dangerous content**
Hate speech, predatory behavior, graphic violence, malicious attacks, and content that promotes harmful or dangerous behavior isn't allowed on YouTube.
- Harmful or dangerous content policies
- Violent or graphic content policies
- Violent criminal organizations policy
- Hate speech policy
- Harassment & cyberbullying policies
- COVID-19 medical misinformation policy

**Regulated goods**
Certain goods can't be sold on YouTube. Find out what's allowed—and what isn't.
- Sale of illegal or regulated goods or services policies
- Firearms policy

YouTube's community guidelines as seen on Google's website[26] as per May 2021

YouTube has, over the recent years, started to demonetise channels, withdrawing the opportunity for channels to monetize from ads on their content, in the manosphere, which have violated the community guidelines on the site, and in 2019 they announced a plan to remove thousands of videos that advocate neo-Nazism, white supremacy and other bigoted ideologies and other videos "alleging that a group is superior in order to justify discrimination, segregation or exclusion"

---

[25]I and a research partner explored how social media platforms could be designed more responsibly in correlation with the injustices and oppression of vulnerable groups, and did qualitative research with trans and queer people, tech-experts, a techno-anthropologist and a digital safety worker.
[26]YouTube's community guidelines: https://support.google.com/youtube/answer/9288567?hl=en

(Roose and Conger 2019). But as it is evident, 'far-Right' and misogynist communities still exist on YouTube. Even though content creators may get their channels shut down, they still guest in videos, podcasts and talk shows broadcasted by other content creators. The 'far-Right' and manosphere community is widespread and vast. In some cases, platforms have even warned accounts hours before being removed, as with popular creators Yiannopoulos, Jones, Laura Loomer, and Paul Joseph Watson, who got removed for being 'dangerous individuals', but still was given time to redirect their audience to other platforms (Martineau 2019). The removing of content creators, also called *deplatforming*, will be described further in this thesis. It is therefore unclear under which goals and measures YouTube chooses to deplatform channels.

## 4.5 Existing Censorship: The SESTA/FOSTA Law

Although discourse is often about online censorship and the need to protect it, little is acknowledged about the laws and measures already in effect censoring especially marginalised users online. Whilst some of the extreme discriminatory content and comments seem to exist rather freely, and often are overlooked by moderation and algorithms, other content has been heavily restricted and impacted the past few years. The SESTA/FOSTA law, the acronym for Stop Enabling Sex Traffickers Act (SESTA) and Fight Online Sex Trafficking Act (FOSTA), have been described by many as an internet censorship threat. The Electronic Frontier Foundation describes it as "an unprecedented push towards Internet censorship and does nothing to fight sex traffickers" (Mullin 2018). Legal definitions of sex trafficking often mistarget people voluntarily engaging in sex work or otherwise working with sex-positivity. Research finds that the law "(...)hurts sex workers by taking away the ability to screen clients, (...) forces people back to the streets and into more dangerous situations, (...) heightens risk of arrest, and (...) contributes to sex workers' vulnerability to third-party market facilitators (e.g., traffickers or pimps) among other harms" (Musto et al. 2021:2). In practice, it also means that a lot of content outside the scope of sexual content gets removed, flagged, or demonetised from social media sites. Moderation, community guidelines and algorithms are known to target and disadvantage women and minorities such as BIPOC, trans and queer people, fat people and femmes[27] (Sap et al. 2019;

---

[27] A (self-identified) queer identity of all genders who relate to femininity and doesn't necessarily abide to traditional rules of femininity

Farokhmanesh 2018; Mooghen & Bau Larsen 2019:28; Smith 2019). This adds to a disadvantage for marginalised users in online spaces.

## 4.5.1 Minorities in Online Spaces

Places like Tumblr, known as a sex-friendly social platform, used by many Left-leaning, feminist and trans and queer users, heavily started restricting their websites after the SESTA/FOSTA laws launch around 2017, and eventually experienced a heavy decline in activity. On Tumblr an algorithm was launched to flag pornographic content since the site would now be held responsible for any sex work advertisements on their site, whether they knew about the content being posted or not because of the SESTA/FOSTA law (Cyboid 2018). But the algorithms detecting content are not always successful in their tasks. Many minorities have time and time again complained over the double standards of algorithms on social media platforms. Fat users have complained about algorithms flagging their photos as "nudity", even when being dressed because they have a larger percentage of skin on their photos (Richman 2019). Several trans YouTubers have complained of being demonetised and otherwise had their content restricted because of being trans or queer (Farokhmanesh 2018).

Tweet by trans YouTuber @Chaseross frustratingly documenting how the algorithm demonetises his videos when the titles include the word "transgender" from May 30, 2018[28]

A study from 2019 analysing racial bias in algorithms showed that AAVE[29] tweets were twice as likely as being labelled as offensive compared to other tweets (Sap et al. 2019:1672). Recently, as new measures of the SESTA/FOSTA law got rolled out, several social media platforms updated their community guidelines, enforcing even stricter rules to its users. In practice, this has meant that common words have been forced to be replaced with new internet lingo, for example writing "seggs" instead of "sex" to not get flagged. Dr Francesca Sobande, a lecturer of Digital Media at Cardiff University said in an interview:

"Online suppression is symbolically violent (...) Certain people are demonized in our society, especially LGBTQ, black women and transgender people. (...) [FOSTA-SESTA] inhibits people's ability to use these platforms, and that can result in losses of income that exacerbate already precarious living situations."
(Holland 2019)

Not only does this have consequences to the already marginalised groups as they are not able to create their content or earn money from it on the social media platform, but it also means that the online spaces in which we exist become less diverse. At the same time, if content in the manosphere or other Right-leaning content with extremist tendencies do not get censored or removed, there will exist an overrepresentation of this content. The effects of moderation bias may play a part in why YouTube has a much larger cluster of channels in the manosphere compared to narratives which counter these, which will be explored later in the next section *Mapping the Misogynist YouTube*. As seen in this technology review, algorithms are not neutral, and as Angela Noble writes "The people who make these decisions hold all types of values, many of which openly promote racism, sexism, and false notions of meritocracy (...)" (Noble 2018:1-2), which is why we should not consider algorithms as value neutral and assess where the machine learning algorithms act in harmful ways.

---

[28] Tweet accessed May 3, 2021: https://twitter.com/ChaseRoss/status/1001922360600137728

[29] African American Vernacular English

## 4.6 Conclusion to Technology Review

In YouTube's quest to design an algorithm that promotes and recommends content to users which they might not have searched for, but which could, based on the interests of other users and demographic information, be of interest, the algorithm shows some problematic side effects. Although the YouTube algorithm is not proven to intentionally promote content from channels on YouTube with extremist tendencies, the infrastructure of the algorithm can end up promote content to users who might not align politically with the channels recommended. Since the algorithm aims to maximise, users watch time, it may functionally work well, but it can also cause unwanted or harmful consequences, which can be prevented or minimised, as seen with the effort that YouTube has put into removing disinformation around the coronavirus.

At the same time, moderation bots, unclear or seemingly biased community guidelines, along with laws like the SESTA/FOSTA law make it difficult for marginalised users to use the platform, which can create an unfair advantage for other channels.

Because these algorithmic codes are inaccessible for the public and researchers, it is impossible to analyse exactly what the algorithm is coded to do, and therefore we must rely on simulations and data, which test how the algorithm *functions* in practice, and use the sparse information released by Google itself, which is outdated (from 2016). Many sources criticise the recommendation algorithm for being harmful, and whilst YouTube denies responsibility for its algorithm, it is clear that if harmful communities are allowed to exist on YouTube, the algorithm will inevitably also recommend and promote them.

# 5 Mapping the Misogynist YouTube

In this section the YouTube's manosphere will be explored. Here, I will also analyse if YouTube's recommendation algorithm can create 'filter bubbles' of (potentially) dangerous communities and explore cases where the online manosphere has been connected to violence and terrorism.

## 5.1 YouTube's Manosphere: Clusters of Communities

The concept of the manosphere is somewhat broad and exists in a big network connecting to other content, groups and spaces rooted in a general distrust and critique of modern society, politics, and media within the 'far-Right'. In Figure 3 we see a network of YouTube channels visualised by an algorithm in a research project by Kaiser and Rauchfleisch (2018) for The Media Manipulation Initiative at Data & Society.
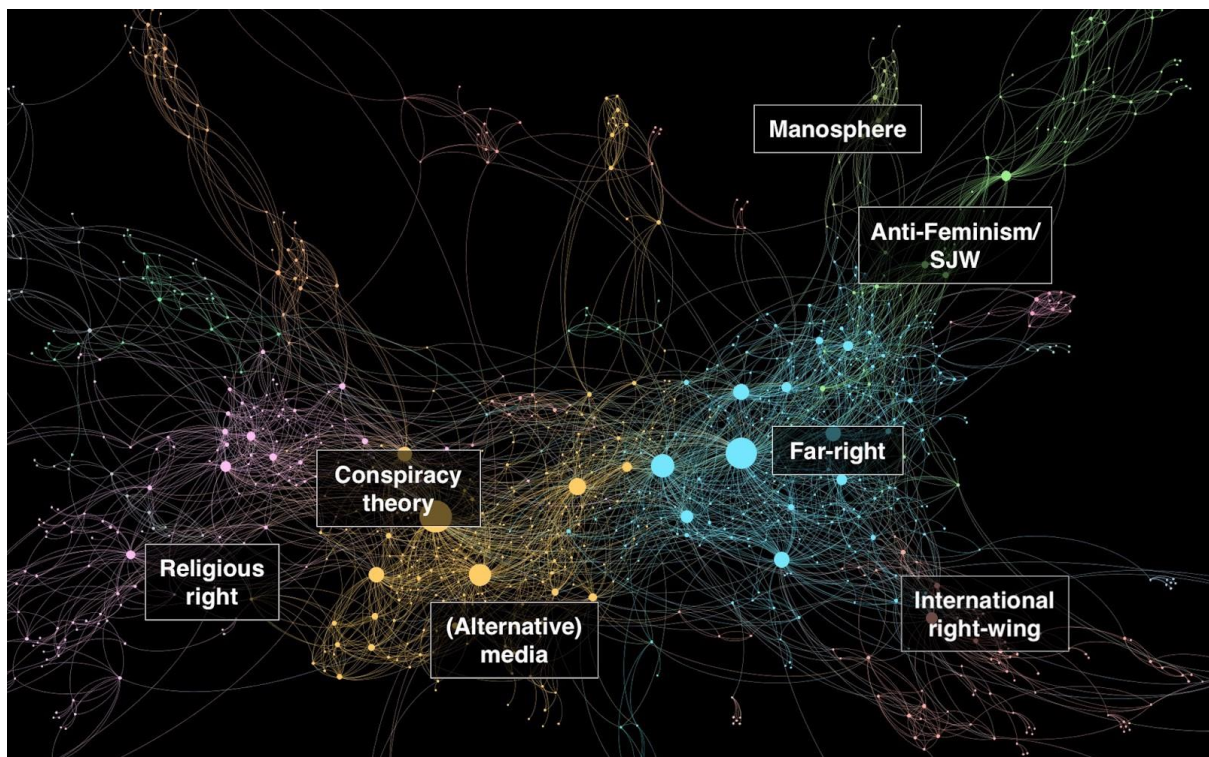


Figure 3: YouTube channel network with labelled communities (node size = indegree/amount of recommendations within network; community identification with Louvaine) (Kaiser and Rauchfleisch 2018)

In this visual network, every channel is a node (the dots), and every recommendation from one channel to another is an edge (lines between the dots) (Kaiser and Rauchfleisch 2018). The sizes of the nodes indicate how much their videos are recommended by other channels. In the study, 13,529 channels on YouTube were analysed to map out how these were clustered (grouped) and how they were connected to each other. A cluster in these visual maps are groups of channels that are "closely knitted", meaning there is a high amount of recommendation between these channels. For example, videos from a musical channel will often recommend videos from other music channels. If a lot of channels do this with each other they create a cluster, which then can be qualitatively labelled. But even though music channels often recommend videos from other music channels, they also recommend videos from channels in other clusters. The more proximity a cluster has to another, the more the recommendations on their videos are intertwined. The channels used in the Kaiser and Rauchfleisch research (2018) are mainly from the US but are relevant in a Danish setting as well since international YouTubers are referred to by Danish influencers in the manosphere and Danish users mention US channels as well in interviews about which content they watch (DareGender and Cybernauterne 2020:7). A similar study could be relevant in Denmark to map the Danish manosphere.

The clusters we see in Figure 3 is from a larger network where all channel clusters have been collected and labelled (see Figure 4 below), from which the channels relating to the 'far-Right', and conservative media and conspiracy theories are clustered together. Here, we have two main clusters "Far-right", and "(Alternative) media". The "Far-right" cluster consists of "alt-Right" and "alt-light" channels as well as white nationalists (Kaiser and Rauchfleisch 2018). The "(Alternative) media)" consists mostly of media outlets, but what is worrying here is that large media outlet channels like the GOP (The Republican Party) and Fox News are closely linked to conspiracy theorists like Alex Jones. This means that if a user starts from a YouTube video on Fox News channel, the recommendation system can cluster the channels together as seen on Figure 4 below, creating a "filter bubble", an algorithmically created bubble that "alters the way we encounter ideas and information." (Pariser 2012 in Kaiser and Rauchfleisch 2018).

In Figure 4 we see how the larger YouTube universe is connected through clusters. Although clusters like "Kid's shows" and the "Far-right" are far apart, the centred clusters and especially the "Mainstream and progressive media & politics" which contains popular channels like The Young

Turks, CNN, Barack Obama, Elizabeth Warren, and Bernie Sanders are very close to the extremist, conspiracy theorist and anti-feminist channels. Though there is a "Far-left" seen on the network, this cluster is not particularly big in comparison to the 'far-Right' and conspiracy clusters (Kaiser and Rauchfleisch 2018). The 'far-Right' communities use YouTube to stream videos, radio talk shows, and podcasts for hours and hours a day (The New York Times 2020; Kaiser and Rauchfleisch 2018), leaving Left-leaning content at a disadvantage if they have less frequent and shorter posting habits. The Right-wing content creators on YouTube are strongly mobilised, and often invite each other in as guests in their videos, and refer to other YouTubers, creating a tight-knit cluster (Roose et al. 2020).
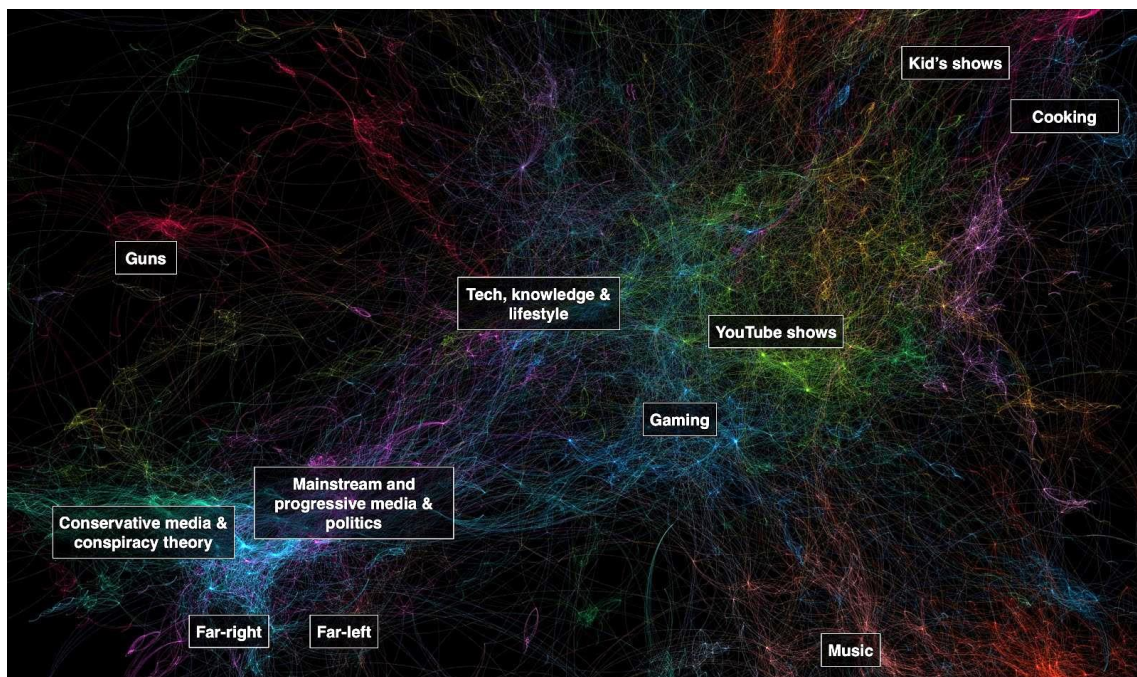


Figure 4: YouTube channel network with labelled communities (node size = indegree / amount of recommendations within network; community identification with Louvaine) (Kaiser and Rauchfleisch 2018)

As the research study from Kaiser and Rauchfleisch shows how the channels recommend each other, it is important to note that the recommendation algorithm on YouTube may possibly direct users to different channels due to the machine learning algorithm. As it considers how the users interact with the site, their demographic details and more, it is not a guarantee that a video from the Fox News channel will direct the audience into the manosphere, and neither is it guaranteed that a user watching a music video from a popular music channel will not get recommended a

video from the manosphere. What is seen from this study is that there exists a clear cluster, which I would label as having concerning ideologies, and which are likely to be recommended to users watching videos from less radical or harmful videos and channels. As the manosphere cluster is somewhat isolated, and 'counter' narratives, which could be found in the cluster "Far-left", is very small compared to the right-leaning cluster, there is a possibility of users falling into rabbit holes. Here, they will get more of the same content recommended, and possibly experience more content with extremist tendencies, which can lead to radicalisation.

## 5.1.1 The Danish Manosphere

In a Danish context worth mentioning are groups such as ethno-nationalist like Generation Identitær and Stram Kurs, anarcho-capitalists (wanting to eliminate the state and let the free market roam) like Lars Kragh Andersen, who also argues that Denmark would be better if women could not vote (DareGender and Cybernauterne 2020:10), 'New Heathens' and traditionalists, neo-Nazis like Nordfront (The Nordic Resistance[30]), fighting for the Nordic countries only being populated by white people, and conspiracy theorist groups such as JFK 21, and the new anti-COVID-measures movement Men In Black. These groups and communities heavily depend on the internet to spread their ideologies and recruit new members, and many of these through YouTube.

Rasmus Paludan, the front person of the new Danish right-wing party Stram Kurs (Hard Line) was only 0.2% votes from being elected in the last parliamentary election in 2019 and has had major success through YouTube. By burning the Qur'an and having confrontations and demonstrations filmed and streamed on YouTube, the channel went viral in Autumn 2018 and Spring 2019 (DareGender and Cybernauterne 2020:20). YouTube deplatformed the channel by deleting it in September 2019 because of not respecting the community guidelines, but by then the channel already had 40,900 subscribers and had over 31 million views on their 634 videos (DareGender and Cybernauterne 2020:20). Both The Nordic Resistance and Stram Kurs were removed by YouTube for having racist and anti-Semitic content (DareGender and Cybernauterne 2020:21), but quite late in terms of how wide-spread their messages had reached.

---

[30] Translated from Danish: Den Nordiske Modstandsbevægelse

## 5.1.2 Children and Youths

Something which is studied in a Danish context in the NCPE report is juvenile online radicalisation. Other sources also point to educating children and youths in order to prevent chances of political grooming (Mogensen and Helding Rand 2020).

Some channels and videos on YouTube are aimed especially at children and youths, and sometimes also created by youths themselves. Danish expert in digital culture, Katrine K. Pedersen, says there is a trend to make the content as extreme as possible, often with explicit and extreme titles to make people click on them, and boosting their exposure through the algorithm (through click-bait), and having content making jokes about paedophilia, Hitler, incest, and misogyny (Høst 2019:12:20). These channels and cultures are often interlinked from gaming culture and forums like Twitch and Discord, where it is common to make "reaction"-videos of something between playtime when popular gamers live stream. This links children and youths to different YouTubers and 'edgy' meme cultures (Høst 2019). PewDiePie, who will also be introduced later, is the world's most subscribed YouTuber and has gained a lot of success with his young audience through a mixture of gaming-related videos and 'edgy' humour, oftentimes boundary seeking and crossing (Høst 2019). Pedersen furthermore says that many adults and parents do not know what is going on in children's online social life, and it can be difficult to identify healthy communities and when the jokes change to something boundary crossing. The NCPE report has conducted a desk research on juvenile online radicalisation, and mentions several online phenomenons such as political grooming, propaganda, mis- and disinformation, dehumanisation and hate speech, and conspiracy theories. The term grooming is known from research on sexual abuse and describes a process where an abuser, often without explicit use of violence or threats, slowly convinces, seduces, and manipulates a child to sexual activity (Stender Petersen and Alberg Peters 2020:13). Similarly, political grooming is the process of slow and gradual attachment and normalisation of the introduced thoughts, feelings, and ideas which children and youths meet online. Children and youths could be especially vulnerable to radicalisation and extremist communities, in their search of inclusion and belonging (Stender Petersen and Alberg Peters 2020:13-14). Mis- and disinformation are also especially dangerous to children and youths, who may not have the skillset to critically analyse the sources and facts of the information that they are given. Laura Bates, the founder of the Everyday Sexism Project, explains that she often hears young boys express propaganda when she goes to elementary schools to teach sexism (Ford 2020). Examples can be mentioning false statistics about men being the highest

percentage of domestic abusive victims and other anti-feminist propaganda (Ford 2020). When children get these narratives from idolised internet personalities or hear this discourse in their online communities, it creates both normalisation and peer pressure to align and accept the behaviour (Stender Petersen and Alberg Peters 2020).

Because children and youth are vulnerable when navigating the internet and engaging in communities, like gaming cultures, which are not harmful by default, it is important to find proper solutions and preventative measures. Through 'edgy' humour and border seeking nature in youth, it is important for children and youths to not be politically groomed or get harmful ideas normalised through their online lives. As for now, YouTube's guidelines restrict children below 13 years to use the platform, although it does not affect how the platform is used by children (Høst 2019). This guideline creates a responsibility gap for YouTube, since they can claim that their content is not eligible or allowed to be seen by people under 13 years (Høst 2019). Some of the measures suggested at the end of this thesis aim specifically for children and youths to prevent them from getting radicalised, as these can act in a proactive way to prevent future radicalisation and terrorism. Knowledge of harmful online cultures for parents, teachers, and other social workers can detect early radicalisation, as well as giving children and youths skillsets and tools to avoid harmful ideologies, communities, and rabbit holes online.

## 5.2 Rabbit Holes: Cases of Falling

To understand which role YouTube and its algorithm plays in radicalisation, the *rabbit holes* will be explained 'Rabbit hole' is a term used to describe the extended YouTube binges that occur for users, who usually start out by watching something unrelated to political videos or moderate content but end up going gradually "deeper" down the more radical political YouTube videos. The videos in these rabbit holes usually spread propaganda and offer no counterarguments, creating what can be called *echo chambers*. Because these videos in rabbit holes exist in the same clusters or near more radical clusters of videos, it is easy getting into more and more radical YouTube binges. Mixed with the powerful algorithm that works to maximise users' watch time, it is not difficult to imagine how some users, especially those who feel marginalised, alone or humiliated by their surroundings, get radicalised. The term 'pipeline' is also used in explaining how "lighter" content and humour which is more mainstream and accessible can work as desensitisation to more radical and harmful ideas.

Projects like DareGender and Cybernauterne's report (2020) and The New York Times podcast *Rabbit Hole* by Kevin Roose et al. (2020) have interviewed people who have been affected by radicalisation in YouTube's rabbit holes. In the New York Times podcast, rabbit hole indulged Caleb Cain's YouTube watch and search history from 2015 and through 4 years is explored (Roose et al. 2020). Cain, a 26-year-old man from West Virginia, USA, was somewhat socially isolated and found comfort and friends through the internet. At the beginning of 2015, Cain identified politically with Barack Obama, and from his YouTube watch history, it can be seen how he got interested in self-help videos and was then recommended to Stefan Molyneux, a Canadian 'far-Right' white nationalist. As Cain says: "Stef [Stefan Molyneux] just was in the sidebar one day, and I clicked on it." (Roose et al. 2020: Episode 1). This led Cain to be referred to by many other prominent YouTube personalities from the manosphere and 'far-Right'. These content creators are also known by Danish users (DareGender and Cybernauterne 2020), why the case of Caleb Cain may be transferable to a Danish context. Some of these videos he got recommended were Molyneux's series "The truth about…", which often discussed controversial topics like "The truth about the Native American genocide", "The truth about slavery", "The truth about immigration". These "truths" are played out to be an unbiased, honest, and uncensored versions of what is taught through mainstream media (Roose et al. 2020: Episode 2). DareGender and Cybernauterne also describe how this is a well-known strategy and phenomenon in the manosphere, and that the manosphere gives simple answers to complex problems:

"The communities we describe throughout this report all have in common that they give answers to existential questions of identity and affiliation. They explain how the world should be, be it a white ethnostate or a stateless society, and they simultaneously urge for a return to a traditional understanding of gender where there are firm and natural roles for men and women. (…) As one of our interviewees expressed, the world was a simpler place, and he was less doubtful back when he was still a part of the [manosphere]" (DareGender and Cybernauterne 2020:26)

A sense of belonging, answers, and explanations from people who look like you, and possibly have some of the same life experiences as having felt marginalised, humiliated, lacked a community, and felt like they did not belong are factors which Karacan explain as some which can be present in people who have been radicalised or have become terrorists (Karacan in interview with author, February 2021). Oftentimes, the content creators in YouTube's manosphere are funny, well-spoken, and "professional" looking men, who look and sound very different to the image of a

stereotypical neo-Nazi or white supremacist. These factors can act in a way which politically grooms the audience. Some users are then "lured" into these rabbit holes, getting radicalised and desensitised by people who look "normal", who look like them:

| Interviewee (from DareGender and Cybernauterne 2020:25-26) | I watched Stephen Crowder [known from the viral meme "Male Privilege is a Myth/Change My Mind" ed.] and Gavin McInnes [leader of the violent, neo-fascist group Proud Boys]. I did not know who they were. I never researched them. If I actually had researched them, I would never have found them appealing. I started with the atheist Richard Dawkins, but he has sent me to the Right, to people who were deeply religious. I first found that out when I started doing background checks of those I watched. |
| --- | --- |

This interviewee watched content from individuals with extremist ideologies, which ended up acting as a radicalising factor, although the interviewee expresses that they would never have watched it, if they had known. Here, it can be seen how background checks can be a powerful tool to provide users with critical thinking abilities when gaining information from online spaces. If these videos perhaps had content warnings or were flagged with hate speech or harmful ideologies, users may be more careful consuming this kind of content. In these cases, we also see how the recommendation algorithm leads users into finding more and more radicalising content. This sort of political grooming and normalisation also contains dog whistles, which in many cases end up going unnoticed by moderation and audience who do not have knowledge of these strategies. When content creators in the manosphere use the argument of addressing problems which nobody "dares to speak about", often controversial topics such as race realism or restoring society through traditional gender roles, it can excite and interest people, who already feel marginalised or misunderstood by society, and who can find meaningful, and simple, "solutions" to their issues and suffering.

## 5.3 The Danger of the Manosphere: Murderers and Memes

Online radicalisation can have consequences manifesting outside the online sphere, which is yet another reason why this is an urgent problem needing attention and solutions. Forums and communities in the manosphere are known to breed violent terrorism and extremism. Incels were some of the first ones to appear on the radar of personas existing in what we now know to be the vast universe of the manosphere due to several attacks with ties to the incel-communities.

### 5.3.1 Terror Attacks

Canadian Alek Minassian drove a van into pedestrians killing 10 people and injuring 16, and prior to the attack allegedly wrote: "Private (Recruit) Minassian Infantry 00010, wishing to speak to Sgt. 4chan please. C23249161. The Incel Rebellion has already begun! We will overthrow all the Chads and Stacys[31]! All hail the Supreme Gentleman Elliot Rodger!" in an incel-community on Facebook (Tierney and Lamoureux 2018). Elliot Rodger is used here as a reference to the 22-year-old English/Chinese-Malaysian man who killed 7 people including himself in Isla Vista, California in 2014. Prior to the attack, Rodger had posted a YouTube video titled "Elliot Rodger's Retribution" and sent out a manifesto titled "My Twisted World: The Story of Elliot Rodger" in which he spews several overtly racist and misogynist ideas. In his manifesto, he describes his frustration of being different from white kids because of his mixed Asian descent and writes stories about his childhood and youth. In both his manifesto and his YouTube video, it is obvious how blatant misogyny and white supremacist ideologies intersect in the incel-communities. Examples of Rodger's ideologies are seen in his manifesto where he writes about his disdain for mixed-race couples. This is rooted in his confusion of Black and Asian men being able to have relationships with white women, when he "could never even have their attention", writing i.e.: "How could an ugly Asian attract the attention of a white girl, while a beautiful Eurasian like myself never had any attention from them?" (Rodger 2014:121). In his video, he chillingly explains how he will "go into the hottest sorority house of UCSB" and "slaughter every spoiled, stuck-up, blond slut he sees inside there" (CNN 2014). In the video he further states how involuntary celibacy and the lack of romantic interest from girls and women throughout his life has made him feel alienated and outraged (CNN 2014),

---

[31] "Chads" and "Stacys" refer in the incel community to attractive, non socially awkward people.

| Elliot Rodger: | I'm 22 years old and I'm still a virgin. I've never even kissed a girl. (...) You girls have never been attracted to me. I don't know why you girls aren't attracted to me, but I will punish you all for it. It's an injustice, a crime, because... I don't know what you don't see in me. I'm the perfect guy and yet you throw yourselves at these obnoxious men instead of me, the supreme gentleman. I will punish all of you for it. (...) All those popular kids who live such lives of hedonistic pleasures while I've had to rot in loneliness for all these years. They've all looked down upon me every time I tried to go out and join them, they've all treated me like a mouse. Well, now I will be a God compared to you. You will all be animals. You are animals and I will slaughter you like animals. And I will be a God. Exacting my retribution on all those who deserve it. You do deserve it. Just for the crime of living a better life than me. |
|---|---|

After his mass murder, Elliot Rodger gained heroic status in the incel-community, and his attack was celebrated on 4chan and Reddit, where he was named "The Supreme Gentleman" (Karacan and Crone 2020:34). New vocabulary has developed after his initials like "Going ER", aka doing an Elliot Rodger inspired act of violence, and violent incel attackers being "hERos". Places like the Reddit subpages /r/Incels and /r/TheRedPill have now been deplatformed, but many men have had the opportunity to converse with and be exposed to these misogynist online communities. Male supremacist terrorism, also called gender-based terrorism, has existed prior to the internet. In 1989, the Montreal Massacre happened where Canadian Marc Lépine killed 14 women and further injured 14 people in an act of "fighting feminism", writing in his suicide note that "I have decided to send the feminists, who have always ruined my life, to their Maker" (Bindel 2012). Though this type of male supremacist terrorism has existed prior to the internet, the executive director of the Institute for Research on Male Supremacism, Alex DiBranco, says that incel-related violence seems to be on the rise having killed somewhere along 50 people in the US and Canada since 2014 (Beckett 2021). Through internet culture, social media platforms, and 'edgy' humour, ideologies are spread and, in certain forums, normalised and the violence minimalised.

### 5.3.2 Dangerous 'Edgy' Humour and Censorship

Many right-wing male mass shooters and terrorists like Anders Breivik, Elliot Rodger, Alek Minassian, Brenton Tarrant, Patrick Crusius and more have used the internet as a central co-actor

to gain support, spread their ideologies, sharing manifestos, planning, videos or live streaming their attacks. In these, the murderers often refer to other mass murders, male-centric cyberspaces, or communities like incels, or, as seen with the Christchurch mosque shooter Brenton Tarrant, to popular and accessible content creators like PewDiePie (Felix Kjellberg) who has 110 million subscribers on YouTube (per May 2, 2021). Moments before killing 49 people and hospitalising more than 20 others The Christchurch shooter said, "Remember lads, subscribe to PewDiePie" (Cuthbertson 2019). The "Subscribe to PewDiePie"-meme was started by the channel owner himself, Kjellberg, in an attempt to maintain the channel as the number 1 subscribed YouTube channel in the world, but shortly after the Christchurch shooting, he urged it to end ("Ending the Subscribe to Pewdiepie Meme" 2019). The meme had also prior been graffitied to a World War II memorial. Whilst he, in the video where he urges the meme to be ended, strongly condemns the terror attack and his association to such act ("Ending the Subscribe to Pewdiepie Meme" 2019), Kjellberg has been known to create "bordering" content on his YouTube channel, having dressed up in military uniforms whilst watching a Hitler speech, said the N-word during a live stream, worn fake Hitler moustache, and posted a controversial video in which he had paid two Indian men from the website Fiverr[32] to hold up a sign which said, "Death to all Jews" (Mahdawi 2017). Kjellberg says these are "just jokes" and that he doesn't identify with white supremacism or other harmful ideologies, but through this, he normalises and desensitises his audience, mainly children and teenagers, with a staggering 27 billion views of his videos ("PewDiePie's YouTube Stats - Social Blade Stats" from May 2, 2021). Through this, 'edgy' humour gets normalised to a broad audience, and jokes on the expense of minorities, seen as harmless.

---

[32] A website where independent freelancers offer their services for low prices

Screenshot of Kjellberg dressed in uniform whilst listening to and showing clips from a Hitler speech from his YouTube channel PewDiePie ("I'M RACIST?" 2016)

PewDiePie is an example of the blurry line between fun content, and harmful jokes PewDiePie is, as well, such a popular YouTuber, that his content is almost guaranteed to get recommended in the algorithm, which may lead users to radical content. As seen in the mapping in figures 3 and 4, the interconnectedness of these communities is tightly knitted. Meme culture and 'edgy' internet humour are big parts of the manosphere and can act like a pipeline to channels and communities.

Users' and content creator's annoyance with censorship and "silencing" if they get deplatformed can make it attractive to migrate to smaller platforms. Here, they can speak more freely, share memes, and discuss conspiracy theories. In the illustration below (Figure 5), we see how YouTube and other popular, mainstream social media platforms can act as a pipeline to smaller platforms, which often have less (or no) moderation, and are more closed. Users can end on platforms or websites like BitTube and Nordfront which almost unrestricted and very directly spread racist, anti-Semitic, Nazi and other extreme ideologies.
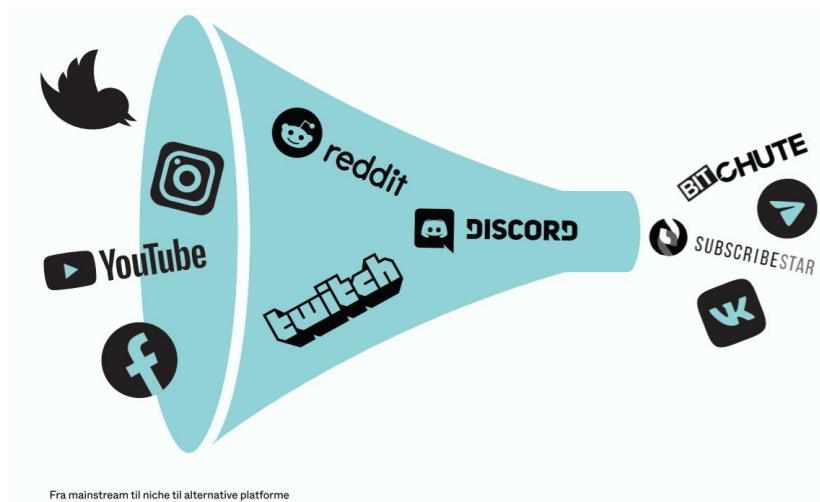
Figure 5: "From mainstream to niche to alternative platforms" (DareGender and Cybernauterne 2020:21)

The discourse on free-speech and censorship from communities in the 'far-Right' and the manosphere has made it, as Angela Nagle says in her book *Kill All Normies: Online Culture Wars from 4chan and Tumblr to Trump and the Alt-right* (2017), so that being on the 'far-Right' is "something exciting, fun and courageous for the first time since... well, possibly ever" (Nagle 2017:118). The feeling of rebellion to the "thought censorships" and (imaginary) Leftist and feminist supremacy of the 21st century seems to be a commonality for the people identifying with the so-called anti-"PC" culture. Nagle describes how abundant mocking videos of 'SJW cringe compilations' were made from YouTubers and 'alt-light' celebrities like Milo Yiannopoulos and these have built careers from "exposing" the "absurdities" of the real-life feminist political changes which have unfolded in recent years like safe space campus politics (Nagle 2017:45), or new Canadian law to respect people's pronouns. This imagery of feminist and Leftist culture consisting of "krænkelsesparate" individuals, is a politicised insult, which has entered the mainstream political scene. In recent years in Denmark, the word "krænkelsesparathed" (ready to get offended) is thrown around in the public debate as a criticism of political correctness and of those who are criticising modern-day discrimination and oppression. The discourse of trolling armies of the online manosphere has seeped into our offline lives.

## 5.4 Conclusion to Extreme YouTube Communities

In this section, research of the YouTube's communities in the manosphere has been explored, which shows that the algorithm can work in their favour, without necessarily intending to do so. The manosphere is used as an umbrella term for harmful communities and creators, often with content rooted in misogyny, but which also contains 'far-Right' and racist ideologies. The communities in the manosphere can be distinguished, but overlap, especially in how they are clustered on YouTube with the algorithm. Because of the recommendation system, users risk getting recommended videos and channels with extreme tendencies, and cases have shown that people indeed have pointed to the recommendation algorithm as having worked as an actor in radicalising them. These users risk falling into radicalising YouTube binges, called *rabbit holes*. 'Edgy' humour, dog whistles, and political grooming can make it difficult to identify harmful content. Studies of online radicalisation for children and youths indicate that many adults fail to properly identify harmful content their kids watch. The Danish and the American manosphere are intertwined, and in the worst cases, individuals linked to these conduct terror attacks. There are clear links of terrorism to the memes and language found in the manosphere, which also should be studied further in a Denmark to provide knowledge on this in a national setting. Overlooked, the misogynist, racist and discriminatory cultures and discourses found in these corners of the internet can then be a cause for real-life danger and harm. I map out terror attacks and content creators here, to stress the danger of these online communities and ideas available and accessible on YouTube.

# 6 Preventative Measures: Safer Online Spaces

In this section, I will point to different types of suggested solutions when it comes to the question of making online spaces safer and prevent the radicalisation from the online manosphere. The solutions suggested in this report are technological, societal, and governmental, and are by no means a completed list, but rather i.e., suggestions that have been referred to in literature, research and through my interviews. I will present the possible effects and concerns of different suggested prevention tools.

## 6.1 Deplatforming as a Tool

Deplatforming is the removal of accounts on social media when they break the platform rules, or community guidelines (Rogers 2020:214), and something which was discussed in the interviews I conducted, as well in other literature. Most famous is probably the deplatforming of the former U.S. President Donald Trump when his Twitter account got shut down for violating the Twitter Rules.

In the paper *Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media* (2020) Rogers describes the deplatforming as a tool which recently is "gaining attention as an antidote to the so-called toxicity of online communities and the mainstreaming of extreme speech (...) that "push the boundaries of acceptable norms of public culture" (Pohjonen and Udupa 2017 in Rogers 2020:214). In recent years, mainstream social media platforms have started to suspend and remove individuals or groups, oftentimes 'white nationalists', 'anti-Semites', 'alt-Rights', 'neo-Nazis', 'hate groups', and others (Kraus 2018; Rogers 2020; Karacan 2021). Deplatforming impacts visibility, the maintenance of fan bases, and income streams (Rogers 2020: 214), and significantly restricts the opportunity for reaching out to new audiences (Karacan 2021). Because it disrupts information flow, and online meeting places, it can make it difficult to maintain communication within the communities. Famous YouTubers such as conspiracy theorist Alex Jones, who was deplatformed from YouTube in 2018, lost his audience of 2,4 million subscribers, and the opportunity of getting his videos recommended on the largest video content site in the world. So, whilst deplatforming can heavily restrict the potentially large outreach available from the mainstream social media, how effective is it in practice, and are there unforeseen consequences with the method?

### 6.1.1 Freedom of Speech: A Fundamental Question

As mentioned above, deplatforming both restricts hateful individuals and communities from platforms and can set a tone for acceptable social life on the websites. Rogers writes that a study on deplatforming subcommunities (subreddits) on Reddit indeed "found that both the subreddits and the platform more generally saw a decline in the type of harassment found on [the now deplatformed hateful subreddits]" (Rogers 2020:215), making the overall atmosphere less toxic and hateful. But social media platforms have also received criticism for being more reactive than preventive (Karacan 2021). Oftentimes, social media will deplatform users and groups to evade legal responsibility for the content posted on their sites (Karacan 2021). Inconsistency, as described in *Existing Censorship: The SESTA/FOSTA LAW* of proper moderation leads to deplatforming of users which are not harmful. Because of moderation bots, community guidelines, and underpaid and overworked human moderators, sometimes deplatforming, banning and restrictions happen on to the 'wrong' users and groups. An example is, when Danish journalist Torben Sangilds profile got removed from Facebook and deplatformed from his site with 10,000 followers for having written a factual and critical post about the conspiracy theory and political movement QAnon (Jørgensen 2021).

If the goal is to shut down individuals and communities with concerning, extremist, or in other ways harmful discourse, language, and ideas, this raises some fundamental questions on what is allowed, and what is not. The limiting of freedom of speech is something which is often criticised as being a feminist or leftist ideology, and something which endangers society. So, whilst deplatforming does effectively shut down 'recruiting' opportunities for communities with extremist tendencies, much content lies in the grey zone of what can be deemed extreme and what cannot. In the NCPE report, the authors write that if authorities and social media want to remove and fight hate speech efficiently, it is necessary to use precise terminology so there is a possibility to legally remove content and prosecute those who commit hate speech (McGonagle 2012 in Stender Petersen and Alberg Peters 2020:22). YouTube has its defined community guidelines and rules, where some content, like nudity and violence, is removed systematically and efficiently, but content with hate speech, especially if the content creators use dog whistles and other suggestive language not considered explicit enough by moderators, will be overlooked or accepted by the platform.

When Susan Wojcicki, the CEO of YouTube, was asked about why, while YouTube had acted efficiently on removing disinformative content about the coronavirus and directing users to proper sources, they would not take the same approach on other subjects on the platform, she answered:

"We want to enable a broad range of views. We want to make sure that the full spectrum of political views or social ideas is represented there. But in this case [with the coronavirus], it's very clear that there's an authority and information behind that authority." (Susan Wojcicki in Roose et al. 2020: Episode 4)

Karacan says in her interview that it is a difficult and philosophical discussion on what should be allowed to be said online, and who should decide what is and what is not (Karacan in interview with author 2020:7-8). The strategic report plan for Denmark's technological diplomacy between 2021-2023 (Udenrigsministeriet 2021), lists several objectives planned to be worked towards within the Danish context (see *Literature Review*:9). Another of these is "[Preserving] the open, global, and free internet", and is quite interesting to examine. How is it possible to preserve an open and free internet if we also want to protect minorities and human rights as also stated as an objective in strategic report? Questions of how the internet is moderated, and who should make these choices are essential. I wrote to both the current Danish tech ambassador Anne Marie Engtoft Larsen, and later the Danish Ministry of Justice to get answers, unfortunately without any responses. I wanted to know how the responsibility between tech companies and governments is to be delegated and negotiated in terms of limiting access to platforms and free speech online, and which concrete steps have been planned to prevent radicalisation in online spaces.

The public discourse from Right-leaning communities, politicians, and certain online spaces have used the argument of "free speech" to allow racist, anti-feminist, Islamophobic and other extreme and discriminatory discourse (Nagle 2017) like the many discussions on whether Rasmus Paludan (from Hard Line party) is a "defender of the free-speech" (Skovm 2019), an argument he often makes himself as well (as seen in Mustafa 2019). Often feminism, 'SJW' and the "liberal big tech" are blamed for silencing free speech which can make spaces like the anonymous and unmoderated posting site 4chan attractive (Rogers 2020; Nagle 2017). When opinions get normalised, borders are also pushed for what is acceptable or not. When our elected officials have opinions with extremist tendencies, it can get even more difficult to acknowledge what is problematic and what is not. In her interview (Kahlke Lorentzen in interview with author, March

2021:5), Kahlke Lorentzen expresses how public discourse and opinion can shape which problems we focus on, and how:

| Kahlke Lorentzen: | He [Rasmus Paludan] has been elected, (...) and half of the political system primarily spends time talking about "Yes it's terrible what he says, but it is important that we support him in his freedom of speech", well how can you then have a serious action against right-wing extremism? Then there's this pushing of borders: "oh, but Nazis are also just *losers*", like with Nordfront - we don't really take it seriously. (...) It's like people like to call Rasmus Paludan a 'loony' instead of seeing it *as it is..* (...) Sometimes people from big political parties say things, which 15 years ago would have been categorised as neo-Nazism (...). There's been an 'erosion', so my guess would be (...) if you want to work against radicalisation and Right-extremism, then you have to look at what's being legitimized in the mainstream |
|---|---|

Without proper agreement, specified guidelines, rules, and laws to what is deemed hate speech, and what is not, restricting harmful content online will continue to be a difficult task. When communities and individuals are deplatformed, it can also reinforce a 'them-versus-us'-narrative, and make certain opinions and ideologies seem more attractive and interesting, leading to further radicalisation (Karacan 2021).

## 6.1.2 Consequences of Deplatforming

In Rogers' (2020) paper on deplatforming of extremist internet celebrities, it concludes that there is lacking research on the smaller platforms to which extreme users may turn (Rogers 2020:215), like the ones seen in Figure 5 which users can get 'pipelined' to. Some sources also suggest that it can be an advantage to not deplatform groups and individuals, and instead monitor and follow the activities, be actively present, and engage with those who post extremist content, but this suggestion is deemed to be very resource demanding (Bjørgo 2018 in Stender Petersen and Alberg Peters 2020:30). Many alternative or niche platforms are currently in use such as the Twitter alternative Gab, which has been described as a 'haven for white supremacists' and for its defence of free speech (Ohlheiser and Shapira 2018; Zannettou et al. 2018). Donald Trump migrated to Gab after having his profile banned from Twitter, and Alex Jones, deplatformed from YouTube,

now broadcasts his shows on InfoWars.com. In a way, social media somewhat decentralises, and as Rogers asks: does it prompt the individuals to migrate to other platforms with more welcoming and 'oxygen-giving' extreme spaces? Does the hate speech intensify, and lead to more harmful communities and radicalising forums? (Rogers 2020:215). A research on Gab concludes that the hate speech is 2,4 times higher than on Twitter, although still 2,2 times lower than on 4chan's "politically incorrect board" /pol/ (Zannettou et al. 2018). But although some users might migrate, it is not everyone who does so. Karacan explain in our interview (Karacan in interview with author, February 2021:10):

| Karacan: | When they [the content creators] get deplatformed they might migrate to BitTube, and it just requires much more for the audience: that you are a follower, and you *then* 'move' to BitTube. It will be more difficult for you, and it is a platform you don't know, so you really have to be dedicated to following them there. And so, they lose quite a lot, and that helps. |
|---|---|

Deplatformed YouTubers like Alex Jones and Milo Yiannopoulos have reported a significant drop in their visibility, audience and income streams, and Yiannopoulos even claimed to have gone bankrupt by deplatforming (Rogers 2020:214).

## 6.1.3 Reflections on Deplatforming

Deplatforming can be a tool to both demonstrate a shift in what is considered acceptable on social media (Rogers 2020:226), and to limit outreach for harmful groups and individuals, disrupt information flow, income and create an overall less-harmful atmosphere on platforms which successfully deplatform content (Rogers 2020). At the same time, it should be more preventative, and less reactive (Karacan 2021), since deplatforming also can cause users to follow celebrities or large communities to alternative platforms (like incels.net which exists as an alternative to the deplatformed subreddit /r/incels), exposing users to get more radicalised (Rogers 2020; Zannettou et al. 2018). There is still a need for further research on users who migrate to alternative and niche platforms, and deplatforming can make it more difficult to monitor and research communities and individuals with extremist tendencies (Stender Petersen and Alberg Peters 2020). Rogers writes:

"For those arguing that it does not work, deplatforming is said to draw attention to suppressed materials (Streisand effect), harden the conviction of the followers, and put social media companies in the position of an arbiter of speech. For those arguing that deplatforming is effective, it is said that it detoxes both subspaces (such as subreddits) as well as platforms more generally, produces a decline in audience and drives extreme voices to spaces that have less oxygen-giving capacity, thereby containing their impact."

(Rogers 2020:215)

It is difficult to measure or conclude how well deplatforming works as a preventative tool for online radicalisation, which further research might be able to answer more clearly. Literature is critically discussing whether the responsibility of deplatforming or removal of content online lies with the social media or the authorities (Stender Petersen and Alberg Peters 2020:29), which also needs to be delegated properly or discussed further. At the same time, deplatformed individuals or groups have the opportunity of creating new accounts shortly after being deplatformed, which is also difficult to restrict or control due to data protection laws (Hjortkjær Henningsen 2021).

## 6.2 Responsible Algorithms, Transparency and Law Regulations

Whilst deplatforming is a tool which relies on moderation bots, human moderators, and possibly cultural norms and definitions of what is appropriate content, the machine learning algorithms affecting the behaviour of users of sites, is a different matter. So how can we create or monitor machine learning algorithms to make them less harmful?

As shown throughout this report, it is difficult to analyse and criticize algorithms and algorithmic biases because of the lack of transparency to access the actual codes. This means that researchers or authorities cannot access and test the effects of algorithms that are in use. Tech companies can deny allegations and responsibility when faced with criticism, like spokespeople from YouTube have done when being confronted with allegations of biases or discrimination on their platforms (Høst 2019; Farokhmanesh 2018; Bryant 2020). In Shoshanna Zuboff's *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, she writes that one expert (Tutt 2016) proposes the creation of a government agency to "oversee the development, distribution, sale, and use of complex algorithms", arguing that existing laws "will prove no match for the difficult regulatory puzzles algorithms pose" (Zuboff 2019:484). In Høst (2019: 20:08) digital culture expert Katrine K. Pedersen describes the current age of social media as an era

where we do not really act preventatively and lack the proper legislation to manage what issues exist in handling the social media space. To make machine learning algorithms more ethical, how do we want to control and create them?

## 6.2.1 Government Agencies and Law Suggestions

"Algorithmic regulation will require federal uniformity, expert judgment, political independence, and pre-market review to prevent—without stifling innovation—the introduction of unacceptably dangerous algorithms into the market." (Tutt 2016:83)

As of now, the insights and problem solving of machine learning algorithms are in the hands of the companies which own them. Unforeseen challenges caused by algorithms are, as shown in this report, a current matter which needs tendance to. Since machine learning algorithms also affect other parts of new societal technological solutions outside of the social media sphere, Andrew Tutt suggests having a dedicated governmental agency which supervises the development, deployment and use of algorithms (Tutt 2016:90). His article concludes that "regardless of the path we take, there is now a need to think seriously about the future of algorithms and the unique threats they pose." (Tutt 2016:91). In the strategic report plan for Denmark's technological diplomacy between 2021-2023, there is also criticism of the little transparency that authorities and governments have to the algorithms used on social media (Udenrigsministeriet 2021:2;5). The front person for TechDK Kommissionen, Stine Bosse, says that there is a need for far stronger law regulations and resources managing how algorithms are used today (Overgaard 2020). Although we do have Datatilsynet (The Data Protection Authority) in Denmark, and the European Data Protection Board (EDPB) within EU, Stine Bosse says that there is a need for the same kind of attention on AI (machine learning algorithms etc.), as Finanstilsynet, Danish Financial Supervisory Authority, has over the financial sector (Overgaard 2020).

In the TechDK commission's report on democratic interventions for the challenges caused by digital technologies, a variety of legal suggestions and changes are recommended (TechDK Kommissionen 2020). Amongst these are:

1) Inspecting algorithms like with public employees
2) Creating transparent collaboration between social media, researchers, and journalists

Both suggestions demand legal specifications and changes to combat issues of the lack of control and transparency of machine learning algorithms. The first suggestion proposes that it should be ensured that algorithms do not incorporate data or apply logics which are not acknowledged by the democratically legitimized state power (TechDK Kommissionen 2020:17). To do this, the commission suggests a liability system where algorithms, just as public employees, can be prosecuted if they repeatedly commit serious acts of misconduct or neglect the rules. This would mean the creation of a "new, clear and easily-understandable" liability system to ensure the proper care and accountability which is applicable for algorithms (TechDK Kommissionen 2020:17). Restricting or controlling algorithms will be a resourceful task and finding ways to avoid unexpected and unwanted side effects will demand knowledge, testing, and financial resources. The second suggestion proposes access to data and new partnerships sharing and mobilising knowledge, expertise, technology, and financial resources (TechDK Kommissionen 2020:30). Journalists and independent researchers must be able to access what is now protected as tech companies "trade secrets" to measure and research aspects of the platforms. The commission mentions that a complication can be the different data protection laws, ensuring the rights of privacy for the users on the site.

By now, there is no definitive answers to which laws or regulations are needed to combat the unforeseen consequences and misuses of machine learning algorithms, but what several sources (Tutt 2016; Høst 2020; TechDK Kommissionen 2020; Zuboff 2019) suggest are some forms of legal specifications, and legal, transnational, and governmental focus on regulating machine learning algorithms. Tutt (2016) stresses that with new and future technologies depending on machine learning algorithms to make critical decisions (self-driving cars, medical examiner bots like Watson, and more), society is proceeding with an inevitable need to discuss and decide how we want to manage these new decision-making technologies.

## 6.2.2 Interdisciplinary and Diverse Code Writings

As the TechDK commission suggests, a collaborative and interdisciplinary approach to monitoring and checking algorithms could be beneficial. The term interdisciplinary refers to "problem-solving activities that involve interactively, and to some extend [sic] integrate, at least two different disciplinary perspectives to the problem at hand" (Klein 2010:17 in Børsen 2013:38) and combines different disciplines' conceptual and theoretical frameworks to solve and analyse problems. Some sources in my literature suggest that an interdisciplinary approach could be beneficial to create algorithms which are more responsible, reflective, and more well-functioning (Høst 2020; Noble 2018). The discussion of morality in technology is broad and philosophical, and I argue that technology cannot be neutral, but that the effects and consequences of (any) technology lie in the hands of those who use them, and therefore can be used and misused as pleased by different actors, either intentionally or unintentionally. We can therefore strive to create algorithms which are reflected upon and analysed from many different angles and tested to see which unforeseen consequences they have.

An interdisciplinary approach can therefore help analyse these issues from many different angles. Danish digital culture expert Katrine K. Pedersen suggests that the YouTube algorithm would look different if the coding room had had teachers, pedagogues and other humanities professionals involved in the process to avoid especially young kids getting unwanted consequences from watching YouTube (Høst 2020). In the interview, she speaks about the decision-making process of choosing videos that young kids are presented with on the platform, even on YouTube Kids, which may not be psychologically healthy at such a young age. Perhaps the in-design and recommendation system would be different with an interdisciplinary approach to optimise YouTube for younger children. In *Algorithms of Oppression*, discussing how search engine algorithms and other machine learning algorithms disadvantage racial minorities, especially Black women, author Safiya Umoja Noble writes:

"This book opens up new lines of inquiry using what I believe can be a black feminist technology studies (BFTS) approach to Internet research. BFTS could be theorized as an epistemological approach to researching gendered and racialized identities in digital and analogue media studies, and it offers a new lens for exploring power as mediated by intersectional identities." (Noble 2018:171-172)

Including minorities and people affected by oppression and discrimination in society and mainstream media, which Noble tries to show are instantiated in digital algorithms (Noble 2018:171), could create an opportunity to analyse and shine a light on the different unintended side effects in machine learning algorithms more efficiently. It can, though, be at a disadvantage for the absolute goals that tech companies want their algorithms to do on the platforms. Examples can be if YouTube restricted watch time, suggested and/or insisted on showing counter-narratives to users or directing them to authoritarian informative sites, or not letting kids navigate the site themselves on YouTube Kids. This would probably decrease the overall watch time from users, which conflicts with the overall goal maximising the users' watch time.

It could be assumed that there would have to be some law regulations for this as well if we as a society wanted to enforce transdisciplinary creation and maintenance of machine learning algorithms. If tech companies were to include interdisciplinary and diverse teams in their coding processes, and the changes suggested or needed to make less harmful machine learning algorithms, there would probably also be a need for legislation to demand the changes in order to value change over profit.

## 6.2.3 Report Mechanisms

So, what could happen, if authoritarian agencies were established, either national, transnational, or global, and firm and concrete laws created? These institutions should be able to research allegations and ensure changes where necessary. This could potentially create user involvement too, since we, as users, are forced to give tacit permission to the ways in which machine learning algorithms work as of now. Currently, if users are dissatisfied or concerned about the machine learning algorithms on social media, we do not have much choice but to try to alter the way the algorithms affect them through various user interventions (to the extent possible). This can be through browser add-ons, changing profile information, using code language in order not to be flagged or getting posts deleted, and more. But even through these interventions, users are still left almost completely uninvolved with the way that social media works.

If other influential bodies had the agency to involve, change and test algorithms and their effects on users and the platforms, there could be established proper responsibility from report mechanisms, so users, institutions, or nations, could make sure that they live up agreed upon demands. Zuboff sums this up effectively as:

"A century ago, workers organized for collective action and ultimately tipped the scales of power, and today's "users" will have to mobilize in new ways that reflect our own unique twenty-first-century "conditions of existence." We need synthetic declarations that are institutionalized in new centres of democratic power, expertise, and contest that challenge today's asymmetries of knowledge and power. This quality of collective action will be required if we are finally to replace lawlessness with laws that assert the right to sanctuary and the right to the future tense as essential for effective human life." (Zuboff 2019:485)

Zuboff here argues that collective action can demands for better circumstances for those affected, and that today's users could do the same for the today's technologies.

## 6.2.4 Reflections on Law and Government Control

The current law regulation and governance of machine learning algorithms, and democratic demands for these to work with ethical reflections are still found lacking and unclear by several sources presented in this section. It is said that the tech industry generally develops faster than the law (TechDK Kommissionen 2020:33), and CEOs from tech companies in Silicon Valley like former Intel CEO Andy Grove, and former Google CEO Eric Schmidt have claimed that high tech runs three-times faster than normal business, and the government runs three-times slower than normal business, creating a nine-times gap (Zuboff 2019:104). This is used as an argument angled as a hindrance to technological evolution. *Business Insider* covered Schmidt's remarks and wrote "When asked about government regulation, Schmidt also said that technology moves so fast that governments really shouldn't try to regulate it because it will change too fast, and any problem will be solved by technology. "We'll move much faster than any government," Schmidt said." (Gobry 2011). Intertwined with the financial success of tech companies, current legislation and the continuous need for research and knowledge of the possible consequences of our digital lives, there will no doubt be difficulty establishing further legislative framework in Denmark. The Danish TechDK commission still stresses that it is "crucial that legislation not only accommodates favourable conditions for the tech industry and surveillance capitalist companies that want to digitize more and faster", but works to bolster democracy, and consider the problems which follow with rapid technological development. (TechDK Kommissionen 2020:33).

User involvement and (more) interdisciplinary involvement involved in managing, creating, and analysing the algorithms are also to be recommended, to get less harmful machine learning algorithms on social media sites. As Harding's feminist standpoint epistemology argues, starting

from marginalised lives help critically and systematically interrogate the unchecked effects and consequences of problems (Harding 1992:442;462), meaning to incorporate critical voices in the processes of analysing these technologies. There are indeed no concrete or absolute answers on how to create governmental bodies or other institutional authorities to monitor the machine learning algorithms used on social media, and the rest of society, but an urgent need to analyse and critically assess the existing and future machine learning algorithms.

## 6.3 Out of the Rabbit Hole: Dialogues and Knowledge

Previous radicalised individuals in the literature I use, and by Kalhke Lortenzen and Karacan have in interviews listed reasons on how they became deradicalised. Moving on from the two previous sections, which focused on technological and legislative preventative measures, this section will focus on what can be done on a more social level: locally and individually. These tools and methods aim to get people out of the rabbit hole, or ultimately prevent them from 'falling down'. The suggestions here are also some which can be more easily implemented in comparison to building legislative bodies and frameworks. Naturally, the reasons for radicalisation and the radicalisation process itself are situational and varies from person to person, and there are no absolute measures that guarantee deradicalisation, but there are preventative methods that helps deradicalising *some* individuals.

### 6.3.1 Counter-narratives

Being lured into communities with extremist tendencies and following individuals or being part of groups where harmful discourse, humour and ideologies are being normalised, seem to be big factors of radicalised individuals (DareGender and Cybernauterne 2020). Interviewees from both Roose et al. (2020) and DareGender and Cybernauterne (2020) who had themselves been radicalised or falling into rabbit holes, sometimes point to counter-narratives or critical dialogues between opposing 'parts' as reasons to why they got out of the rabbit hole and deradicalised. An example is Caleb Cain, presented earlier in this report, who had gotten radicalised through YouTube and its recommendation algorithm (Roose et al. 2020: Episode 3):

| Kevin Roose: | Yeah, so when did you feel your views shifting again? |
| Caleb Cain: | So, a big thing was Destiny |
| Kevin Roose: | Can you just explain who Destiny is? |
| Caleb Cain: | Destiny is a Twitch streamer |

Destiny, real name Steven Kenneth Bonnell II, streams himself playing video games through Twitch and uploads the videos to YouTube. In these, he also sometimes invites people on his channel, and while playing video games, debates them. He is known for debating several prominent individuals known from the manosphere and the 'far-Right' communities of YouTube, and for *owning* them: presenting factual arguments, and at the same time as making fun of the opponents, and expose their hypocrisy, dis- or misinformative arguments and arguments used to justify the upholding of discriminatory beliefs. Both Cain and another former radicalised person from DareGender and Cybernauternes' report (2020:28) mention Destiny as their way out of the rabbit hole. Because he presents counter-narrative and counterarguments to the ideologies and discourse presented in the manosphere communities on YouTube and has prominent individuals in his shows, his videos attract viewers. This may also make the recommendation algorithm link Destiny's videos to videos in the manosphere. DareGender and Cybernauterne's study points to counter-narratives and open dialogues with people from "outside" the rabbit hole as being a recurrent reason for individuals becoming deradicalised (DareGender and Cybernauterne 2020:27-28).

When presenting counter-narratives and counter-arguments studies point to several factors which need to be considered. The arguments of counter-narratives must be exact and know the audience of the radicalised group addressed including the vulnerabilities, frustrations and feelings of marginalisation found in individuals who engage in them, not speak into an "us-versus-them"-narrative and need trust to those presenting the counter-narratives (Stender Petersen & Alberg Peters 2020:31). Studies also point to avoiding direct counterpropaganda which either has no effect or further radicalises into extremism (Gemmerli 2015). In the case of the Twitch streamer Destiny, he films with individuals where both parts are prepared and have a 'fair' discussion, which is still playful and respectful. Both parts are engaged, and at times the prominent Right-wing individuals do change their minds or agree with arguments presented by Destiny, who is Left-leaning, which means that radicalised viewers hear more complex narratives and ideologies from

people they trust. Speaking to Kahlke Lorentzen about content creators like Destiny, who 'dedicate' themselves to creating dialogue in a time often known as one of the most polarised political eras in the West, she talks about why there may not be many of such influences generally, especially in the Danish manosphere (Kahlke Lorentzen in interview with author, March 2021:18):

| Kahlke Lorentzen: | I think it can be rough; researching, debunking things - it takes longer time than writing some bullshit [without factual claims]. The thing about looking at something and asking: "What is it this person is saying here? Where is the dog whistle? What do the statistics *actually* say? (...) And like he [Hbomberguy, another "debunk" debater] says, it takes him a month to make those videos if they must be as precise as they should. And if you really want to debunk something, you really must know your data. So maybe there is a need for a Destiny [in the Danish manosphere], but maybe there's also a need for a research team. |

## 6.3.2 Local Initiatives

In the NCPR research on the radicalisation of children and youths, their report points to local actors to be part of the preventative methods. When talking about local actors and local communities, it is defined as municipalities, local societies, city areas, specific housing areas, schools, and individual actors within these such as teachers, pedagogues, friends, family, and the like (Stender Petersen & Alberg Peters 2020:39). These can both help children and youths find (non-harmful) communities to be a part of, and act like trusted voices in the local areas. The NCPR report writes that the Danish Social- and Integrationsministeriet (The now Ministry of Social Affairs and Senior Citizens and the Ministry of Immigration and Integration) points to the collaboration between youths, home, school, and after-school initiatives as a decisive effect for fruitful preventative work (Social- og Integrationsministeriet 2012). A study (Staniforth and Nitsch 2014) analyses a concrete method of educating local voices, which can be positive and strong representatives for their communities and communicate positive and trustworthy messages and counter-narratives (Stender Petersen & Alberg Peters 2020:40). Stainforth and Nitsch (2014:554) also find that local initiatives like strengthening local voices can be a part of building different online communities' resilience, meant as resistance to hateful and harmful online rhetoric. The Aarhus model in Denmark works similarly with radicalised people to guide and

mentor them back into non-harmful ideologies and create healthy communities locally which individuals can engage in and be a part of. As described, I have tried throughout the duration of this thesis, to contact Aarhus Municipality to explore how their exit-program and mentoring program work for people radicalised in the online manosphere works but without answer, it is difficult to conclude and analyse the existing measures.

### 6.3.3 Expert Reliance

To guide local initiatives, or even to guide the government in this field, the question of expert status arose in my research. Because the researchers studying this, often are what Haraway describes as *marked bodies*, I asked Kahlke Lorentzen in our interview on whether she had the feeling about their organisation holding expert status. As explained in *Objectivity and Standpoint*, there are several obstacles that can be described as *minority taxation*, when working from a framework or identity deemed *marked*, and therefore biased. In this conversation, it is seen how the expectation of objectivity can hinder research (Kahlke Lorentzen in interview with author, March 2021:13):

| Author: | Do you have any expert status from the government's side? Is there an attitude that you are respectable experts? |
|---|---|
| Kahlke Lorentzen: | We were in a public procurement notice process, and it was our clear impression that we were too political to give us the assignment, despite it being an area that we have worked with for a long time (...) We are Leftist. It's not something we write on our website, but it's evident: it's not something that we hide either. (...) We aren't neutral, but we can still make good research. Nobody's neutral. |

In this case, it can be described as getting *minority taxed*, somewhat in measurable ways (of extra work trying to get an assignment, worse feedback, and scepticism), and in some unclear ways (feeling discriminated against, feeling discredited). In another case, spoken about in the interview, Kahlke Lorentzen speaks on the lack of proper research of the manosphere and the scepticism of those researching it (Kahlke Lorentzen in interview with author, March 2021:3):

| Kahlke Lorentzen: | When I get asked if I can refer to a researcher who specialises in this [the manosphere], there are incredibly few, but at the same time many who focus on Islamic State or other types of Islamic radicalisation, which naturally is equally important, but there has not been a focus [on the manosphere]. I had a chocking conversation with someone who researches radicalisation in Denmark, and he did not know i.e., Nordfront and asked me to map out different Danish actors, and I think that says something about how radicalisation and extremism are researched within a Danish context. (...) Redox [a Leftist, anti-fascist research site] are the only one who has mapped these movements in Denmark properly, but every time I refer [other] journalists to their work, they don't want to use Redox as a reference, because they are seen as 'Left-extremist', even though they are part of the Danish Press Council and make good journalism. |
|---|---|

It can be difficult to go "against" public discourse and normativity, and as affiliation with strong, political views can be seen as a hinderance for objectivity or "good" research, as Kahlke Lorentzen describes in this example. When feminism, or other kinds of searches for egalitarianism are seen as threats, and when public discourse follows, it makes it difficult for experts who work with these frameworks to get funding and acknowledgement. As seen in the study on the news media reporting on terrorism in Denmark (Parker 2018), male supremacist terrorism has tendency to be framed as a "lone acting" and blamed on mental illness, compared to Islamic terrorism which often is framed as organised and affiliated with a community. Although there is a rise in research on the online manosphere, and the danger for radicalisation, extremism, and terrorism in a Danish setting, it is still new, and though there are suggestions for preventative methods, there is still lacking implementation of preventative methods suggested in literature, and evaluation of these when they have been implemented. If we need to educate local workers or spread awareness through national campaigns, there is a need for trusted actors or bodies to convey the messages, and a need for continuous research to be funded, and research findings considered by authorities.

### 6.3.4 Expanded Understanding of Gender

Another suggestion from the research project, *The Angry Internet* (Mogensen and Helding Rand 2020), assessing the extent of online misogyny and anti-feminism in the Nordic countries suggests cultural change and education. Here, they recommend expanding ideas and understanding of gender, and teaching it to children from an early age. They write that this can be achieved through a greater focus on gender, gender identity, gender development, and gender expression, and having these themes as mandatory courses in the education of teachers and other professionals working with children and youth (Mogensen and Helding Rand 2020:32). Furthermore, they write that the concept of masculinity and "masculinity ideal" should be wider and more inclusive, breaking down the relatively constrictive and stereotypical ways of "being a man" that boys and young men are currently presented with (Mogensen and Helding Rand 2020:32). Through creating cultural changes and breaking down ideas of a dichotomy between masculinity and femininity, there is an opportunity to combat the underlying issues of male loneliness, isolation, and social inability to express 'vulnerable' emotions (Mogensen and Helding Rand 2020:33), which may lead men into radicalisation and harmful online communities. This could possibly also teach youth to be critical of harmful narratives found in online communities. Here, anti-feminist discourse though has to be challenged, and this would in some areas also demand a cultural shift, which may be a resourceful task.

### 6.3.5 Reflections on Societal Initiatives

Social and local initiatives can build a bridge between the online world and the offline world where issues like harmful online communities, 'edgy' humour, and other aspects of the online manosphere are taken seriously, investigated, and intervened with. There are both need for social interventions, but also more research and education on how these online cultures look, so this can be mediated to professionals or the public to spread awareness. As seen throughout this thesis, harmful aspects of the cultures found in the online manosphere might not be transparent to 'outsiders' looking in, and therefore there are needs to educate those who might be subjected to online radicalisation of the manosphere, and those helping individuals to either getting deradicalized or preventing them from radicalisation. A combination of technological, governmental, and social solutions will probably work best as a preventative measure, so most people can be reached both online and offline.

## 6.4 Conclusion to Prevention Strategies

In this section, I have presented and analysed suggestions which can help combatting radicalisation in YouTube's manosphere. The suggested preventative measures are both be technical, legislative, and social, but should ideally be implemented collaboratively. Literature researching preventative measures for online radicalisation point to there still being a need for further research in the field. There is a gap in research on preventative measures for the radicalising manosphere, and a need for further research in a Danish setting as well. This thesis argues that proper prevention of radicalisation in online spaces should start on a governmental level, which defines male-supremacist online radicalisation and relies on and funds expert knowledge to continue the research in this problem field. The lack of legislation and governmental control of tech companies' machine learning algorithmic infrastructures on social media make it hard, if not impossible, to demand changes to the sites. Some social preventative measures can more easily be implemented, like education and counter-narratives both online and offline, but there is no guarantee that this will reach the users who are socially isolated, outside the education system, or who are already radicalised and engaging in online communities with extremist tendencies.

# 7 Discussion and Reflections on Positioning

In this thesis, the current issue of male-supremacist violence and extremism, stemming from online radicalisation have been examined. Here, I have focused on YouTube as a platform, due to its specific machine learning algorithm and the allegations of its bias, often seen in mainstream media and other studies. YouTube has, as well, been listed as a starting point for people who have been radicalised or engaged in communities with extremist tendencies in the manosphere. This thesis' collecting knowledge on the technicalities of the machine learning algorithm on YouTube, online radicalisation in the manosphere, and combining these with concise suggestions of preventative measures is, to my knowledge, the first study which does so from a Danish setting.

Other studies have previously theorised different definitions on radicalisation, terrorism, and extremism, which can lack the proper nuances needed to theorise these issues within the manosphere. Though this thesis does give a simple theoretical framework to limit the problem field, further theorisation, and clear definitions of 'Right-extremism' and the manosphere radicalisation are necessary. Here, it is needed to theorise how to best draw lines between acceptable ideologies and those which are dangerous and harmful. At the same time, this thesis has focused solely on male-supremacist radicalisation and terrorism, whereas studies could also focus on 'Left-extremist' or feminist radicalisation. I will though argue that intersectional feminist and certain branches of 'Leftist' radicalisation most often are for the better and speak into Harding's feminist standpoint epistemology where problems are defined from marginalised voices. Of course, not all Leftists are feminists, and Leftism is also not a monolith, easily defined. The topic of the manosphere, anti-feminism, and male-supremacist terrorism has been chosen, because it, from my perspective, is urgent and necessary.

Over the duration of my thesis, from February-June 2021, a lot of new research have seemed to emerge, or studies which had only just been released. My initial hypothesis of there not being any relevant studies or reports was proven wrong, but still, literature and strategies were difficult to find, especially from governmental sources. Though there are national initiatives, these seemed to have either specific focuses (like juvenile radicalisation) and existing literature often gave broad and unconcise suggestions and lacked clear strategic plans or specifications. Many of these, (Stender Petersen and Alberg Peters 2020; Mogensen and Helding Rand 2020; TechDK

Kommissionen 2020) mention several preventative measures, but fail to expand on, nor analyse these through implementations or trials. This also points to expert knowledge being overlooked, or preventative measures failing to get implemented.

I see that there are lacking strategies and governmental knowledge on male-supremacist radicalisation, at least from what has been able to be studied during my thesis research. This does not mean that these do not exist, but that they have not been able to find in the duration of this thesis. There are many preventative suggestions in different reports and studies, and my thesis has solely focused on some of those released from papers in either Denmark or the U.S. Especially with the technical suggestions to preventative measures, further research in collaboration with machine learning engineers could give deeper insights to the ways algorithms can work more responsibly. As there are reports (like the NCPE's by Stender Petersen and Alberg Peters 2020), which focus on mapping, analysing, and discussing various preventative measures, my thesis did not aim to reproduce a similar listing of all (or the many) possible solutions, but aimed to address the specific issue of the online manosphere, the role of machine learning algorithms in radicalisation processes, and what can be done in a national setting. Here, knowledge on radicalisation from algorithms, memes, dog whistles, and anti-feminist rhetoric seems overlooked in research within a Danish scope, and therefore only contains little research on. Although some studies on this from the U.S. could be transferable, the solutions may not.

When much new research is emerging in this problem field, I feel that techno-anthropology adds a unique strength to assessing these problems. Mediating and building bridges between different disciplines are necessary when having to navigate a complex problem, which this is. There are no simple answers to this thesis' problem field, as solutions probably are many, and may look different all depending on where they are implemented: in legislation, politics, institutions, schools, platform designs, machine learning designs, etc.

The knowledge gap which seems to have been highlighted mostly in the literature used in this thesis about preventative measures for radicalisation, have been that from a post-implementing phase. Implementation of concrete measures to prevent online radicalisation of the manosphere is lacking, or non-transparent, and therefore difficult to research. Outside the corona pandemic, it might have been possible to conduct field work with the few institutions which have exit-

programs in order to study their methods for specifically male-supremacist radicalised individuals and add knowledge to this field.

Research from literature used in my thesis have both included data from digital methods, scraping, and collecting data from YouTube and their communities (AlgoTransperency; Kaiser and Rauchfleisch 2018), and qualitative data from interviews of radicalised individuals (Roose et al. 2020; DareGender and Cybernauterne 2020; amongst knowledge from Karacan), but further research in a Danish context is still needed. Especially the digital data scraping in a Danish setting similar to the Kaiser and Rauchfleisch study, could map the Danish manosphere on YouTube. Generally, more knowledge could be expanded of the Danish digital cultures in the manosphere and Danish male-supremacist radicalisation. Studies on which knowledge social workers, teachers, youth workers, and other relevant social professionals have and need of online radicalisation could also help map out what is needed to detect radicalised individuals early in their process. Knowledge about this, I would argue, should also be publicly accessible so parents, family, and friends can detect and help their loved ones.

I feel that this thesis, and research on male-supremacist radicalisation and terrorism add knowledge and shine lights on issues which are, despite the current momentum of (often) independent researchers or research institutions, still heavily overlooked. The threats from anti-feminist discourse, ideologies, and politics are some which have real life effects and consequences on those of us who are marginalised, and who stand up against oppression. Therefore, I could also argue that Harding's point on starting from marginalised voices to analyse and identify problems does in fact produce 'good' science, and that examining these problems are only possible with the help and involvement from marginalised individuals and communities. To examine male-supremacy, misogyny, anti-feminism, and algorithmic biases, we exactly must listen to the voices of those affected by their effects. Through funding research, and trusting experts with knowledge on this, there would also be many different narratives of *situated* studies and research, which will be beneficial to shining lights on the many issues in this problem field, and the many possible solutions needed.

Lastly, I will say that as much relevance the issue of incel-related terrorism and radicalisation have, as well as from other harmful communities in the manosphere, there is an equal necessity to

continue and conduct research in racist, transphobic, Islamophobic and other online (and offline) communities in which there are a rising of hate speech and radicalisation.

# 8 Conclusion

This thesis has studied how the YouTube algorithm radicalise in anti-feminist ways, and which preventative solutions can be made in Denmark. *Anti-feminism* has been theorised as an ideology found in the manosphere which frames feminism as, amongst other things, threatening free speech, society, the traditional gender roles and family constellations. Especially the threat of masculinity is seen in anti-feminist ideologies in the manosphere. The *manosphere* has, at the same time, been used as a broad term for the communities existing in a YouTube cluster where misogyny, racism, 'far-Right' ideologies, and other content with extremist tendencies exist. These communities all contain anti-feminist rhetoric, discourse, and ideologies. Especially the groups and individuals characterised as *incels* have been highlighted, as they have been connected to several male-supremacist terror attacks, specifically targeting women and feminists, often due to a frustration of being involuntarily celibate, misogyny, and anti-feminism.

YouTube's recommendation algorithm has been analysed, and though it cannot be proven that the algorithm has a clear bias towards anti-feminist content, the algorithm does unintentionally seem to guide users towards communities with extremist tendencies, most strongly seen from the starting point of mainstream political or news channels like CNN, Fox News, and Barack Obama. The algorithm is designed to maximise users' viewing time and when users get directed towards content with extremist tendencies, they often are recommended more of the same content, creating a filter bubble, which can make them fall into *rabbit holes* and radicalising them. With few counter narratives, which for some individuals can deradicalize and possibly prevent radicalisation, the algorithm can be seen as an actor in the radicalisation process. Moderation on the site not being as efficient as it could, and at times targeting marginalised users and not effectively removing harmful content, as for example in the manosphere, may also create a disadvantage for counter-narratives on YouTube.

Analysing YouTube's machine learning algorithm, also called the *Deep Neural Network*, is not accessible as a researcher. Therefore we, for now, must rely on an outdated paper provided by Google itself from 2016 (Covington, Adams and Sargin) describing the design of the algorithm. For now, algorithmic simulations can provide knowledge and analysis of YouTube's recommendation algorithm. Through using literature from strategic reports and national

research institutions and centres, the thesis has found that there exist many different suggestions for preventative measures, especially pedagogical (as seen with the NCPE report by Stender Petersen and Albert Pers (2020). Many of these measures are though yet to be properly implemented and monitored afterwards. The legislation on algorithms is also found lacking, which makes the assessment and monitoring of machine learning algorithms difficult, if not impossible due to them being labelled as 'trade secrets' from the tech companies owning and using them. It has not been possible during the scope of this thesis to get proper answers from governmental bodies as to what they are planning or working towards for legislative or governmental measures and solutions. Throughout this thesis, the Ministry of Justice, The Danish Tech Ambassador, and Aarhus' Municipality's office for efforts against radicalisation have been contacted, unfortunately without providing further knowledge. It is therefore advised that legislative measures and governmental bodies should ensure ability to monitor, test, and analyse algorithms to ensure or working towards less harmful algorithms.

At the same time, there also seems to be a need for governmental definitions of radicalisation, which may help identify individuals who are radicalised, or in risk of getting radicalised, possibly also clear definitions solely targeting male-supremacist radicalisation. The dilemma of limiting free speech and *deplatforming* individuals and groups is also something which needs more research and assessment since there are both benefits, risks, and challenges with these choices and issues correlating to male-supremacist radicalisation.

This thesis has aimed to provide knowledge of the danger of the online manosphere, amongst suggestions analysed from an interdisciplinary approach, which can be expanded on further through research and implementation of suggested preventative measures. Though the findings indicate that there are many different suggestions to preventative measures, it also raises questions of how and which of these should be implemented, and how to follow up afterwards, and measure the changes. A combination of governmental, technological, and social initiatives is therefore to be recommended to ensure the most efficient preventative strategies as it looks now.

# References

## Literature

Andersen, Niels Åkerstrøm. 1999. *Diskursive Analysestrategier: Foucault, Koselleck, Laclau, Luhmann.* København: Nyt Fra Samfundsvidenskaberne. https://doi.org/10.7146/politica.v32i3.68399

Berwick (Breivik), Andrew. 2011. *2083: A European Declaration of Independence.* https://info.publicintelligence.net/AndersBehringBreivikManifesto.pdf

Bjørgo, Tore, ed. 2018. "Høyreekstremisme I Norge: Utviklingstrekk, Konspirasjonsteorier Og Forebyggingsstrategier." PHS Forskning; 2018: 4: Politihøgskolen. https://phs.brage.unit.no/phs-xmlui/handle/11250/2568904

Bryant, Lauren Valentino. 2020. "The YouTube Algorithm and the Alt-Right Filter Bubble." *Open Information Science* 4 (1): 85–90. https://doi.org/10.1515/opis-2020-0007

Børsen, Tom. 2013. "Chapter 1: Identifying Interdisciplinary Core Compe-Tencies in Techno-Anthropology: Interactional Expertise, Social Responsibility Competence, and Skills in Anthropology-Driven Design." In *What Is Techno-Anthropology?*, edited by Tom Børsen and Lars Botin, 35–66. Aalborg: Aalborg Universitetsforlag.

Covington, Paul, Jay Adams, and Emre Sargin. 2016. "Deep Neural Networks for YouTube Recommendations." *Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16.* https://doi.org/10.1145/2959100.2959190

Crenshaw, Kimberlé. 1989. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics," *University of Chicago Legal Forum*: Vol. 1989 , Article 8. 139-167.

# References

https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1052&context=uclf

DareGender, and Cybernauterne. 2020. *Under Indflydelse - Veje Ind I Ekstreme Digitale Fællesskaber Gennem Køn Og Maskulinitet.* https://cybernauterne.dk/blog/under-indflydelse-rapport-ekstreme-digitale-faellesskaber/

Domingos, Pedro. 2015. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World.* Penguin Books. https://doi.org/10.5860/choice.194685

Harding, Sandra. 1992. "*Rethinking Standpoint Epistemology: What Is "Strong Objectivity?"* Michigan State University Press. *The Centennial Review* 36 (3): 437–70.

Haraway, Donna. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." Feminist Studies 14 (3): 575–99. https://doi.org/10.2307/3178066

Karacan, Telli Betül, and Manni Crone. 2020. "Incels, Kvindehad Og Den Nye Antifeminisme". DIIS. *Udenrigs* (August): 32–36. https://pure.diis.dk/ws/files/3975746/Udenrigs_2_2020_Incels_kvindehad_og_den_nye_antifeminisme.pdf

Kezar, Adrianna. 2003. "Transformational Elite Interviews: Principles and Problems." *Qualitative Inquiry* 9 (3): 395–415. https://doi.org/10.1177/1077800403009003005

Klein, Julie Thompson (2010). "A Taxonomy of Interdisciplinarity". In R. Frodeman, J. T. Klein, & C. Mitcham (Eds.), The Oxford Handbook of Interdisciplinarity (pp. 15-30).

References

Oxford: Oxford University Press.

Lumsden, Karen, and Emily Harmer, eds. 2019. *Online Othering Book: Exploring Digital Violence and Discrimination on the Web*. 1st ed. Vol. Cham: Palgrave Macmillan. XIX– 407. https://doi.org/10.1007/978-3-030-12633-9

Lynggaard, Kennet. 2015. "Dokumentanalyse". In *Kvalitative Metoder: En Grundbog*, edited by Svend Brinkmann and Lene Tanggaard, 153–67. Hans Reitzels Forlag.

Marwick, Alice, and Rebecca Lewis. 2017. "Media Manipulation and Disinformation Online." Data & Society. https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf

Mogensen, Christian, and Stine Helding Rand. 2020. "The Angry Internet: A Threat to Gender Equality, Democracy & Well-Being." Centre for Digital Youth Care. https://cfdp.dk/wp-content/uploads/2020/11/CFDP_the_angry_internet_ISSUE.pdf

Mooghen, Justice Monir and Alberte Bau Larsen. 2019. *Edge of Margins: When Social Media Oppress*. Aalborg University, Copenhagen: Student project.

Musto, Jennifer, Anne E. Fehrenbacher, Heidi Hoefinger, Nicola Mai, P. G. Macioti, Calum Bennachie, Calogero Giametta, and Kate D'Adamo. 2021. "Anti-Trafficking in the Time of FOSTA/SESTA: Networked Moral Gentrification and Sexual Humanitarian Creep." *Social Sciences* 10 (2): 58. https://doi.org/10.3390/socsci10020058

Nagle, Angela. 2015. "An Investigation into Contemporary Online Anti-Feminist Movements.". http://doras.dcu.ie/22385/3/Angela_Nagle_PhD_thesis.pdf

Nagle, Angela. 2017. *Kill All Normies: The Online Culture Wars from Tumblr and 4chan to the Alt-Right and Trump*. Winchester, UK; Washington, Usa: Zero Books.

# References

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York: New York University Press, Cop.

Padilla, Amado M. 1994. "Research News and Comment: Ethnic Minority Scholars; Research, and Mentoring: Current and Future Issues." *Educational Researcher* 23 (4): 24–27. https://doi.org/10.3102/0013189x023004024

Pariser, Eli. 2012. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think.* New York, N.Y.: Penguin Books.

Parker, David, Julia M. Pearce, Lasse Lindekilde, and M. Brooke Rogers. 2018. "Press Coverage of Lone-Actor Terrorism in the UK and Denmark: Shaping the Reactions of the Public, Affected Communities and Copycat Attackers." *Critical Studies on Terrorism* 12 (1): 110–31. https://doi.org/10.1080/17539153.2018.1494792

Pohjonen, Matti, and Sahana Udupa. 2017. "Extreme Speech Online: An Anthropological Critique of Hate Speech Debates." International Journal of Communication. https://www.researchgate.net/publication/321097183_Extreme_Speech_Online_An_Anthropological_Critique_of_Hate_Speech_Debates

Richards, Julian. 2019. "Propaganda I En Ekstremistisk Sammenhæng". Nationalt Center for Forebyggelse af Ekstremisme. https://stopekstremisme.dk/ekstremisme/videnspublikationer/propaganda-i-en-ekstremistisk-sammenhaeng

Rogers, Richard. 2020. "Deplatforming: Following Extreme Internet Celebrities to Telegram and Alternative Social Media." *European Journal of Communication* 35 (3): 213–29. https://doi.org/10.1177/0267323120922066

Sager, Maja, and Diana Mulinari. 2018. "Safety for Whom? Exploring Femonationalism and Care-Racism in Sweden." *Women's Studies International Forum* 68 (): 149–56.

# References

https://doi.org/10.1016/j.wsif.2017.12.002

Sap, Maarten, Dallas Card, Saadia Gabriel, Yejin Choi, Noah Smith, and Paul Allen. 2019. "The Risk of Racial Bias in Hate Speech Detection." Association for Computational Linguistics. https://homes.cs.washington.edu/~msap/pdfs/sap2019risk.pdf

Skadegård Thorsen, Tess S. 2019. "Minoritetsbeskatning - et Værktøj Til at Forstå Opretholdelse Af Strukturelle Uligheder I Dansk Akademia." *Kvinder, Køn & Forskning*, no. 1-2 (July): 31–43. https://doi.org/10.7146/kkf.v28i1-2.116115

Smith, Jo. 2019. "'When I Saw Women Being Attacked … It Made Me Want to Stand Up and Fight': Reporting, Responding to, and Resisting Online Misogyny" in *Online Othering*, eds. Harmer and Lumsden, 287-308. New York: Palgrave Macmillan.

Social- og Integrationsministeriet. 2012. "Metoder i Arbejdet med Radikalisering.". https://uim.dk/publikationer/metoder-i-arbejdet-med-radikalisering

Staniforth, Andrew, and Holger Nitsch. 2014. "Preventing Terrorism Together: A Framework to Provide Social Media Anti-Radicalization Training for Credible Community Voices." In *Emerging Trends in ICT Security*, edited by Babak Akhgar and Hamid R. Arabnia, 549–56. Amsterdam ; Boston: Morgan Kaufmann/Elsevier. https://ebookcentral.proquest.com/lib/aalborguniv-ebooks/detail.action?pq-origsite=primo&docID=1562332

Stender Petersen, Kit, and Rikke Louise Alberg Peters. 2020. "Desk Research: Kortlægning Af Viden Om Forebyggelse Af Ekstremisme Online Blandt Børn Og Unge." Center for Forebyggelse af Ekstremisme. https://stopekstremisme.dk/ekstremisme/videnspublikationer/desk-research-kortlaegning-af-viden-om-forebyggelse-af-ekstremisme-online-blandt-born-og-unge

TechDK Kommissionen. 2020. "TechDK Rapport 2 - Demokrati." . Djøf.

https://www.djoef.dk/-/media/documents/djoef/t/techdk/techdk_rapport_2_demokrati_a4_1119.ashx

Tutt, Andrew. 2016. "An FDA for Algorithms." *SSRN Electronic Journal*, 83–123. https://doi.org/10.2139/ssrn.2747994

Udenrigsministeriet. 2021. Review of *Strategi for Danmarks Teknologiske Diplomati 2021-2023*. https://techstrategi.um.dk/

Zannettou, Savvas, Barry Bradlyn, Emiliano De Cristofaro, Haewoon Kwak, Michael Sirivianos, Gianluca Stringini, and Jeremy Blackburn. 2018. "What Is Gab? A Bastion of Free Speech or an Alt-Right Echo Chamber?" *International World Wide Web Conferences Steering Committee*, March, 1–8. https://doi.org/10.1145/3184558.3191531

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. London Profile Books. 1-691.

## Links

Beckett, Lois. 2021. Review of *The Misogynist Incel Movement Is Spreading. Should It Be Classified as a Terror Threat?*. The Guardian. March 3, 2021. Accessed May 2, 2021. https://www.theguardian.com/lifeandstyle/2021/mar/03/incel-movement-terror-threat-canada

Bindel, Julie. 2012. "The Montreal Massacre: Canada's Feminists Remember." *The Guardian*, December 3, 2012, sec. World news. Accessed May 2, 2021. https://www.theguardian.com/world/2012/dec/03/montreal-massacre-canadas-feminists-remember

CNN. 2014."Transcript of Video Linked to Santa Barbara Mass Shooting". *CNN*, May 28, 2014. Accessed April 6, 2021. https://edition.cnn.com/2014/05/24/us/elliot-rodger-video-transcript/index.html

## References

Crenshaw, Kimberlé. 2016. "The Urgency of Intersectionality." TED Talk. Accessed

      May 12, 2021.

      https://www.ted.com/talks/kimberle_crenshaw_the_urgency_of_intersectionality#t-

      1116240

Cuthbertson, Anthony. 2019. *"Subscribe to PewDiePie": What Did Christchurch Mosque*

      *Gunman Mean in Final Words before Shooting?* Independent. March 15, 2019.

      Accessed May 2, 2021.

      https://www.independent.co.uk/life-style/gadgets-and-tech/news/new-zealand-

      shooting-attack-pewdiepie-subscribe-youtube-mosque-a8825326.html

Cyboid, Cookie. 2018. "Want to Know Why Tumblr Is Cracking down on Sex? Look to

      FOSTA/SESTA". Medium. The Establishment. December 25, 2018. Accessed May 1,

      2021. https://medium.com/the-establishment/want-to-know-why-tumblr-is-cracking-

      down-on-sex-look-to-fosta-sesta-15c4174944a6

"Ending the Subscribe to Pewdiepie Meme." 2019. YouTube Video. *YouTube*. Accessed

      May 2, 2021. https://www.youtube.com/watch?v=Ah5MYGQBYRo

Farokhmanesh, Megan. 2018. "YouTube Is Still Restricting and Demonetizing LGBT Videos

      — and Adding Anti-LGBT Ads to Some." The Verge. June 4, 2018. Accessed May 1,

      2021. https://www.theverge.com/2018/6/4/17424472/youtube-lgbt-

      demonetizationads-algorithm

Ford, Clementine. 2021. "Big Sister Hotline: S2E1 Feat. Laura Bates" *Clementine Ford's Big*

      *Sister Hotline*. Podcast audio, January 30, 2021.

      https://podcasts.apple.com/au/podcast/big-sister-hotline-s2e1-feat-laura-

      bates/id1499175131?i=1000507081882

Gibson, Caitlin. 2019. "'Do You Have White Teenage Sons? Listen Up.' How White

      Supremacists Are Recruiting Boys Online." *The Washington Post*, September 17, 2019.

      Accessed May 21, 2021.  https://www.washingtonpost.com/lifestyle/on-parenting/do-

# References

you-have-white-teenage-sons-listen-up-how-white-supremacists-are-recruiting-boys-online/2019/09/17/f081e806-d3d5-11e9-9343-40db57cf6abd_story.html

Gobry, Pascal-Emmanuel. 2011. "Eric Schmidt to World Leaders at EG8: Don't Regulate Us, or Else." Business Insider. May 24, 2011. Accessed May 19, 2021. https://www.businessinsider.com/eric-schmidt-google-eg8-2011-5?op=1&r=US&IR=T

Hjortkjær Henningsen, Laurits. 2021. "Voldtægtskultur På Facebook: Danske Mænd Dyrker Kvindehad I Private Grupper." Politiken. March 27, 2021. Accessed May 23, 2021 https://politiken.dk/kultur/medier/art8139946/Voldt%C3%A6gtskultur-p%C3%A5-Facebook-Danske-m%C3%A6nd-dyrker-kvindehad-i-private-grupper

Holland, Kate. 2019. "New Bill Calls for Examination of Anti-Trafficking FOSTA-SESTA Law". ABC News. December 20, 2019. Accessed May 3, 2021. https://abcnews.go.com/Politics/bill-calls-examination-anti-trafficking-fosta-sesta-law/story?id=67831743

Høst, Marie. 2019. "Episode 4: Fra 'Gurli Gris' til 'How To Buy A Bazooka'". *Algoritmen*. Podcast audio, October 13, 2019. https://podimo.com/dk/shows/3445619b-3c4d-4121-a328-eddfe89ed311

"I'M RACIST?" 2016. YouTube Video. *YouTube*. Accessed May 2, 2021. https://www.youtube.com/watch?v=QGX2mJ6IbS0

Jørgensen, Anna Sol. 2021. "Torben Sangild Er Blevet Smidt Af Facebook Uden Nogen Forklaring." Journalisten. January 25, 2021. Accessed May 13, 2021. https://journalisten.dk/torben-sangild-er-blevet-smidt-af-facebook-uden-nogen-forklaring/

Kaiser, Jonas, and Adrian Rauchfleisch. 2018. "Unite the Right? How YouTube's Recommendation Algorithm Connects the U.S. Far-Right. Medium. D&S Media Manipulation: Dispatches from the Field. Accessed April 11, 2018.

# References

https://medium.com/@MediaManipulation/unite-the-right-how-youtubes-recommendation-algorithm-connects-the-u-s-far-right-9f1387ccfabd

Karacan, Telli Betül. 2021. "Deplatforming Virker, Men Udfordrer Demokratiet Og Ytringsfriheden". Jyllands Posten. April 18, 2021. Accessed May 5, 2021. https://jyllands-posten.dk/debat/breve/ECE12903114/deplatforming-virker-men-udfordrer-demokratiet-og-ytringsfriheden/

Kraus, Rachel. 2018. "2018 Was the Year We (Sort Of) Cleaned up the Internet." Mashable. December 26, 2018. Accessed May 12, 2021. https://mashable.com/article/deplatforming-alex-jones-2018/?europe=true

Lindberg, Kristian. 2019. "Vi Krænkes Som Aldrig Før, Men Hvorfor Blev vi Så Sarte? Her Er et Bud På En Forklaring." Berlingske.dk. February 26, 2019. Accessed May 12, 2021. https://www.berlingske.dk/aok/vi-kraenkes-som-aldrig-foer-men-hvorfor-blev-vi-saa-sarte-her-er-et-bud-paa-en

Lewis, Paul, and Erin McCormick. 2018. "How an Ex-YouTube Insider Investigated Its Secret Algorithm." *The Guardian*, February 2, 2018, sec. Technology. Accessed April 11, 2021. https://www.theguardian.com/technology/2018/feb/02/youtube-algorithm-election-clinton-trump-guillaume-chaslot

Lund, Rikke. 2019. "Gymnasielever Trodser Krænkelseskultur: - Er Du Krænkelsesparat? Så Skal Du Nok Ikke Møde Op." TV2 Fyn. January 25, 2019. Accessed May 12, 2021. https://www.tv2fyn.dk/odense/gymnasielever-trodser-kraenkelseskultur-er-du-kraenkelsesparat-sa-skal-du-nok-ikke-mode-op

Mahdawi, Arwa. 2017. "PewDiePie Thinks "Death to All Jews" Is a Joke. Are You Laughing Yet?" *The Guardian*. February 15, 2017. Accessed May 2, 2021. https://www.theguardian.com/commentisfree/2017/feb/15/youtube-pewdiepie-thinks-death-to-all-jews-joke-laughing-yet

## References

Maislis, Vi. 2020. "Why Is Everybody Suddenly Sharing These Alt-Right Memes?" Alma.
December 16, 2020. https://www.heyalma.com/why-is-everybody-suddenly-sharing-
these-alt-right-memes/

Martineau, Paris. 2019. "Facebook Bans Alex Jones, Other Extremists—but Not as
Planned." Wired. May 2, 2019. Accessed May 14, 2021.
https://www.wired.com/story/facebook-bans-alex-jones-extremists/.

Mullin, Joe. 2018. "House Vote on FOSTA Is a Win for Censorship." Electronic Frontier
Foundation. February 27, 2018. Accessed April 30, 2021.
https://www.eff.org/deeplinks/2018/02/house-vote-fosta-win-censorship

Mustafa, Bave. 2019. "Uønsket I Norge: Ingen Adgang Til Paludan." Ekstrabladet.dk.
Ekstrabladet. August 13, 2019. Accessed June 9, 2021.
https://ekstrabladet.dk/nyheder/politik/danskpolitik/uoensket-i-norge-ingen-
adgang-til-paludan/7748100

Nationalt Center for Forebyggelse af Ekstremisme. 2020. "Ekstremisme Og Radikalisering".
May 28, 2020. Accessed May 6, 2021.
https://stopekstremisme.dk/ekstremisme/ekstremisme

Newton, Casey. 2017. "How YouTube Perfected the Feed." The Verge. The Verge. August
30, 2017. Accessed April 11, 2021.
https://www.theverge.com/2017/8/30/16222850/youtube-google-brain-algorithm-
video-recommendation-personalized-feed

Newton, Casey. 2020. "YouTube Moderators Are Being Forced to Sign a Statement
Acknowledging the Job Can Give Them PTSD." The Verge. January 24, 2020.
Accessed May 9, 2021.

# References

https://www.theverge.com/2020/1/24/21075830/youtube-moderators-ptsd-accenture-statement-lawsuits-mental-health

O'Donovan, Caroline, Charlie Warzel, Logan McDonald, Brian Clifton, and Max Woolf. 2019. "We Followed YouTube's Recommendation Algorithm down the Rabbit Hole." BuzzFeed News. January 24, 2019. Accessed June 8, 2021. https://www.buzzfeednews.com/article/carolineodonovan/down-youtubes-recommendation-rabbithole

Ohlheiser, Abby, and Ian Shapira. 2018. "Gab, the White Supremacist Sanctuary Linked to the Pittsburgh Suspect, Goes Offline (for Now)." *Washington Post*, October 29, 2018. Accessed May 14, 2021. https://www.washingtonpost.com/technology/2018/10/28/how-gab-became-white-supremacist-sanctuary-before-it-was-linked-pittsburgh-suspect/

Overgaard, Linda. 2020. "Lovgivning Skal Bremse Algoritmernes Magt.". Djøfbladet. January 17, 2020. Accessed May 17, 2021. https://www.djoefbladet.dk/artikler/2020/1/lovgivning-skal-bremse-algoritmernes-magt.aspx

Skovm, Steffen. 2019. "Klumme: Rasmus Paludan - Vor Tids Forsvarer Af Ytringsfriheden." 24NYT. May 1, 2019. Accessed June 9, 2021. https://24nyt.dk/klumme-rasmus-paludan-vor-tids-forsvarer-af-ytringsfriheden/

The New York Times. 2020. "Three: Mirror Image,", sec. Rabbit Hole Podcast. Accessed May 1, 2021 https://www.nytimes.com/2020/04/30/podcasts/rabbit-hole-internet-youtube-virus.html?action=click&module=audio-series-bar&pgtype=Article%C2%AEion=header&showTranscript=1

"PewDiePie's YouTube Stats (Summary Profile) - Social Blade Stats." n.d. Socialblade.com.

# References

https://socialblade.com/youtube/user/pewdiepie

Richman, Jessica. 2019. "This Is the Impact of Instagram's Accidental Fat-Phobic
    Algorithm." Fast Company. October 15, 2019. Accessed May 3, 2021.
    https://www.fastcompany.com/90415917/this-is-the-impact-of-instagrams-accidental-
    fat-phobic-algorithm

Rodger, Elliot. 2014. "My Twisted World: The Story of Elliot Rodger". Accessed April 6,
    2021. http://files.crimescenedb.com/pdf/Elliot-Rodger-Manifesto-My-Twisted-
    World.pdf

Roose, Kevin. 2020. "Welcome to the 'Rabbit Hole.'" *The New York Times*, April 11, 2020,
    sec. Technology. Accessed April 11, 2021.
    https://www.nytimes.com/2020/04/16/technology/rabbit-hole-podcast-kevin-
    roose.html

Roose, Kevin, Andy Mills, Julia Longoria, Sindhu Gnanasambandan, Larissa Anderson, and
Wendy Dorr. 2020. *Rabbit Hole*. Episode 1-8. New York Times. April 11, 2020. Accessed
    April 11, 2021
    https://www.nytimes.com/2020/04/16/podcasts/rabbit-hole-internet-youtube-
    virus.html?action=click&module=audio-series-
    bar&pgtype=Article%C2%AEion=header&showTranscript=1

Roose, Kevin, and Kate Conger. 2019. "YouTube to Remove Thousands of Videos Pushing
    Extreme Views". The New York Times. June 5, 2019. Accessed May 9, 2021.
    https://www.nytimes.com/2019/06/05/business/youtube-remove-extremist-
    videos.html

Tierney, Allison, and Mack Lamoureux. 2018. "Everything We Know about Alek
    Minassian, the Alleged Toronto van Attacker". *Vice*, April 24, 2018. Accessed April 6,
    2021. https://www.vice.com/en/article/59j83n/everything-we-know-about-alek-
    minassian-the-alleged-toronto-van-attacker

References

"YouTube: Candidates Favored by the Algorithm." n.d. Algotransparency.org. Accessed
April 14, 2021.
https://data.algotransparency.org/methodology.html?candidat=Marine%20Le%20Pen
&file=ytrecos-presidentielle-2017-06-10