

Ontwikkelingspaden en handelingsperspectieven

# Rechtsextremisme op sociale mediaplatforms?



# Rechtsextremisme op sociale mediaplatforms?

Ontwikkelingspaden en handelingsperspectieven

Utrecht, augustus 2023



# Inhoud

Samenvatting	4	4	Aanbevelingsalgoritmen als sociale mediafuik	67
Summary	12		4.1. Introductie over algoritmen	67
1	19		4.2. Aanbevelingsalgoritmen op YouTube en TikTok	69
1.1. Probleemschets van rechtsextremisme in het online domein	19		4.3. Sociale mediafuik en (rechts)extremisme?	74
1.2. Doelstelling en onderzoeksvragen	22		5	79
1.3. Opzet en uitgangspunten voor het onderzoek	23		5.1. Introductie	79
1.4. Leeswijzer	25		5.2. Opzet en uitvoering van de trial op YouTube	83
2	27		5.3. Opzet en uitvoering van de trial op TikTok	93
2.1. Introductie	27		5.4. Handelingsperspectieven platforms	100
2.2. Risicofactoren en profielkenmerken	29		6	103
2.3. Beschermende factoren en weerbaarheid	33		6.1. Introductie	103
2.4. Radicaliseringsmechanismen	35		6.2. Radicaliseringscasussen: online invloeden	104
2.5. Verschillen radicaliseringsmechanismen per ideologie?	38		6.3. Online en offline rekrutering	106
2.6. Voorspelt online radicalisering gewelddadig rechtsextremisme?	41		7	109
3	44		7.1. Nuances omtrent de rol van aanbevelingsalgoritmen bij rechtsextremistische radicalisering	109
3.1. Introductie	44		7.2. Moderatie van algoritmen en content door sociale mediaplatforms	111
3.2. Definieren en identificeren van (rechts)extremistische content	45		7.3. Sociale mediabeleid in Nederland en de Europese Unie	116
3.3. Rechtsextremistische content: online strategieën	49		7.4. Handelingsperspectieven online	121
3.4. Normalisering, acclimatisering en dehumanisering	52		7.5. Handelingsperspectieven offline	127
3.5. Rechtsextremistische content: mainstream en alternatieve platforms	54		8	131
3.6. Rechtsextremistische content: gameplatforms	58		Bijlage 1 Glossary	137
3.7. Subculturen als online 'gateways'	59		Bijlage 2 Referenties	141

# Samenvatting

## Bevindingen op hoofdlijnen

Dit onderzoek heeft in kaart gebracht wat verschillende ontwikkelingspaden van rechtsextremistische radicalisering zijn en wat hierbij de rol is van sociale mediaplatforms, met daarbij specifieke aandacht voor de werking van aanbevelingsalgoritmen. Ook is gekeken naar de rol van de offline context in ontwikkelingspaden van rechtsextremistische radicalisering. Op basis van de bevindingen zijn mogelijke haalbare handelingsperspectieven beschreven voor de overheid om rechtsextremistische radicalisering via online platforms te voorkomen.

De belangrijkste bevindingen van het onderzoek zijn de volgende:

- Een groot aantal risicofactoren op individueel-, groeps- en samenlevingsniveau dragen bij aan het ontstaan van radicaliseringsprocessen. Ontwikkelingspaden van radicalisering zijn uiterst complex, waarbij zowel online als offline factoren een rol spelen en op elkaar inwerken. Om de invloed van de online context op extremistische overtuigingen te kunnen begrijpen, dient te worden gekeken naar deze wisselwerking.
- De rol van aanbevelingsalgoritmen bij rechtsextremistische radicalisering binnen de online context is volgens recent wetenschappelijk onderzoek beperkt. Persoonlijke keuzen van gebruikers zijn minstens zo belangrijk, zoals het delen van links naar extremistische content; het abonneren / volgen van bepaalde kanalen op sociale mediaplatforms en op die manier selectief geconfronteerd worden met meer extreme content; het ontmoeten van gelijkgestemden in echokamers; en persoonlijk surf/zoekgedrag. Mensen gaan uit zichzelf op zoek naar extremere content en voeren hiertoe doelbewust bepaalde zoektermen in, of klikken door op de meer extreme aanbevelingen.
- Mainstream sociale media platforms lijken de moderatie van extremistische (illegale) content beter op orde te krijgen. Gebruikers krijgen via aanbevelingsalgoritmen minder extremistische content gesuggereerd. Echter, er kunnen een aantal disclaimers geplaatst worden bij deze conclusies:
  - Extremistische content is verplaatst naar alternatieve 'alt-tech' platforms.
  - Extremistische content lijkt op mainstream sociale media platform grotendeels vervangen te zijn door polariserende borderline content (ook wel aangeduid als 'legal yet harmful' content). Een groot probleem met de aanpak van borderline content is de beoordeling wanneer deze content illegaal is, vanwege het gebruik van gecodeerde taal, symbolen, ambigue taalgebruik, beeldspraak en ironie waarin extremistische content wordt verpakt. Het vereist steeds meer expertise om deze te identificeren. Het gebrek aan deze expertise is een zwakke schakel in de strijd tegen borderline content op de platforms.
  - Platformgebruikers kunnen nog steeds in aanraking komen met extremistische content door het eigen online gedrag, zoals het delen van weblinks, het abonneren of volgen van bepaalde kanalen en het zoeken/vinden van gelijkgestemden in bepaalde online (gesloten) chatgroepen of op fora.
- Oplossingsrichtingen om (online) rechtsextremistische radicalisering te voorkomen zijn dan ook het centraal stellen van het probleem van niet-illegale extreme (borderline) content en handvatten om deze content beter te begrijpen en modereren; meer aandacht te besteden aan illegale content op alt-tech mediaplatforms; én de verantwoordelijkheid al dan niet gedeeltelijk terug te leggen bij gebruikers zelf, aangezien zij door het eigen gedrag eerder in aanraking kunnen komen met borderline- of illegale content.
- Het handelingsperspectief van de Nederlandse overheid ten aanzien van de zeer grote online platforms is beperkt, aangezien de Europese Commissie respectievelijk het land van vestiging hier de rol van toezichthouder heeft. Meer handelingsperspectieven lijken mogelijk ten aanzien van de minder grote platforms.

- De handelingsperspectieven zijn te categoriseren langs vier lijnen:
  - Inzet op dialoog met kleinere platforms om extreme, illegale content sneller te verwijderen.
  - Investeren in samenwerking tussen overheden, online platforms en onderzoekers op het gebied van meer transparantie over aanbevelingsalgoritmen; toewerken naar meer co-regulering tussen platforms en overheid; investeren in het beter begrijpen van de problematiek rondom borderline content en faciliteren dat meer duidelijkheid komt over mogelijke schadelijke effecten en illegaliteit van borderline content.
  - Investeren in ruimte voor experimenteren en mogelijkheden te onderzoeken of steun te geven aan initiatieven voor lokaal werkende online platforms of non-profit sociale media.
  - Inzet op offline aanpakken door meer te investeren in digitale weerbaarheid en mediawijsheid. Hiertoe zijn inmiddels meerdere geëvalueerde lesprogramma's beschikbaar welke breder ingezet kunnen worden.

## Samenvatting en conclusies

### Doelstellingen van het onderzoek

Het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) heeft aan het Verwey-Jonker Instituut en TextGain gevraagd een onderzoek uit te voeren om meer inzicht te krijgen in de ontwikkelingspaden voor het optreden van rechtsextremistische radicalisering, met daarin centraal de rol van online sociale media platforms, aanbevelingsalgoritmen en de wisselwerking tussen de online- en offline contexten. De doelstellingen van het onderzoek waren de volgende:

- Wat zijn de ontwikkelingspaden voor het optreden van rechtsextremistische radicalisering en wat is daarbij de rol van online sociale mediaplatforms?
- Welke invloed hebben de aanbevelingsalgoritmen die op sociale mediaplatforms worden toegepast (nog) op rechtsextremistische radicalisering van gebruikers van deze platforms?

- Vindt er naast online radicalisering ook offline rekrutering plaats bij rechtsextremistische radicalisering (wisselwerking online en offline contexten)?
- Welke haalbare handelingsperspectieven zijn er om rechtsextremistische radicalisering via online sociale mediaplatforms tegen te gaan?

### Welke informatie staat in het rapport?

Het onderzoek heeft geresulteerd in een rapportage waarin op toegankelijke wijze een overzicht wordt gegeven van:

- Theorieën en actuele inzichten omtrent ontwikkelingspaden van radicalisering, in het bijzonder rechtsextremistische radicalisering in de online context, waaronder:
  - Risico- en beschermende factoren die een rol spelen bij de ontvanbaarheid voor radicalisering.
  - Radicaliseringsmechanismen in het online domein.
  - Relatie tussen online rechtsextremistische radicalisering en het gebruik van geweld.
- Definities, afbakening, inhoudelijke kenmerken en narratieven van (rechts)extremistische content.
- Online strategieën om rechtsextremistische content binnen de leefwereld van personen te brengen, zoals:
  - Gebruik van humor, ironie en dog whistles.
  - Mainstreaming van (rechts)extremistische content en red-pilling.
  - Normalisering, acclimatisering en dehumanisering.
  - Rechtsextremistische content op mainstream, alternatieve en game platforms.
- Online subculturen en influencers als 'gateways' naar rechtsextremistische content, waaronder: Conspiritualiteit en Qanon; Manosfeer en masculine subculturen; Incels subcultuur; Vikingcultuur; Preppers community.

- Aanbevelingsalgoritmen als sociale mediafuik?
  - Wat zijn aanbevelingsalgoritmen en hoe werken deze?
  - Aanbevelingsalgoritmen op YouTube en TikTok.
  - Is er bewijs voor aanbevelingsalgoritmen die leiden tot sociale mediafuiken en radicalisering?
- Trials aanbevelingsalgoritmen op YouTube en TikTok op basis van gepersonaliseerd zoekgedrag binnen online subculturen (voorbeelden: Manosfeer/masculiene subcultuur en Vikingcultuur).
- Beschrijving van enkele casuïstieken van rechtsextremistische radicalisering waarbij online sociale mediafuiken mogelijk een rol hebben gespeeld.
- Nieuwe regelgeving en handelingsperspectieven
  - Nuances omtrent de rol van aanbevelingsalgoritmen bij rechtsextremistische radicalisering.
  - Wat doen sociale mediaplatforms al aan moderatie van algoritmen en content?
  - Sociale mediabeleid in Nederland en de Europese Unie, nieuwe mogelijkheden door de Digital Services Act.
  - Handelingsperspectieven online en handelingsperspectieven offline.

### **Opzet van het onderzoek**

De methodische uitvoering van dit onderzoek bestond uit verschillende onderdelen. Het eerste onderdeel was een uitgebreide state-of-the-art literatuurstudie naar inzichten in de wetenschappelijke literatuur over ontwikkelingspaden van online (rechtsextremistische) radicalisering en de rol hierin van verschillende 'typen' sociale media. Het tweede onderdeel betrof interviews met experts over de werking en effecten van aanbevelingsalgoritmen, de invloed van hiervan ten opzichte van andere mechanismen in de online – en offline context, en haalbare mogelijkheden voor de overheid om online rechtsextremistische radicalisering tegen te gaan. De conceptbevindingen van het rapport zijn voorgelegd in een gesprek met twee juristen met kennis over de nieuwe Europese wetgeving omtrent sociale media platforms (Digital Service Act) en in twee expertsessies aan deskundigen op het gebied van de werking en effecten van aanbevelingsalgoritmen; media en digitale samenleving; en sociale media en de maatschappelijke impact hiervan. In deze expertsessies is gereflecteerd op de resultaten van de literatuurstudie en – vooral – het gezamenlijk met de experts doordenken van haalbare en (daarmee) praktisch inzetbare handelingsperspectieven voor de overheid. Deze onderzoeksactiviteiten zijn uitgevoerd door het Verwey-Jonker Instituut. Het derde onderdeel van de studie is uitgevoerd door TextGain en betrof de uitvoering van online trials op YouTube en TikTok waarin is onderzocht in hoeverre algoritmen op basis van gepersonaliseerd zoekgedrag door gebruikers leiden tot suggesties voor extreme(re) content. Wij voerden deze trials uit als een aanvulling op de literatuurstudie. Het onderzoek is begeleid door een begeleidingscommissie met vertegenwoordigers van Ministerie J&V en het WODC en wetenschappers vanuit Universiteit van Amsterdam, Universiteit van Maastricht/Vrije Universiteit Amsterdam en Universiteit Leiden.

## Resultaten

Dit onderzoek heeft in kaart gebracht wat verschillende ontwikkelingspaden van rechtsextremistische radicalisering zijn en wat hierbij de rol is van sociale mediaplatforms. Daarbij was specifiek aandacht voor de werking van aanbevelingsalgoritmen. Ook is gekeken naar de rol van de offline context in ontwikkelingspaden van rechtsextremistische radicalisering. Op basis van de bevindingen zijn handelingsperspectieven beschreven die haalbaar geacht worden voor de overheid om bij te dragen aan het voorkomen van rechtsextremistische radicalisering via online platforms.

### Grote variëteit aan risicofactoren en ontwikkelingspaden

Uit de wetenschappelijke literatuur blijkt dat een groot aantal risicofactoren op individueel-, groeps- en samenlevingsniveau kunnen bijdragen aan het ontstaan van radicaliseringsprocessen. Daarbij zijn er veel meer overeenkomsten dan verschillen in de kenmerken, risicofactoren en mechanismen van radicaliseringsprocessen met uiteenlopende ideologische achtergronden (links- en rechtsextremisme, religieus gemotiveerd en single-issue extremisme). Ontwikkelingspaden van radicalisering zijn uiterst complex, waarbij zowel online als offline factoren een rol spelen en op elkaar inwerken. Om de invloed van de online context op extremistische overtuigingen en geweld te kunnen begrijpen dient te worden gekeken naar deze wisselwerking. Op basis van empirische studies blijkt dat de meeste personen die online radicaliseren, niet overgaan tot gewelddadig gedrag. Er zijn zeer uiteenlopende manieren waarop rechtsextremistisch gedachtengoed de online leefwereld van personen kan binnendringen. Diverse online strategieën van rechtsextremistische individuen en netwerken kunnen bijdragen aan de blootstelling en acceptatie van rechtsextremistische content. Een belangrijke strategie daarbij is het 'vermommen' van rechtsextremistische content door het gebruik van humor, ironie en dog whistles. Een langdurige blootstelling aan rechtsextremistische content – en strategieën kan personen steeds ontvankelijker maken hiervoor. In de literatuur worden hiertoe meerdere zogeheten 'cognitieve fasen' geïdentificeerd. Het gaat om de volgende fasen: normalisering, acclimatisering en dehumanisering. Deze fasen overlappen elkaar, zijn niet exclusief of strikt lineair, maar veeleer cyclisch van aard.

### Gateways naar rechtsextremistische radicalisering

Radicaal- en rechtsextremistische groepen zijn actief op verschillende sociale mediaplatforms, maar hebben nog altijd het grootste bereik via mainstream-platforms zoals Facebook, Twitter, YouTube en in iets mindere mate Instagram. Deze grote platforms hebben in de afgelopen jaren streng opgetreden tegen kanalen en pagina's die gelinkt zijn aan rechtsextremistische groepen of alternatieve media. Alternatieve ofwel 'fringe' platforms zoals Gab, Telegram en 4chan zijn mede daardoor populairder geworden onder rechtsextremistische groepen, omdat hier meer extreme content kan worden gepost. Aangezien het bereik op alternatieve platforms kleiner is, wordt op de grotere platforms meer 'gematigd' materiaal gepost dat aan de voorwaarden van het platform voldoet. Wel blijkt dat gebruikers via deze 'gematigde' content soms via links doorgeleid worden, of zelf op zoek gaan, naar meer extreme content elders. Ook online subculturen die niet noodzakelijk rechtsextremistisch zijn (zoals conspiritualiteit, manosphere en viking subcultuur) kunnen fungeren als gateways naar rechtsextremistisch gedachtengoed. Zoals beschreven in paragraaf 3.7 kunnen personen binnen deze subculturen eerder met rechtsextremistisch gedachtengoed geconfronteerd worden. Sommige narratieven die binnen deze subculturen worden verspreid hebben overeenkomsten met rechtsextremistische narratieven. Een andere mogelijke ingang naar extremistische content wordt gevormd door sommige populaire online influencers. Online is sprake van talrijke invloedrijke figuren die via grote sociale mediaplatforms of gameplatforms een ingang kunnen zijn voor (jonge) personen om zich online verder te verdiepen in bepaald gedachtengoed of bepaalde gemeenschappen.

### **Online en offline rekrutering**

Vanuit de wetenschappelijke literatuur is betrekkelijk weinig bekend over casuïstieken waarbij is vastgesteld dat een sociale mediafuik heeft bijgedragen aan rechtsextremistische radicalisering of rekrutering. Het vaststellen hiervan is ook niet eenvoudig, omdat de werking van aanbevelingsalgoritmen en filterbubbels zich niet een geïsoleerd universum voltrekt. Er zijn immers ook andere online- en offline mechanismen die bij een persoon kunnen bijdragen aan zijn of haar radicalisering. Daarbij is het moeilijk te meten of iemand specifiek door de sociale mediafuik is geradicaliseerd of dat andere factoren een grotere bijdrage hebben geleverd. Het is kortom nauwelijks mogelijk om te concluderen dat een individu uitsluitend is geradicaliseerd door consumptie van eenzijdige online content op sociale mediaplatforms en dat diegene zonder deze media niet geradicaliseerd zou zijn. Desalniettemin zijn er anekdotische aanwijzingen (casussen) waarin aannemelijk is dat de sociale mediafuik een rol heeft gespeeld in het radicaliseringsproces. Vraag is dan echter wel in hoeverre deze fuik veroorzaakt wordt door bijvoorbeeld aanbevelingsalgoritmen op sociale mediaplatforms, of dat er sprake is van een ogenschijnlijke fuik die veeleer te wijten is aan het eigen online gedrag (keuzes) van gebruikers.

### **Rol aanbevelingsalgoritmen bij rechts-extreme radicalisering is beperkt**

Zoals hierboven opgemerkt, voltrekt het mechanisme van een "algoritme gestuurd aanbod van rechtsextremistische content op sociale mediaplatforms dat kan leiden tot een rechtsextremistische sociale media fuik" zich allesbehalve in een vacuüm. De rol van aanbevelingsalgoritmes en sociale mediafuiken bij rechtsextremistische radicalisering lijkt kleiner dan in het publieke discours soms wordt voorondersteld. Personen kunnen zowel online als offline op allerlei manieren in aanraking komen met polariserende of extremistische content. De verscheidene persoonlijke keuzes die gebruikers maken zijn daarbij relevant:

- Mensen delen binnen hun sociale netwerk links naar extremistische content met elkaar;
- Mensen abonneren / volgen bepaalde kanalen op sociale mediaplatforms en worden op die manier selectief geconfronteerd met meer extreme content;
- Gelijkgestemden ontmoeten elkaar in echokamers zoals chatgroepen op Telegram;
- Zoekgedrag: mensen gaan uit zichzelf op zoek naar extremere content en voeren hiertoe doelbewust bepaalde zoektermen in, of klikken door op de meer extreme aanbevelingen

Ook zijn vraagtekens te plaatsen bij de rol van internet en sociale mediaplatforms bij gewelddadige radicalisering. Zoals eerder gesteld is er weinig empirisch bewijs dat radicalisering in de online context leidt tot gewelddadig extremisme noch dat gewelddadig extremisme rechtstreeks voortvloeit uit online radicalisering. Kortom, er zijn veel online én offline factoren en mechanismen die kunnen bijdragen aan rechtsextremistische radicalisering. De rol van aanbevelingsalgoritmen lijkt hierbij beperkt. Al met al betekent dit dat het tegengaan van rechtsextremistische radicalisering een aanpak vereist waarbij aan meerdere knoppen tegelijkertijd gedraaid moet worden om tot gewenste resultaten te kunnen komen.

### **Moderatie extremistische content is verbeterd, ook gaan focussen op zoekgedrag gebruikers**

Uit de literatuurstudie én uit de trials op YouTube en TikTok komt naar voren dat deze mainstream sociale media platforms de moderatie van extremistische (illegale) content beter op orde lijken te krijgen. Gebruikers krijgen via aanbevelingsalgoritmen minder extremistische content gesuggereerd. Een belangrijke bevinding van de trials is dat de algoritmen op basis van neutrale zoekwoorden geen extremistische (illegale) content en beperkt polariserende borderline content aanbevelen. Om meer suggesties voor borderline content te krijgen moet gezocht worden door gebruikers met niet-neutrale zoekwoorden. Een aanbeveling naar de platforms is dat zij dus óók moeten focussen op ook het gedrag van gebruikers zélf.



### **'Kuil' van borderline content**

Effecten van contentmoderatie van illegale extremistische content op de platforms zijn zichtbaar. Desalniettemin blijft er een grijs gebied van content over: content die ervaren kan worden als onwenselijk, maar die niet op grond van wettelijke kaders verwijderd kan worden. In dit verband wordt er gesproken van 'borderline content'. Polariserende borderline content heeft op de platforms grotendeels de plaats ingenomen van extremistische content. Borderline content tast de grenzen van de contentmoderatie af die door het gebruik van gecodeerde taal, symbolen en ambigue taalgebruik steeds meer expertise vereist om deze correct te identificeren. Het gebrek aan deze expertise is op dit ogenblik een zwakke schakel in de strijd tegen borderline content op de platforms. De bevindingen uit de literatuurstudie en trials suggereren dat er zeker nog ruimte is voor verdere verbetering om borderline content accurater te detecteren en te weren uit de tijdlijn van de gebruikers. Een effectieve aanpak van borderline content staat of valt met de correcte identificatie ervan. Een rol hierbij kan worden gespeeld door Trusted Flaggers (conform de nieuwe DSA wetgeving: instanties of personen met bijzondere expertise en competentie voor het opsporen, identificeren en melden van illegale content of andere overtredingen van de wet op onder meer sociale mediaplatforms).

### **Mogelijke handelingsperspectieven voor de overheid**

De nieuwe Europese wetgeving, in het bijzonder de Digital Services Act (DSA), en daarnaast de Digital Markets Act (DMA) en de European Media Freedom Act, geeft overheden meer mogelijkheden dan in het verleden om de verspreiding van extreme – en borderline content tegen te gaan. De handelingsperspectieven richten zich op het aanpakken van de volgende problemen:

1. Polariserende borderline content lijkt grotendeels de plaats van extremere content ingenomen op mainstream platforms. Er is sprake van tekortschietende moderatie van borderline content op de grote sociale mediaplatforms. Er bestaan onduidelikheden over wat er onder borderline content moet worden verstaan.
2. Er is sprake van een gebrek aan transparantie over de werking van aanbevelingsalgoritmen op de grote sociale mediaplatforms.
3. Er is sprake van tekortschietende moderatie van extreme content op alt-tech mediaplatforms.
4. Het zoek/surfgedrag van gebruikers zelf speelt een belangrijke rol bij blootstelling aan radicale content.

Het handelingsperspectief van de Nederlandse overheid naar de zeer grote online platforms of overige online platforms met hoofdvestiging in het buitenland is beperkt, aangezien de Europese Commissie respectievelijk het land van vestiging hier de rol van toezichthouder heeft. Wel kan Nederland 'Trusted Flaggers' identificeren, betrekken en ondersteunen; en trachten invloed uit te oefenen op de Europese Commissie bij de implementatie van DSA wetgeving, zoals op prioriteiten toezicht of handhaving.

Meer handelingsperspectieven lijken mogelijk ten aanzien van minder grote platforms. De handelingsperspectieven zijn te categoriseren langs vier lijnen:

1. Het sneller verwijderen van extremistische, illegale content op de kleinere platforms. De DSA lijkt vooral gericht op de hele grote platforms, terwijl er ook risico's bestaan bij de minder grote en kleine platforms. Een mogelijke aanbeveling ter verbetering is dat vooral de kleinere platforms meer zouden kunnen doen om sneller illegale content te verwijderen, en de overheid zou hierin ook een rol kunnen spelen. De DSA lijkt mogelijkheden te bieden voor de beoogd coördinator (ACM) om hierover in gesprek te gaan met kleinere online platforms, in het bijzonder de in Nederland gevestigde online platforms.

2. Samenwerking met platforms Investeren in samenwerking tussen overheden, online platforms en onderzoekers. Op dit moment hebben overheden en gebruikers geen of weinig inzicht in de aanbevelingsalgoritmen die sociale mediaplatforms gebruiken. Relevant is daarom de bepaling in de DSA die voorschrijft dat sociale mediaplatforms transparanter dienen te zijn richting haar gebruikers, overheden en onderzoekers. Dit vraagt bijvoorbeeld om transparantie over hun beleid omtrent contentmoderatie en het gebruik van aanbevelingsalgoritmen. Om een beter beeld te krijgen van de problematiek, ook omtrent de aanpak van borderline content, is het aan te bevelen dat zowel overheden als platforms investeren in een goede samenwerking met onderzoekers.

- Toewerken naar meer co-regulering tussen platforms en overheid waarbij de overheid en online platforms samenwerken om problemen aan te pakken.
- Er zijn momenteel nog weinig handvatten omtrent de omgang met borderline content. Het is ingewikkeld voor online platforms als de overheid zegt dat zij borderline content moeten aanpakken, terwijl duidelijkheid ontbreekt over wat precies aangepakt moet worden. De overheid zou kunnen investeren in het beter begrijpen van de problematiek rondom borderline content en faciliteren dat hierover meer duidelijkheid komt. Aangezien de overheid zelf terughoudend dient te zijn in het beoordelen van 'harmful content' is hier primair een rol weggelegd voor 'trusted flaggers' met kennis van zaken over mogelijke schadelijke effecten en illegaliteit van border line content. Een groot probleem met de aanpak van borderline content is immers de beoordeling wanneer deze content illegaal is, zoals bij bijvoorbeeld beeldspraak en ironie waarin extremistische content wordt verpakt. 'Freedom of speech' is echter niet hetzelfde als 'freedom of reach'. Online platforms kunnen zulke borderline content weliswaar niet verbieden maar ze kunnen wel de amplificatie ervan belemmeren, wat ze in de praktijk overigens ook al doen. Als betrouwbare flaggers meer duidelijkheid scheppen over de illegaliteit en schadelijkheid van bepaalde borderline content, dan zal

deze vorm van contentmoderatie door online platforms makkelijker ingezet kunnen worden.

- Om te voorkomen dat contentmoderatie minder effectief is bij niet-Engelstalige content, is een mogelijk handelsperspectief voor de overheid om hierover in gesprek te gaan met de online platforms, transparantie te vragen hoe hiermee wordt omgegaan en hen zonodig te stimuleren hier meer aandacht aan te geven.
  - Centraal stellen van publieke waarden. Het beleid omtrent sociale media platforms zou niet alleen gebaseerd moeten zijn op economische en juridische uitgangspunten, maar ook op publieke waarden. Hier zou meer discussie over moeten zijn en er zou meer ruimte en randvoorwaarden moeten worden gecreëerd om te experimenteren. De overheid zou dit kunnen faciliteren.
3. Samenwerking tussen overheden / organisaties / ruimte voor experimenteren
- Het up-to-date-houden van kennis omtrent online platforms, algoritmen en contentmoderatie is belangrijk en vergt afstemming tussen ministeries. Liever geen losstaande onderzoeken maar deze met elkaar in verband brengen (continue kennisopbouw).
  - Als alternatief voor de grote internationaal werkende online platforms is het te overwegen om de mogelijkheden te onderzoeken of steun te geven aan initiatieven voor lokaal werkende online platforms of non-profit sociale media. De grote platforms hebben enorme aantallen gebruikers; de schade neemt toe zodra extreme- en borderline content grootschalig kan worden verspreid. Binnen een ecosysteem van kleinere lokale platforms speelt dit probleem minder. Bovendien kunnen gebruikers dan kiezen voor lokale platforms waar effectief en/of volgens bepaalde ethische normen wordt gemodereerd en de kans dus kleiner is dat ze met extreme en borderline content in aanraking komen. Het advies is hierbij wel omwille van de haalbaarheid aan te sluiten bij bestaande initiatieven, niet om zelf als overheid nieuwe initiatieven te initiëren.

- Proactief betrekken van jeugdwerkers. De rol van jeugdwerkers in de EU-wetgeving is onderbelicht. Zij weten het beste wat bij de jeugd actuele polariserende en radicaliserende content is. Zij zouden betrokken kunnen worden als trusted flaggers. De overheid kan hier proactief een rol in spelen door hierover in gesprek te gaan met beroepsorganisaties en 'koplopers' van jeugdwerkorganisaties.
  - Online zouden talrijke influencers die via grote sociale mediaplatforms of gameplatforms een ingang kunnen zijn voor (jonge) personen om zich online verder te verdiepen in bepaald gedachtegoed. Ook subculturen kunnen een 'gateway' zijn naar rechtsextremistisch gedachtegoed. Deze invloedrijke rol van influencers en subculturen (en de risico's hiervan) als 'gateways naar rechtsextremistisch gedachtegoed' kan met nader wetenschappelijk worden onderzocht om meer bewustzijn te creëren rondom dit thema.
4. Inzet op offline aanpakken.
- Een handelingsperspectief voor de overheid kan zijn om meer te investeren in kritisch denken, digitale weerbaarheid en mediawijsheid. Het belang van individuele weerbaarheid van (in het bijzonder) jongeren, en dan in het bijzonder het vermogen om kritisch en flexibel te kunnen denken, neemt in de online context alleen maar toe. Extremistische content zal immers altijd beschikbaar zijn op het onuitputtelijke internet, wat maakt dat het in de eerste plaats uiterst belangrijk is om mensen weerbaar te maken voor deze en andere radicale/borderline content. Inmiddels zijn er meerdere geëvalueerde lesprogramma's beschikbaar die breder ingezet kunnen worden online weerbaarheid en mediawijsheid te vergroten.
  - Dat jongerencultuur altijd deels onbegrepen zal blijven voor (volwassen) gezagsdragers betekent niet dat we ons hier bij moeten neerleggen. Overheden kunnen investeren in het bevorderen van meer inzicht bij ouders, docenten, jeugdprofessionals en gezagsdragers in de online leefwereld van jongeren. Zo kunnen we beter begrijpen wat de aantrekkingskracht van rechts-extremisme kan zijn en hoe jongeren zo'n narratief en/of netwerk ingelokt kunnen worden.
  - Het bevorderen van meer bewustwording bij werkgevers omtrent het belang van vergroten van online vaardigheden van jeugdwerkers.
  - Opvallende bevinding in de trials op YouTube was dat de meest radicale content te vinden was onder (in de commentsecties) van artikelen in de nieuwsmedia. Dergelijke berichten kunnen aantrekkingskracht hebben op personen of bepaalde reacties van hen uitlokken. Er worden veel mediaberichten becommentarieerd; opnieuw geplaatst; en er staan boze of opruiende comments onder video's waarin enkel nieuws wordt weergegeven. Suggestie zou kunnen zijn dat de nationaal coördinator hierover het gesprek aangaat met nationale of internationale nieuwsmedia om hen hiervan bewuster te maken. Platforms en media kunnen onder berichtgeving weblinks naar extremere content verwijderen.

# Summary

## Main findings

This research has mapped different development paths of right-wing extremist radicalisation and the role played in these processes by social media platforms, with a specific focus on the workings of recommendation algorithms. We also examined the role of the offline context in development paths of right-wing extremist radicalisation. Based on the findings, possible viable action perspectives have been described for the government, to prevent right-wing extremist radicalisation via online platforms.

The study's most important findings are the following:

- A large number of risk factors at individual, group and societal levels contribute to the onset of radicalisation processes. Development pathways of radicalisation are extremely complex, involving, and interacting with, both online and offline factors. Understanding the influence of the online context on extremist beliefs requires an examination of this interaction.
- The role of recommendation algorithms in right-wing extremist radicalisation within the online context is limited, according to recent academic research. The personal choices of users are at least as important, such as sharing links to extremist content; subscribing to/following certain channels on social media platforms and thus being selectively confronted with more extreme content; meeting like-minded people in echo chambers; and personal surfing/searching behaviour. People search for more extreme content on their own accord and, to this end, deliberately enter certain search terms, or click through on the more extreme recommendations.
- Mainstream social media platforms seem to be getting better at moderating extremist (illegal) content. Users get less extremist content suggested via recommendation algorithms. However, some disclaimers can be added to these conclusions:
  - Extremist content has been moved to alternative, 'alt-tech' platforms.
  - On mainstream social media platforms, extremist content seems to have been largely replaced by polarising borderline content (also referred to as 'legal yet harmful' content). A major problem in dealing with borderline content is assessing when it crosses the line and becomes illegal due to the use of coded language, symbols, ambiguous language, imagery and irony in which extremist content is packaged. It requires increasing expertise to identify these. The lack of this expertise is a weak link in the fight against borderline content on platforms.
  - Platform users may still come into contact with extremist content through their own online behaviour, such as sharing web links, subscribing to or following certain channels and seeking/finding like-minded people in certain online (closed) chat groups or forums.
- Possible solutions to prevent (online) right-wing extremist radicalisation therefore include focusing on the problem of non-illegal, extreme (borderline) content and tools to better understand and moderate this content; paying more attention to illegal content on alt-tech media platforms; as well as placing the responsibility, in part or in whole, back with users themselves, since their own behaviour can make them more likely to come into contact with borderline or illegal content.
- The action perspective of the Dutch government with respect to the very large online platforms is limited, as the role of supervisor here is played by the European Commission and the country of establishment, respectively. More action prospects seem possible with regard to the smaller platforms.

These perspectives can be categorised along four lines:

- Commitment to a dialogue with smaller platforms, to get extreme illegal content removed faster.
- Investment in collaboration between governments, online platforms and researchers on greater transparency about recommendation algorithms; working towards more co-regulation between platforms and the government; investment in a better understanding of the issues surrounding borderline content and facilitating greater clarity about the potentially harmful effects and illegality of borderline content.
- Investment in space for experimentation, and exploration of opportunities or support for initiatives for locally operating online platforms or non-profit social media.
- A focus on offline approaches by investing more in digital resilience and media literacy. To this end, several evaluated teaching programmes are now available that can be used more widely.

## Summary and conclusions

### Research objectives

The Research and Documentation Centre (WODC) has asked the Verwey-Jonker Institute and TextGain to conduct a study to gain a better understanding of the development pathways for the occurrence of right-wing extremist radicalisation, focusing on the role of online social media platforms, recommendation algorithms and the interaction between online and offline contexts.

The objectives of the study were as follows:

- What are the development paths for the emergence of right-wing extremist radicalisation and what is the role of online social media platforms in this?
- What influence do the recommendation algorithms applied on social media platforms (still) have on the right-wing extremist radicalisation of the users of these platforms?

- In addition to online radicalisation, is offline recruitment also a part of right-wing extremist radicalisation (an interaction between online and offline contexts)?
- What viable courses of action are there to counter right-wing extremist radicalisation via online social media platforms?

### What information does this report contain?

The study has resulted in a report that provides an accessible overview of:

- Theories about, and up-to-date insights into, development paths of radicalisation, in particular right-wing extremist radicalisation in the online context, including:
  - Risk and protective factors involved in the susceptibility to radicalisation.
  - Mechanisms of radicalisation within the online domain.
  - The relationship between online right-wing extremist radicalisation and the use of violence.
- Definitions, delineation, content characteristics, and narratives of (right-wing) extremist content.
- Online strategies to bring right-wing extremist content to the life world of people, such as:
  - The use of humour, irony and dog whistles.
  - Mainstreaming of (right-wing) extremist content and red-pilling.
  - Normalisation, acclimatisation and dehumanisation.
- Right-wing extremist content on mainstream, alternative, and gaming platforms.
- Online subcultures and influencers as 'gateways' to right-wing extremist content, including: Conspirituality and Qanon; the Manosphere and masculine subcultures; the Incels subculture; the Viking culture; the Preppers community.
- Recommendation algorithms as a social media trap?
  - What are recommendation algorithms and how do they work?
  - Recommendation algorithms on YouTube and TikTok.

- Is there evidence of recommendation algorithms leading to social media traps and radicalisation?
- Trials regarding recommendation algorithms on YouTube and TikTok, based on personalised search behaviour within online subcultures (examples: Manosphere/masculine subculture and Viking culture).
- A description of some case histories of right-wing extremist radicalisation in which online social media traps may have played a role.
- New regulations and action perspectives:
  - Nuances regarding the role of recommendation algorithms in right-wing extremist radicalisation.
  - What are social media platforms already doing in terms of moderation of algorithms and content?
  - Social media policy in the Netherlands and the European Union; new possibilities through the Digital Services Act.
  - Prospects for action, both online and offline.

### Research design

The methodical execution of this research consisted of several components. The first part was an extensive state-of-the-art literature study on insights into the scientific literature on development paths of online (right-wing extremist) radicalisation and the role played in this by different 'types' of social media. The second part involved interviews with experts on the workings and effects of recommendation algorithms, their influence compared to other mechanisms in the online and offline contexts, and feasible options for the government to counter online, right-wing, extremist radicalisation. The draft findings of the report were presented in an interview with two jurists with knowledge of the new European legislation on social media platforms (Digital Service Act) and in two expert sessions, to experts in the field of the workings and effects of recommendation algorithms; media and digital society; and social media and their social impact. During these expert sessions, we reflected on the results of the literature study and - above all - reflected together with the experts on feasible and (thus) practical action perspectives for the government. These

research activities were carried out by the Verwey-Jonker Institute. The third part of the study was undertaken by TextGain and involved conducting online trials on YouTube and TikTok, which investigated the extent to which algorithms based on users' personalised search behaviour lead to suggestions for (more) extreme content. We conducted these trials as a complement to the literature review. The study was supervised by a supervisory committee with representatives from the Ministry of Justice and Security and the WODC, and academics from the University of Amsterdam, Maastricht University/Free University Amsterdam and Leiden University.

### Results

This research mapped out different development paths of right-wing extremist radicalisation and the role played in these by social media platforms. Specific attention was paid to the workings of recommendation algorithms. We also examined the role of the offline context in development paths of right-wing extremist radicalisation. Based on the findings, action perspectives were described that are considered feasible for the government to contribute to the prevention of right-wing extremist radicalisation via online platforms.

#### *Great variety in risk factors and development paths*

The scientific literature shows that a large number of risk factors at individual, group and societal levels may contribute to the emergence of radicalisation processes. In addition, there are many more similarities than differences in the characteristics, risk factors and mechanisms of radicalisation processes from diverse ideological backgrounds (left- and right-wing extremism, religiously motivated and single-issue extremism). Development paths of radicalisation are extremely complex, with both online and offline factors playing a role and interacting in this process. To understand the influence of the online context on extremist beliefs and violence, it is necessary to look at this interaction. Based on empirical studies, it appears that most individuals who radicalise online do not engage in violent behaviour. There are very disparate ways in which right-wing extremist beliefs can enter individuals' online lives.

Various online strategies of right-wing extremist individuals and networks can contribute to the exposure and acceptance of right-wing extremist content. A major strategy here is to 'disguise' right-wing extremist content through the use of humour, irony, and dog whistles. Long-term exposure to right-wing extremist content and strategies can make individuals increasingly susceptible to them. The literature identifies several so-called 'cognitive phases' for this purpose. These include the following stages: normalisation, acclimatisation, and dehumanisation. These stages overlap, are not exclusive or strictly linear, but rather cyclical in nature.

#### *Gateways to right-wing extremist radicalisation*

Radical and right-wing extremist groups are active on various social media platforms, but still have the greatest reach through mainstream platforms such as Facebook, Twitter, YouTube and, to a slightly lesser extent, Instagram. In recent years, these major platforms have taken stern action against channels and pages linked to right-wing extremist groups or alternative media. Alternative or 'fringe' platforms such as Gab, Telegram and 4chan have become more popular among right-wing extremist groups, partly because more extreme content can be posted there. As the reach is smaller on alternative platforms, more 'moderate' material that meets the platform's conditions is posted on the larger platforms. However, it seems that through this 'moderate' content, users are sometimes redirected via links, or search for themselves, to find more extreme content elsewhere. Online subcultures that are not necessarily right-wing extremist (such as Conspirativity, the Manosphere and Viking subculture) can also act as gateways to right-wing extremist thought. As described in section 3.7, individuals within these subcultures may be more likely to encounter right-wing extremist ideology. Some narratives disseminated within these subcultures share similarities with right-wing extremist narratives. Another possible entry point to extremist content is provided by some popular online influencers. Online, numerous influencers can be found who, through major social media platforms or gaming platforms, can present an entry point for (young) individuals to delve further into certain ideas or communities online.

#### *Online and offline recruitment*

From the academic literature, relatively little is known about case histories in which it has been established that a social media trap contributed to right-wing extremist radicalisation or recruitment. Establishing this is not easy either, because the workings of recommendation algorithms and filter bubbles do not take place in an isolated universe. After all, there also are other online and offline mechanisms that can contribute to a person's radicalisation. Besides, it is difficult to measure whether someone has been radicalised specifically by the social media trap or whether other factors have made a greater contribution. In short, it is hardly possible to conclude that an individual has been radicalised solely through the consumption of one-sided online content on social media platforms and that the person would not have been radicalised without it. Nevertheless, there are anecdotal indications (cases) that make it plausible that the social media trap played a role in the radicalisation process. However, the question then is to what extent this trap is caused by, for example, recommendation algorithms on social media platforms, or whether there is an effect that seems a trap, but that is rather more attributable to users' own online behaviour (choices).

#### *The role of recommendation algorithms in right-wing extremist radicalisation is limited*

As pointed out above, the mechanism of an "algorithm-driven supply of right-wing extremist content on social media platforms that can lead to a right-wing extremist social media trap" takes place in anything but a vacuum. The role of recommendation algorithms and social media traps in right-wing extremist radicalisation seems smaller than is sometimes assumed in public discourse. Individual persons may encounter polarising or extremist content in a variety of ways, both online and offline. The various personal choices users make are relevant in this regard:

- Within their social network, people share links to extremist content with each other.
- People subscribe to / follow certain channels on social media platforms and are in this way selectively confronted with more extreme content.

- Like-minded people meet in echo chambers such as chat groups on Telegram.
- Searching behaviour: people search for more extreme content of their own accord and to this end deliberately enter certain search terms, or click further on the more extreme recommendations.

Also questionable is the role an sich of the internet and social media platforms in violent radicalisation. As stated earlier, there is little empirical evidence that radicalisation in the online context leads to violent extremism, nor that violent extremism directly results from online radicalisation. In short, there are many online as well as offline factors and mechanisms that may contribute to right-wing extremist radicalisation. The role of recommendation algorithms seems limited in this regard. All in all, this means that countering right-wing extremist radicalisation requires an approach in which several wheels have to be spun simultaneously in order to achieve desired results.

*Moderation of extremist content has improved, also start focusing on users' search behaviour*

Both the literature review and trials on YouTube and TikTok show that these mainstream social media platforms seem to be getting a grip on the moderation of extremist (illegal) content. Users are suggested less extremist content via recommendation algorithms. An important finding of the trials is that, based on neutral keywords, the algorithms do not recommend extremist (illegal) content and limited polarising borderline content. To get more suggestions for borderline content, users should search with non-neutral keywords. A recommendation to platforms is that they should also focus on the behaviour of the users themselves.

*'Pit' for borderline content*

The effects of content moderation of illegal, extremist content on the platforms are visible. Nevertheless, a grey area of content remains: content that may be perceived as undesirable but cannot be removed on the basis of legal

frameworks. In this context, it is known as 'borderline content'. Polarising borderline content has largely replaced extremist content on platforms. Borderline content tests the boundaries of content moderation which, through the use of coded language, symbols and ambiguous language, requires increasing expertise to identify it correctly. The lack of this expertise is currently a weak link in the fight against borderline content on platforms. The findings from the literature review and trials suggest that there is definitely room for further improvement to detect and keep borderline content out of users' timelines more accurately. An effective handling of borderline content hinges on its correct identification. A role may be played here by Trusted Flaggers (in accordance with the new DSA legislation: bodies or individuals with special expertise and competence for detecting, identifying and reporting illegal content or other violations of the law on social media platforms, among others).

*Possible courses of action for the government*

The new European legislation, in particular the Digital Services Act (DSA), as well as the Digital Markets Act (DMA) and the European Media Freedom Act, gives governments more opportunities than in the past to combat the distribution of extreme - and borderline - content. The action perspectives focus on addressing the following issues:

5. Polarising borderline content seems to have largely replaced more extreme content on mainstream platforms. The moderation of borderline content on major social media platforms is inadequate. There is confusion about what should be understood to be borderline content.
6. Transparency about the workings of recommendation algorithms is lacking on the big social media platforms.
7. There is inadequate moderation of extreme content on alt-tech media platforms.
8. The searching / surfing behaviour of users themselves plays an important role in exposure to radical content.



The Dutch government's action perspective regarding very large online platforms or other online platforms with headquarters abroad is limited, as the European Commission and the country of establishment, respectively, have the role of supervisor here. However, the Netherlands can identify, engage and support 'Trusted Flaggers', and try to influence the European Commission in the implementation of DSA legislation, such as on surveillance or enforcement priorities.

More action perspectives seem possible with respect to smaller platforms. The action prospects can be categorised along four lines:

1. Faster removal of extremist, illegal content on the smaller platforms. The DSA seems mainly focused on the very large platforms, while risks also exist on the less large and small platforms. One possible recommendation for improvement is that the smaller platforms in particular should do more to remove illegal content faster, and the government might also play a role in this. The DSA seems to offer opportunities for the intended coordinator (ACM) to enter into discussions with smaller online platforms on this issue, especially those based in the Netherlands.
2. Cooperation with platforms.
  - Invest in collaboration between governments, online platforms and researchers. Currently, governments and users have little or no insight into the recommendation algorithms used by social media platforms. Relevant, therefore, is the provision in the DSA that requires social media platforms to be more transparent towards their users, governments and researchers. This requires, for instance, transparency about their policies on content moderation and their use of recommendation algorithms. To get a better picture of the issues, also regarding the handling of borderline content, it is recommended that both governments and platforms invest in a good collaboration with researchers.
  - Working towards more co-regulation between platforms and the government, whereby the government and online platforms collaborate to address problems.
- At the moment, there still are few guidelines on how to deal with borderline content. It is complicated for online platforms when the government says they have to take action against borderline content, while clarity on what exactly should be addressed is lacking. The government might invest in a better understanding of the issues surrounding borderline content and facilitate greater clarity on this. Since the government itself should be cautious in assessing 'harmful content', this is primarily a role for 'trusted flaggers' with knowledge about possible harmful effects and the illegality of borderline content. After all, a major problem in dealing with borderline content is assessing when it is illegal, as in the case of, for example, imagery and irony in which extremist content is packaged. However, 'freedom of speech' is not the same as 'freedom of reach'. While online platforms cannot ban such borderline content, they can hinder its amplification, which, incidentally, they are already doing in practice. If trusted flaggers become more articulated about the illegality and harmfulness of certain borderline content, this form of content moderation will be easier to deploy by online platforms.
- To prevent content moderation from being less effective in the case of non-English-language content, a possible action perspective for the government is to engage in a dialogue with online platforms about this, to ask for transparency on how this is handled and, if necessary, encourage them to pay more attention to it.
3. Focus on public values. Policy on social media platforms should not only be based on economic and legal principles, but also on public values. There should be more discussion about this, and more space and preconditions should be created to experiment. The government might facilitate this.
4. Cooperation between authorities / organisations / room for experimentation.
  - Keeping knowledge about online platforms, algorithms and content moderation up-to-date is important and requires coordination between ministries. Rather than separate studies, link them together (continuous knowledge building).

- As an alternative to the large, internationally operating online platforms, consideration should be given to exploring or supporting initiatives of locally operating online platforms or non-profit social media. The large platforms have huge numbers of users; the damage increases once extreme and borderline content can be widely distributed. Within an ecosystem of smaller local platforms, this problem plays less of a role. Moreover, users can then opt for local platforms that are moderated effectively and/or according to certain ethical standards, and are therefore less likely to encounter extreme and borderline content. The advice here is to join existing initiatives for the sake of feasibility, not to initiate new ones yourself as the government.
  - Proactively engage youth workers. In EU legislation, too little attention is paid to the potential role of youth workers. They know best what, among the young, is currently polarising and radicalising content. They might be engaged as trusted flaggers. The government can play a proactive role in this by interacting with professional organisations and 'frontrunners' of youth work organisations on this.
  - Online, numerous influencers operating through major social media platforms or gaming platforms might provide an entry point for (young) individuals to delve further into a certain ideology online. Subcultures can also be a 'gateway' to right-wing extremist thought. This influential role of influencers and subcultures (and the potential risks thereof) as 'gateways to right-wing extremist ideology' can be investigated with further scientific research to create more awareness around this topic.
5. Focus on offline approaches.
- An action perspective for the government could be to invest more in critical thinking, digital resilience and media literacy. The importance of individual resilience of (especially) young people, in particular the ability to think critically and flexibly, is only increasing in the online context. After all, extremist content will always be available on the inexhaustible internet, which makes it extremely important to make people resilient to this and other radical/borderline content in the first place. Meanwhile, several evaluated teaching programmes are available that can be used more broadly to increase online resilience and media literacy.
  - That youth culture will always remain partly misunderstood by (adult) authority figures does not mean we should resign ourselves to this. Governments can invest in promoting greater understanding among parents, teachers, youth professionals and authority figures in the online living environment of young people. This will allow us to better understand what the appeal of right-wing extremism can be and how young people can be lured into such a narrative and/or network.
  - Promote greater awareness among employers about the importance of increasing the online skills of youth workers.
  - A striking finding in the trials on YouTube was that the most radical content was found at the bottom (in the comment sections) of articles in the news media. Such posts may appeal to individuals or elicit certain reactions from them. Many media reports are commented on, and reposted; and there are angry or inflammatory comments under videos showing only news. A suggestion could be for the national coordinator to start a conversation with national or international news media about this to make them more aware of it. Platforms and media can remove web links to more extreme content under news coverage.

# 1 Introductie

## 1.1. Probleemschets van rechtsextremisme in het online domein

Rechtsextremisme lijkt bezig aan een opmars in Nederland. Deze speelt zich voor een belangrijk deel online af en onderscheidt zich daarom op belangrijke punten van het 'traditionele', meer op fysieke ontmoetingen geënte rechtsextremisme. Er is zelfs sprake, zo meldt ook het Dreigingsbeeld Terrorisme Nederland (nr. 55), van een geheel nieuwe generatie rechtsextremisten. Er zijn meerdere factoren die dit tot een zorgwekkende ontwikkeling maken. Allereerst kan online radicalisering zich heel snel voltrekken. Vanwege de anonimiteit die het computerscherm biedt, zijn conversaties binnen chatgroepen veelal extremer, waarbij nieuwkomers in dergelijke groepen zich razendsnel conformeren aan de groepsnormen, zonder daarbij te beseffen dat sommige uitspraken voortkomen uit overdrijvingen en/of stoerdoenerij. Wagenaar (in: Peeters et al., 2022) spreekt in dit verband van flitsradicalisering. Studies laten zien dat algoritmen (waaronder: aanbevelingsalgoritmen) op sociale mediaplatforms kunnen bijdragen aan steeds extremer wordende denkbeelden bij gebruikers, omdat deze kunnen functioneren als zogeheten 'filterbubbels' of 'sociale mediafijken' (O'Callaghan et al., 2015; Ribeiro et al., 2019; Reed et al., 2019; Whittaker et al., 2021; Keller, 2021). Door deze amplificerende werking van deze algoritmen kan iemand in een 'koninginnenhol' van steeds extremere content belanden. Een persoon blijft dan graven 'op zoek naar antwoorden' en vindt als resultaat (nog) extremere content (Van Puffelen, 2021). Belangrijk om echter te noemen is dat dit beeld over de aanjagende rol van algoritmen bij online radicalisering in recente studies wordt genuanceerd. Ontwikkelingspaden naar rechtsextremistische content op sociale mediaplatforms zijn zeer divers.

Zo blijkt ook het eigen (zoek)gedrag van internetgebruikers belangrijk voor het in aanraking komen met radicale of extremistische content (Ledwich & Zaitsev, 2019; Hosseinmardi et al., 2020; Chen et al., 2022; Brown et al., 2022). Ook is er sprake van een bijzonder risico op blootstelling aan extremistische content op zogenoemde 'alt-tech' platforms. Deze platforms werken weliswaar minder of niet met (aanbevelings)algoritmen, maar hier wordt extremistische content – in vergelijking met mainstream platforms zoals Facebook, YouTube en TikTok – minder goed gereguleerd en gemodereerd.

### **Definitie: radicalisering**

De term 'radicalisering' roept veel verschillende beelden op. Deze veelheid aan definities maakt duidelijk dat radicalisering een dynamisch proces is, met verschillende invloeden op verschillende personen in verschillende contexten (Dijkman, 2021; Stephens & Sieckelinck, 2021). Radicale ideeën zijn op zichzelf niet altijd zorgwekkend. Bij jongeren passen radicale ideeën soms bij de zoektocht naar een eigen identiteit (Stephens & Sieckelinck, 2021). Onderscheid kan daarom gemaakt worden tussen 'radicalisering' en 'radicalisering die kan leiden tot extremistisch geweld' (Caramellino et al., 2022). In dit onderzoek gaat het om de tweede variant. In navolging van de Nationaal Coördinator Terrorismebestrijding en Veiligheid hanteren wij de volgende definitie: "Een proces van toenemende bereidheid om de uiterste consequentie uit een denkwijze te aanvaarden en die in daden om te zetten. Deze toenemende bereidheid kan leiden tot gedrag dat andere mensen diep kwetst of in hun vrijheid raakt, kan aanleiding zijn voor individuen of groepen om zich af te keren van de samenleving en kan leiden tot het gebruik van geweld" (NCTV, 2021a).

### **Definitie: rechtsextremisme en -terrorisme**

Wat is rechtsextremisme en hoe verhoudt zich dit tot extreemrechts? Deze definities zijn in de literatuur en beleid niet altijd eenduidig. De Nationaal Coördinator Terrorismebestrijding en Veiligheid (NCTV) definieert extreemrechts als volgt: "Tot extreemrechts gedachtengoed worden denkbeelden gerekend als vreemdelingenhaat (inclusief: antisemitisme en antimoslims) en haat jegens vreemde (cultuur)elementen en ultranationalisme. Hierbij houden extreemrechtse activisten zich veelal aan de wet, overtreden rechtsextremisten de wet en extreemrechtse terroristen plegen geweld gebaseerd op extreemrechtse opvattingen (NCTV, 2021a)." Mudde (2000) voegt aan deze denkbeelden van extreemrechts 'antidemocratisch gedachtengoed' en 'een voorkeur voor een sterke staat' toe. Opgemerkt moet worden dat er in de wetenschap en in het beleid nog discussies plaatsvinden over de precieze definiëring van 'extreemrechts'. Dit heeft mede te maken met de complexe en steeds veranderende context van radicalisering waarin voortdurend nieuwe verschijningsvormen opkomen. Zo zijn er bijvoorbeeld discussies over in hoeverre anti-overheidsradicalisering zou behoren tot of overlappen met extreemrechtse narratieven.

Gekeken naar het rechtsextremisme, wijst de Algemene Inlichtingen- en Veiligheidsdienst (AIVD) op "groeperingen en individuen die vanuit (een van) de [hierboven genoemde] (...) denkbeelden acties ondernemen die een bedreiging vormen voor de democratische rechtsorde" (AIVD, 2023a). Deze bedreiging van de democratische rechtsorde door (rechts)extremisten bestaat uit het nastreven van antidemocratische doelen en/of het inzetten van ondemocratische middelen. Ondemocratische middelen kunnen zowel gewelddadig (bijvoorbeeld geweldplegingen, mishandelingen en ernstiger vormen van geweld) als niet-gewelddadig zijn (bijvoorbeeld het systematisch zaaien van haat, het verspreiden van angst of desinformatie of het bijdragen aan demonisering of intimidatie) (AIVD, 2023b).

In deze literatuurstudie baseren we ons op de volgende driedeling van de AIVD: (1) activisme, dit kenmerkt zich door de afwezigheid van ondemocratische doelen of middelen (en is daarmee niet in strijd met de wet); (2) extremisme, dit kenmerkt zich door het nastreven of gebruiken van ondemocratische doelen en middelen, waardoor het in strijd is met de wet, maar niet per definitie gewelddadig is; en (3) verondersteld wordt dat extremisme in zijn uiterste vorm kan leiden tot terrorisme. Onder terrorisme vallen acties die maatschappij-ontwrichtende schade aanrichten en/of grote groepen mensen angst aanjagen (AIVD, 2023b).

Internet en sociale mediaplatforms kunnen op verschillende manieren een rol spelen in de ontwikkelingspaden van rechtsextremistische radicalisering (Peeters et al., 2022; Liang & Cross, 2020; Marwick et al., 2022; ESS, 2021). Voorbeelden hiervan zijn het genereren van aandacht voor en het verspreiden van rechtsextremistisch gedachtengoed; het normaliseren ('mainstreamen') van rechtsextremistische ideeën; het rekruteren en mobiliseren van leden; het vinden van gelijkgestemden waarmee gemeenschappen, netwerken en subculturen gevormd kunnen worden; het zoeken naar ideologische rechtvaardigheid; het ondermijnen van sociale cohesie, de overheid en institutioneel vertrouwen; en tot slot, het verweven van rechtsextremistische content en 'vermaak' (Marwick et al., 2022; ESS, 2021). In de literatuur is thans discussie over de vraag in hoeverre het gebruik van internet op zichzelf radicalisering kan veroorzaken of dat juist de wisselwerking met 'offline' factoren noodzakelijk is hiervoor (Marwick et al., 2022; Whittaker, 2022). Ook wanneer het laatstgenoemde wordt verondersteld, moet worden gerealiseerd dat langdurige online blootstelling aan extremistische content kan bijdragen aan een proces van aanscherpen van tegenstellingen, isolatie van de oude sociale omgeving en verregaande identificatie met een groep en bijbehorende extremistische ideeën, kenmerken en gebruiken. Gewelddadigheid kan hiervan onderdeel zijn (Meleagrou-Hitchens & Kaderbhai, 2017; Marwick et al., 2022).

Het internet biedt toegang tot internationale rechtsextremistische netwerken – die bovendien na verwijdering vaak weer onder een andere naam opduiken (Hart et al., 2021). Dit heeft ertoe geleid dat het huidige rechtsextremisme in Nederland sterker internationaal georiënteerd is dan voorheen, waarbij de invloed van rechtsextremisme vanuit de Verenigde Staten is toegenomen. We zien dit terugkomen in bijvoorbeeld de steun voor de zogeheten internationale accelerationistische netwerken<sup>1</sup> en in de steun voor de omvolkingstheorie, die de basis vormt voor het accelerationisme. Binnen het accelerationisme wordt gesproken over zaken zoals het ontketenen van een rassenoorlog door middel van terroristische aanslagen. Een paar honderd Nederlandse jongeren zouden onderdeel uitmaken van dergelijke netwerken die zich met name online manifesteren. De NCTV (2022) meldde onlangs in het Dreigingsbeeld Terrorisme Nederland nr. 56 dat het gevaar van een aanslag in Nederland met name in deze hoek schuilt: "Zoals ook in het vorige Dreigingsbeeld vastgesteld is binnen het rechts-extremisme het accelerationisme de stroming in Nederland waar de meeste geweldsdreiging vanuit gaat. Enkele honderden Nederlandse accelerationisten zijn online in beeld gekomen; van enkelen kan dreiging uitgaan. Op diverse online netwerken van deze stroming blijkt een fascinatie voor wapens te bestaan". Naast het accelerationisme zijn er in Nederland ook andere online verschijningsvormen van extreemrechts te onderscheiden, zoals anti-overheidsextremisme, samenzweringstheorieën, antisemitisme/holocaustontkenning en de verschillende subculturen van extreemrechts.

Belangrijk is om op te merken dat het internet niet alleen toegang tot een breed ideologisch spectrum biedt – waaronder diverse rechtsextremistische narratieven, maar ook potentiële mogelijkheden biedt om een proces van gewelddadige radicalisering te doorlopen: van een ondersteunend sociaal (online) netwerk, het kopen van wapens tot het surveilleren van doelwitten

<sup>1</sup> Accelerationisme is een stroming binnen het rechtsextremisme. Hierbinnen wordt het creëren of versnellen van chaos nagestreefd, om zo een rassenoorlog en de vervanging van de democratie door een witte etno-staat te bespoedigen (NCTV, 2021). Zie daarvoor: Paragraaf 2.6.

via Google Maps (Feldman in: Damen en Kraak, 2021). Online radicalisering kan daardoor potentieel bijdragen aan het risico op wat lone actor-aanslagen genoemd worden, uitgevoerd door eenlingen die online zijn geradicaliseerd en geïnspireerd worden door aanslagen op andere plekken in de wereld. Deze lone actors zijn minder afhankelijk van netwerken of leiders bij het plegen van aanslagen maar vinden en krijgen wel veelal steun online<sup>2</sup>. Overigens wordt het 'alleen handelen' van een lone actor inmiddels veelal gezien als een achterhaald idee. De suggestie van een lone actor is namelijk dat iemand een radicaliseringsproces en overstap naar geweld als eenling doormaakt. Dat gebeurt zo goed als nooit. Hoewel een lone actor zélf de aanslag pleegt, zijn in veel gevallen zulke lone actor-aanslagen gepleegd door eenlingen die wel een geschiedenis hebben in extreemrechtse netwerken online en offline (Spaaij, 2021; Marwick et al., 2022).

De gevaren van online rechtsextremisme schuilen, tot slot, ook in het grote bereik ervan. Dit zien we onder meer terug in de mate waarin digitale kanalen aanhangers van rechtsextremistisch gedachtegoed in staat stellen om het maatschappelijk debat te beïnvloeden en bijdragen aan een gepolariseerd en vijandig klimaat richting bijvoorbeeld migranten of overheden. Zo infiltreren rechtsextremistische groepen vaak ook in protestbewegingen, zoals die tegen corona- en stikstofmaatregelen (Wagenaar in: Peeters et al., 2022). Het verspreiden van (rechtsextremistisch geïnspireerde) complottheorieën heeft in veel landen bovendien al tot een toename van invloed geleid voor rechtsextremistische groepen (Mulhall & Khan-Ruf, 2021). In Nederland is dit onder andere zichtbaar door een bredere omarming van de omvolkingstheorie.

<sup>2</sup> Er is sprake van discussie omtrent de vraag in hoeverre lone actors daadwerkelijk alleen acteren. Personen die schijnbaar alleen een aanslag plegen, blijken in de praktijk zelden alleen te opereren (zoals de bekende casus van de 'una bomber') maar zijn wel degelijk veelal betrokken bij (offline/online) groepen. Ook benutten zij in meerdere gevallen sociale media om een fanbase op te bouwen.

In de (nabije) toekomst kan 'online' rechtsextremistisch radicalisme zich mogelijk ook in Nederland verder ontwikkelen als gevolg van de laagdrempeligheid, anonimiteit en ongrijpbaarheid. Vanwege het profiel van rechtsextremistische radicalisering, en dan met name de ongrijpbaarheid die het online domein met zich meebrengt, alsook het relatief recente karakter ervan is het onduidelijk in hoeverre de overheid en haar zorg- en veiligheidspartners terug kunnen vallen op de kennis waarover ze al beschikken ten aanzien van (jihadistische) radicalisering en (offline) rechtsextremisme (Winter et al., 2020). Er is om deze reden grote behoefte aan kennis over hoe rechtsextremistische radicalisering online verloopt, wat de ontwikkelingspaden zijn, wat de rol van aanbevelingsalgoritmen hierbij is en welke mogelijke handelingsperspectieven er zijn.

## 1.2. Doelstelling en onderzoeksvragen

Het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) heeft aan het Verwey-Jonker Instituut en TextGain gevraagd een onderzoek uit te voeren om meer inzicht te krijgen in de ontwikkelingspaden voor het optreden van rechtsextremistische radicalisering, met daarin centraal de rol van online sociale media platforms, aanbevelingsalgoritmen en de wisselwerking tussen de online- en offline contexten. Er wordt daartoe een onderscheid gemaakt tussen mainstream en alt-tech platforms. We hebben bovendien specifiek aandacht voor de werking van aanbevelingsalgoritmen die toegepast worden op deze sociale mediaplatforms en de mate waarin offline contacten – naast online invloeden – een noodzakelijke voorwaarde zijn voor het optreden van rechtsextremistische radicalisering. De doelstelling vatten we in de volgende vier hoofdvragen:

- A. Wat zijn de ontwikkelingspaden voor het optreden van rechtsextremistische radicalisering en wat is daarbij de rol van online sociale mediaplatforms?
- B. Welke invloed hebben de aanbevelingsalgoritmen die op sociale mediaplatforms worden toegepast (nog) op rechtsextremistische radicalisering van gebruikers van deze platforms?

- C. Vindt er naast online radicalisering ook offline rekrutering plaats bij rechtsextremistische radicalisering (wisselwerking online en offline contexten)?
- D. Welke haalbare handelingsperspectieven zijn er om rechtsextremistische radicalisering via online sociale mediaplatforms tegen te gaan?

Op basis van deze hoofdvragen zijn er negen onderzoeksvragen geformuleerd. Deze zijn als volgt:

1. Welk(e) mechanisme(n) zijn er in de wetenschappelijke literatuur bekend over radicaliseringsprocessen van individuen en groeperingen in de richting van rechtsextremistisch, gewelddadig gedachtegoed? Wat is hierbij de rol van online content? In hoeverre en op welke aspecten zijn er overeenkomsten óf verschillen in de mechanismen van rechtsextremistische radicalisering ten opzichte van andere vormen van radicalisering?
2. Wat is er bekend over de voorspellende waarde van online radicalisering naar het optreden van gewelddadig rechtsextremisme? In hoeverre is dit uniek voor rechtsextremisme?
3. Wat is er in de wetenschappelijke literatuur bekend over de wijze waarop sociale mediaplatforms en andere online content kan worden geïdentificeerd als extreemrechts en rechtsextremistisch? Hoe kan dit worden onderscheiden van andere extremistische content? Is hierbij een verschil tussen de diverse sociale mediaplatforms?
4. In welke mate zijn er aanwijzingen dat het algoritme gestuurde aanbod (aanbevelingsalgoritmen) van rechtsextremistische content op sociale mediaplatforms kan leiden tot een rechtsextremistische 'sociale media fuik'? Is er hierbij een verschil tussen de diverse sociale mediaplatforms?

5. In hoeverre leiden aanbevelingsalgoritmen op sociale mediaplatforms op basis van gepersonaliseerd zoekgedrag door gebruikers (persona's / invoer van zoekwoorden) tot contentsuggesties door het platform voor extreme content?<sup>3</sup>
6. Zijn er gevallen bekend waarbij de rechtsextremistische 'sociale mediafuik' een bijdrage heeft geleverd aan radicalisering en uiteindelijke rekrutering van bezoekers van sociale mediaplatforms?
7. Op welke wijze droeg de 'sociale mediafuik' dan bij aan radicalisering en eventuele rekrutering?
8. Zijn er gevallen bekend waarbij naast online radicalisering sprake was van actieve offline rekrutering door rechtsextremisten/-terroristen?
9. Welke haalbare mogelijkheden zijn er (voor de overheid) om rechtsextremistische radicalisering via online sociale mediaplatforms tegen te gaan?

### 1.3. Opzet en uitgangspunten voor het onderzoek

#### Opzet van het onderzoek in drie fasen

De methodische uitvoering van dit onderzoek bestaat uit verschillende onderdelen. Het eerste onderdeel is een studie naar inzichten in de wetenschappelijke literatuur over ontwikkelingspaden van online (rechtsextremistische) radicalisering en de rol hierin van verschillende 'typen' sociale media. Het tweede onderdeel betreft interviews met experts over de werking en effecten van aanbevelingsalgoritmen, de invloed van hiervan ten opzichte van andere mechanismen in de online – en offline context, en haalbare mogelijkheden voor de overheid om online rechtsextremistische radicalisering tegen te gaan. Gesproken is met vijf experts op de domeinen sociale mediaplatforms, werking van aanbevelingsalgoritmen, mediatechnologie en samenleving, en ontwikkelingen in het rechts-extremisme. De conceptbevindingen van het rapport

zijn voorgelegd in een gesprek met twee juristen met kennis over de nieuwe Europese wetgeving omtrent sociale media platforms (Digital Service Act) en aan zeven experts (vier Nederlandse experts en drie experts uit andere Europese landen) in twee expertsessies. In deze expertsessies is gereflecteerd op de resultaten van de literatuurstudie en – vooral – het gezamenlijk met de experts doordenken van haalbare en (daarmee) praktisch inzetbare handelingsperspectieven voor de overheid. De achtergronden van de experts waren als volgt: experts op het gebied van de werking en effecten van aanbevelingsalgoritmen; een hoogleraar media en digitale samenleving; een vertegenwoordiger van het Commissariaat voor de Media; en experts op het gebied van sociale media en de maatschappelijke impact hiervan. Deze onderzoeksactiviteiten (literatuurstudie en expert interviews/sessies) zijn uitgevoerd door het Verwey-Jonker Instituut. Het derde onderdeel van de betref de uitvoering van online trials op YouTube en TikTok, deze is uitgevoerd door TextGain. Hierin is er onderzocht in hoeverre algoritmen op basis van gepersonaliseerd zoekgedrag door gebruikers leiden tot suggesties voor extreme(re) content. Wij voerden deze trials uit als een aanvulling op de literatuurstudie. Een methodologische kanttekening bij bestaande studies naar de werking van aanbevelingsalgoritmen is namelijk dat deze studies de gepersonaliseerde aard van de interactie tussen een gebruiker en content (zoals zoekgedrag door gebruikers) meestal niet meenemen. Als zodanig houden deze studies onvoldoende rekening met personalisatie die de kern vormt voor filterbubbels. In de trials is daarom beter rekening gehouden met de gepersonaliseerde aard van de interactie tussen gebruikers en content. Om die reden is in de trials specifiek gekeken naar aanbevelingen voor video's op YouTube en TikTok op basis van de invoer van zoekwoorden. Het gaat hierbij om zoekwoorden die niet noodzakelijk hoeven te resulteren in extreme content, maar hier wel een aanleiding voor kunnen geven. Het laatste onderdeel in dit onderzoek omvat het schrijven van deze eindrapportage.

<sup>3</sup> Deze onderzoeksvraag is in de loop van het onderzoek toegevoegd op basis van de bevindingen in de literatuurstudie omtrent het potentieel belang van gepersonaliseerd zoekgedrag bij aanbevelingen van rechtsextremistische content.

## Begeleidingscommissie

Het onderzoek is begeleid door een begeleidingscommissie. Deelnemers aan de begeleidingscommissie waren de volgende personen:

- Prof. dr. R.A. Rogers (Universiteit van Amsterdam).
- Prof. dr. J.W. van Prooijen (Universiteit van Maastricht: Vrije Universiteit Amsterdam; NCSR).
- Dr. G.M. van Buuren, (Universiteit van Leiden).
- K.A. van der Heijden, Msc (Ministerie J&V).
- Dr. H.C.J. van der Veen (WODC).

## Uitgangspunten voor het onderzoek

Voor de uitvoering van het onderzoek onderscheiden we drie belangrijke uitgangspunten, welke we baseren op de bevindingen van een eerder uitgevoerde literatuurstudie naar rechtsextremistische radicalisering in de online context (Peeters et al., 2022). Nadrukkelijk moet worden opgemerkt dat ontwikkelingspaden naar extremisme enerzijds een individueel proces zijn, maar anderzijds kan er zowel vanuit de online als offline omgeving van individuen van alles spelen dat invloed heeft op ontwikkelingspaden cq. op het tot stand komen van acties.

1. *Radicalisering heeft veelal dezelfde drijfveren.* Een belangrijke vraag die boven deze studie hangt, is in hoeverre het bij 'online' rechtsextremistische radicalisering gaat om een nieuw fenomeen. Hierop passen twee antwoorden. Enerzijds lijkt er weinig nieuws onder de zon. Er is namelijk een brede consensus over het feit dat de drijfveren van radicalisering, ongeacht het 'type' radicalisering – links, rechts, religieus, anti-overheid et cetera – veelal overeenkomen (Peeters et al., 2022). In de kern worden offline en online radicalisering vaak gedreven door een zoektocht naar zingeving en identiteit, psychosociale problematiek, een verlangen naar sociale binding en een ervaren gemarginaliseerde positie (Kruglanski et al. 2009; Sieckelinck & Gielen, 2020). Anderzijds houdt het onderzoek een kritisch oog en worden eventuele verschillen tussen het rechtsextremisme en andere vormen van extremisme onderzocht.

2. *Het online domein drukt een belangrijke stempel op de ontwikkelingspaden van rechtsextremistische radicalisering en de mechanismen die hieraan ten grondslag liggen.* Op de vraag in hoeverre 'online' rechtsextremistische radicalisering een nieuw fenomeen is, kan óók beredeneerd worden dat het wel degelijk om een nieuw fenomeen gaat (Peeters et al., 2022). Het feit dat het radicaliseringsproces zich ten minste voor een deel naar het online domein heeft verplaatst, heeft namelijk aanzienlijke gevolgen voor de ontwikkelingspaden van rechtsextremistische radicalisering en de mechanismen die hieraan ten grondslag liggen. In dit onderzoek hebben we nadrukkelijk aandacht voor de verschillende functies, effecten en gevolgen van het online domein en online content.
3. *Verwevenheid van de online en offline context.* Het is relevant te wijzen op het gegeven dat de online en offline contexten niet als twee onafhankelijke domeinen begrepen kunnen worden bij processen van radicalisering, en zo ook bij rechtsextremistische radicalisering (Gill et al., 2015; Mattheis, 2019; Whittaker & Herath, 2019; Whittaker, 2022). In de literatuur wordt regelmatig een onderscheid gemaakt tussen 'online' en 'offline' radicalisering. Alhoewel dit onderscheid helpend is in het al dan niet theoretisch beschrijven van deze twee domeinen, wijst de praktijk uit dat een dergelijk onderscheid vrijwel niet gemaakt kan worden (Ducol, 2015; Neumann, 2013; Gill et al., 2017). Bij radicalisering van personen en groepen spelen ontwikkelingen op maatschappelijk niveau, groepsniveau en persoonlijk niveau een rol (Feddes et al., 2015) en op al deze niveaus zijn de online en offline context met elkaar verweven. Het betekent dat de twee domeinen veeleer in het verlengde van elkaar liggen: zij vormen een 'continuüm'. Het bovenstaande betekent dat het denken in termen van deze 'online versus offline'-dichotomie problematisch kan zijn voor de beleidsmatige aanpak van rechtsextremistische radicalisering (Rowa, 2022). Door kenmerken, processen en activiteiten in de online en offline context als afzonderlijk te begrijpen, wordt de cruciale wisselwerking tussen deze contexten onderschat of zelfs over het hoofd gezien (Valentini et al., 2020). Het heeft daarmee weinig zin heeft om alleen te proberen online rechtsextremisme



te bestrijden zonder tegelijkertijd te proberen offline manifestaties ervan te begrijpen en aan te pakken.

4. *Ontwikkelingspaden van radicalisering beperken zich niet alleen tot het individu.* De context of omgeving van een persoon heeft namelijk een belangrijke invloed op het ontwikkelingspad van dit persoon richting radicaal of extremistisch gedachtegoed. Ook zijn er vaak andere factoren uit diens omgeving aanwezig die bijvoorbeeld bijdragen aan het tot stand komen van acties. Denk hierbij aan hulpbronnen, financiering en een infrastructuur om een bepaalde actie te faciliteren. Hoewel ontwikkelingspaden van rechtsextremistische radicalisering in deze studie met name vanuit het individuele oogpunt beschreven worden, zijn deze niet geïsoleerd van de omgeving waarin een persoon zich bevindt.

#### **Lacunes in bestaand onderzoek**

Voordat de bestaande literatuur over (online rechtsextremistische) radicalisering besproken wordt, is het goed om kort stil te staan bij enkele belangrijke tekortkomingen van het bestaande onderzoek. Onderzoek naar online radicalisering wordt doorgaans gekenmerkt door de volgende tekortkomingen: (a) de empirische fundering schiet vaak tekort; (b) het is veelal louter theoretisch en beschrijvend van aard; en (c) het is hoofdzakelijk gebaseerd op secundaire bronnen en casestudies welke de generaliseerbaarheid van resultaten bemoeilijkt (Archetti, 2012; Odağ et al., 2019; Von Behr et al., 2013; Winter et al., 2020; Mølmen & Ravndal, 2021; Hassan et al. 2018). Daar waar er wel onderzoek naar online radicalisering is verricht, valt het op dat de meest beschikbare studies zich voornamelijk richten op jihadistisch en/of islamitisch extremisme en dat er weinig tot geen aandacht is voor (of geen onderscheid gemaakt wordt in) gewelddadig rechtsextremisme en/of –terrorisme. Naast dat er weinig empirisch onderzoek naar online radicalisering voorhanden is, zijn er ook discussies omtrent definities in de literatuur en beleid. Zo is er bijvoorbeeld kritiek op het gebruik van de term ‘online radicalisering’ (Marwick et al. 2022) en ontbreekt er in de literatuur een eenduidige terminologie als het gaat om termen als radicalisering, (rechts)extremisme en (rechts)terrorisme. Dit

bemoeilijkt niet alleen de vergelijkbaarheid van diverse studies (Marwick et al. 2022; Hassan et al. 2018), maar ook de vindbaarheid ervan. Behalve onderzoek naar radicalisering is er ook empirisch onderzoek uitgevoerd naar inhoudelijk verwante thematiek. Deze studies gaan dan niet direct over rechtsextremisme, maar bijvoorbeeld over radicale anti-overheidsopvattingen of polarisatie. Ook voor die literatuur geldt dat de vindbaarheid via literatuuronderzoek wordt bemoeilijkt door uiteenlopende terminologie of definities.

#### **1.4. Leeswijzer**

Dit rapport bestaat uit de volgende hoofdstukken:

- *Hoofdstuk 2 Radicaliseringsprocessen* beschrijft (a) risicofactoren, risicoprofielen en beschermende factoren voor radicalisering, in het bijzonder voor rechtsextremistische radicalisering; (b) mechanismen van online radicaliseringsprocessen; en (c) mogelijke verschillen in deze mechanismen tussen rechtsextremistische en andere ‘typen’ radicalisering. Er wordt, tot slot, gekeken naar (d) de eventuele voorspellende waarde van online radicaliseringsprocessen naar het optreden van gewelddadig rechtsextremisme en/of rechtsterrorisme. Hoofdstuk 2 adresseert onderzoeksvragen 1 en 2.
- *Hoofdstuk 3 Rechtsextremistische content* geeft een overzicht van de rechtsextremistische online leefwereld door in te gaan op (a) definiëren en identificeren van rechtsextremistische content; (b) online strategieën die door rechtsextremisten gebruikt worden om hun doelen en publiek te bereiken; (c) cognitieve fasen (normalisering, acclimatisering en dehumanisering) die een gevolg kunnen zijn van blootstelling aan rechtsextremistische content; (d) mainstream en alternatieve sociale mediaplatforms waarop rechtsextremisten actief zijn; (e) gaming platforms; en (f) een bespreking van online subculturen die kunnen functioneren als ‘gateway’ naar rechtsextremistische content. Hoofdstuk 3 adresseert onderzoeksvraag 3.
- *Hoofdstuk 4 Aanbevelingsalgoritmen als sociale mediafuiik* geeft (a) een introductie op de werking van algoritmen door onder andere te kijken naar ‘typen’ algoritmen en de wijze waarop algoritmen op sociale

mediaplatforms werken; vervolgens wordt ingegaan (b) op de zorgen over sociale mediafauken, filterbubbels, echokamers en amplificatie door aanbevelingsalgoritmen; waarna (c) een overzicht wordt gegeven van de resultaten in empirische studies die deze concepten onderzochten. Hoofdstuk 4 adresseert onderzoeksvraag 4.

- *Hoofdstuk 5 Trials* beschrijft de opzet en uitkomsten van trials om – aanvullend op de reeds bekende resultaten van empirische studies – meer zicht te krijgen op de gepersonaliseerde werking van aanbevelingsalgoritmen op sociale mediaplatforms (YouTube en TikTok) op basis van de invoer van zoekwoorden. Hoofdstuk 5 adresseert onderzoeksvraag 5.
- *Hoofdstuk 6 Online en offline rekrutering* gaat nader in op (a) online invloeden bij bekende casussen van rechtsextremistische radicalisering; (b) de wisselwerking tussen de online context en de offline (fysieke) context bij radicaliseringsprocessen en rekrutering. Hoofdstuk 6 adresseert onderzoeksvragen 6, 7 en 8.
- *Hoofdstuk 7 Handelingsperspectieven* omvat suggesties voor mogelijke (haalbare) handelingsrichtingen voor de overheid om online rechtsextremistische radicalisering via sociale mediaplatforms tegen te gaan en/of te voorkomen. Om hiertoe te kunnen komen, worden allereerst nuances beschreven omtrent de rol van aanbevelingsalgoritmen op sociale mediaplatforms bij rechtsextremistische radicalisering. Hierna wordt ingegaan op wat sociale mediaplatforms zelf al doen aan moderatie van aanbevelingsalgoritmen en content. Vervolgens gaan we in op beleidsmogelijkheden van Nederland en de Europese Unie vanuit nieuwe aangescherpte Europese wetgeving. We sluiten het hoofdstuk af met mogelijke handelingsperspectieven. Daarbij hanteren we een breed perspectief. De gepresenteerde handelingsperspectieven richten niet alleen op online mogelijkheden voor contentmoderatie, maar ook op het offline versterken van de weerbaarheid van gebruikers tegen extremistische content. Hoofdstuk 7 adresseert onderzoeksvraag 9.

- In het afsluitende *Hoofdstuk 8 Conclusies* worden de belangrijkste bevindingen en conclusies van het onderzoek in samenhang met elkaar op een rij gezet en geduid.

Ieder hoofdstuk start met een blok waarin er kort wordt beschreven wat er in het hoofdstuk per paragraaf aan bod komt. Aan het eind van de hoofdstukken zijn er samenvattende blokken opgenomen waarin de belangrijkste punten uit het hoofdstuk besproken worden.

## 2 Radicaliseringsprocessen

Onderzoeksvragen die worden beantwoord in dit hoofdstuk:

1. Welk(e) mechanisme(n) zijn er in de wetenschappelijke literatuur bekend over radicaliseringsprocessen van individuen en groeperingen in de richting van rechtsextremistisch, gewelddadig gedachtegoed? Wat is hierbij de rol van online content? In hoeverre en op welke aspecten zijn er overeenkomsten óf verschillen in de mechanismen van rechtsextremistische radicalisering ten opzichte van andere vormen van radicalisering?
2. Wat is er bekend over de voorspellende waarde van online radicalisering naar het optreden van gewelddadig rechtsextremisme? In hoeverre is dit uniek voor rechtsextremisme?

In dit hoofdstuk komt het volgende aan bod:

- Introductie over radicaliseringsprocessen (Paragraaf 2.1).
- Risicofactoren en profielkenmerken voor radicalisering, in het bijzonder voor rechtsextremistische radicalisering (Paragraaf 2.2).
- Beschermende factoren en weerbaarheid tegen radicalisering, in het bijzonder in het online domein (Paragraaf 2.3).
- Mechanismen van online radicaliseringsprocessen (Paragraaf 2.4).
- Overeenkomsten en verschillen in radicaliseringsprocessen en -mechanismen tussen rechtsextremistische en andere 'typen' radicalisering (Paragraaf 2.5).
- Voorspelt online radicalisering gewelddadig (rechts)-extremisme? (Paragraaf 2.6).

### 2.1. Introductie

In de wetenschappelijke literatuur naar rechtsextremistische radicalisering is er veelvuldig aandacht voor het radicaliseringsproces dat personen doormaken met als uitkomst dat zij geweld verheerlijken, legitimeren of gebruiken. Niet zelden wordt daarin benadrukt dat dit zeer complexe, dynamische en geïndividualiseerde processen zijn (Kleuver, Dijk & Ramirez, 2017; Nickolson et al., 2021; Copeland & Marsden, 2020). Verschillende persoonlijke, sociale, culturele en politieke factoren kunnen namelijk veroorzaken dat een persoon in de richting van gewelddadig of terroristisch gedachtegoed beweegt. Ontwikkelingspaden van radicalisering – die onderscheiden worden op basis van de literatuur – reflecteren daarbij met name op risicofactoren en/of mechanismen die kunnen bijdragen aan de ontvankelijkheid van personen voor dit extremistische gedachtegoed (Copeland & Marsden, 2020). Echter, er is enige voorzichtigheid geboden in het identificeren van dergelijke ontwikkelingspaden. Het denken in termen van 'risicofactoren' of 'risicoprofielen' kan namelijk onterecht bijdragen aan het stigmatiseren van bepaalde groepen die mogelijk vatbaarder zijn voor radicalisering (Peeters et al., 2022). Belangrijk daarin is dat (a) het grootste gedeelte van de personen die 'voldoet' aan één of meerdere risicoprofielen niet radicaliseert; (b) risicofactoren al naar gelang de radicaliseringsfase waarin iemand zich bevindt verschillen; en (c) een dergelijke aanpak in onderzoek vrijwel nooit iets oplevert. Determinanten van radicalisering zijn doorgaans erg algemeen. Dit is begrijpelijk, omdat radicalisering al dan niet ten dele een gevolg is van algemene menselijke ontwikkelingspaden, bijvoorbeeld een zoektocht naar identiteit, betekenisgeving en vriendschappen. Het is, tot slot, relevant te noemen dat ontwikkelingspaden naar (rechts)radicalisering vaak beschreven worden vanuit het individu. De omgeving van dit individu heeft daarentegen een belangrijke invloed op het verloop van diens ontwikkelingspad, maar ook op het faciliteren van of bijdragen aan gewelddadige of niet-gewelddadige acties door dit individu.

## Radicaliseringsproces in fasen

Een mogelijkheid waarop radicalisering geïnterpreteerd kan worden, is vanuit een procesmatige benadering. Hierbinnen worden verschillende radicaliseringsfasen onderscheiden die een persoon of groep doormaakt tijdens het radicaliseringsproces. Gebaseerd op bestaande modellen van radicalisering (waaronder: Borum, 2011; Moghaddam, 2005; Precht, 2007; Sageman, 2004; Schmid, 2013; Silber & Bhatt, 2007; Wiktorowicz, 2004) identificeren Feddes et al. (2015) vier fasen in het radicaliseringsproces van een persoon:<sup>4</sup>

1. *Fase van gevoeligheid* voor radicalisering: deze fase gaat in feite vooraf aan de daadwerkelijke radicalisering. Gevoeligheid voor radicalisering kan allerlei oorzaken hebben. Een deel hiervan wordt in paragraaf 2.2 besproken.
2. *Fase van verkenning*: het actief op zoek gaan naar informatie gerelateerd aan de ideologie, en het ontwikkelen van een radicale interpretatie.
3. *Fase van participeren* in een (reële of virtuele) radicale groep: deze fase wordt gekenmerkt door het verder ontwikkelen van ideologisch gedachtegoed, indoctrinatie en mogelijk een toenemende bereidheid om geweld als een geoorloofd middel te zien. Als het laatste het geval is dan is in feite de afslag richting het extremisme genomen en zou een individu de stap kunnen zetten naar de volgende fase.
4. *Fase van actie*: in deze fase worden handelingen daadwerkelijk ondernomen, zoals het voorbereiden, plannen en uitvoeren van een geweldsdaad. Gezien de eerdere definities valt deze fase in feite niet meer binnen het proces van radicalisering, maar betreft het hier extremisme.

4 Belangrijk om hierbij te noemen, is dat Feddes et al. (2015) redeneren vanuit primair het radicaliserende individu. Sommige andere auteurs begrijpen het radicaliserende individu eerder in en vanuit zijn/haar omgeving en zien het isoleren hiervan als een cruciaal onderdeel van het radicaliseringsproces. Er ontstaat dan een wederzijdse afstand tussen een geradicaliseerd persoon en diens omgeving. Dit is een belangrijk element om radicalisering te begrijpen, maar ook om inzicht in geschikte handvaten te krijgen in relatie tot de verschillende fasen in een dergelijk proces.

Hoewel een procesmatige benadering van radicalisering een lineair verloop uit lijkt te stippelen, is dit niet noodzakelijk het geval: het proces resulteert immers niet altijd in het onderschrijven van extremistisch of gewelddadig gedachtegoed (Bartlett, Birdwell, & King, 2010; De Wolf & Doosje, 2010/2015; Feddes et al., 2013; King & Taylor, 2011; Kleuver, Van Dijk & Ramirez, 2017). Het is bovendien lang niet altijd een gestage ontwikkeling van gevoeligheid voor radicalisering tot een extremistische of terroristische actie. Het proces kan versnellen, vertragen of zelfs helemaal omkeren. Een (tijdelijke of permanente) beweging terug naar een eerdere fase is mogelijk. Ook is het mogelijk dat een fase niet of zeer snel wordt doorlopen, zoals bij de zogeheten 'flitsradicalisering', welke vaker plaatsvindt in de online context. Om deze reden is er niet één ontwikkelingspad van radicalisering te onderscheiden, hetgeen reflecteert op de keuze om te spreken van verschillende ontwikkelingspaden (Kleuver, Van Dijk & Ramirez, 2017; Marwick et al., 2022). Desalniettemin blijkt uit onderzoek dat het risico op radicalisering wel toeneemt wanneer er sprake is van een combinatie en/of cumulatie van risicofactoren (Spee & Reitsma, 2015; Nickolson et al., 2021).

## Rechtsextremistische radicaliseringsprocessen

Om meer inzicht te verkrijgen in het verloop van (online) rechtsextremistische radicalisering, kan er gekeken worden naar drie aspecten die ontwikkelingspaden van radicalisering inkleuren: (1) risicofactoren die een persoon, ongeacht het 'type' radicalisering, ontvankelijk maken voor extremistisch gedachtegoed of handelen; (2) profielkenmerken van personen die herhaaldelijk terugkomen bij rechtsextremistische radicalisering; en (3) mechanismen die online plaatsvinden en kunnen bijdragen aan rechtsextremistische radicalisering. Omwille van de overzichtelijkheid van deze literatuurstudie worden risicofactoren en profielkenmerken in Paragraaf 2.2 apart beschreven van de mechanismen van online radicalisering in Paragraaf 2.3. Echter, in de praktijk zijn deze risicofactoren, profielkenmerken en radicaliseringsmechanismen met elkaar verweven.

## 2.2. Risicofactoren en profielkenmerken

De wetenschappelijke literatuur toont een grote veelzijdigheid in het identificeren van risicofactoren die kunnen bijdragen aan het ontstaan van radicaliseringsprocessen (Schröder et al., 2022; Copeland & Marsden, 2020; Schmid, 2013; Nickolson et al., 2021; Van den Berg, 2022). Risicofactoren blijken hierin vaker overeen te komen dan te verschillen tussen verschillende 'typen' radicalisering – zoals links, rechts, religieus, anti-overheid et cetera (Alonso et al., 2008; Sikkens, et al., 2017; Schröder et al., 2022). Paragraaf 2.2 focust daarom allereerst op risicofactoren waarvan bekend is dat zij kunnen bijdragen aan de ontvankelijkheid van personen voor extremistisch gedachtegoed, waaronder rechtsextremistische radicalisering. Om een volledig beeld te geven van de wijze waarop rechtsextremistische radicalisering tot stand komt, wordt daaropvolgend beschreven hoe risicofactoren terugkomen in profielkenmerken van rechtsextremistische radicalisering. De paragraaf wordt afgesloten met een korte beschrijving van beschermende factoren die op basis van literatuur geïdentificeerd zijn.

### Risicofactoren

Er zijn verschillende manieren waarop risicofactoren voor radicalisering ingedeeld en/of geordend kunnen worden – onder andere met behulp van modellen, kaders, perspectieven en niveaus (Van den Berg, 2022). Een veelvoorkomende manier waarop risicofactoren geordend worden, is door te kijken naar de werkzaamheid ervan in relatie tot het radicaliseringsproces. Deze benadering onderscheidt pushfactoren, pullfactoren en triggerfactoren (Van den Berg, 2022; Feddes et al., 2015; Sikkens et al., 2017; Korteweg et al., 2010; Schmid, 2013; Bjørgo en Horgan, 2009). Radicalisering van personen is een complexe mix van pushfactoren en pullfactoren die op elkaar inwerken en beïnvloed kunnen worden door triggerfactoren – daar waar laatstgenoemde zowel een keerpunt als katalysator kan zijn in het radicaliseringsproces. Pushfactoren behelzen onderliggende triggers en drijfveren die personen richting radicale groepen 'duwen' (Sikkens et al., 2017; Schmid, 2013). Voorbeelden hiervan zijn

persoonlijke uitsluiting, relatieve deprivatie of marginalisatie, sociale isolatie, het ervaren van discriminatie, problematische thuissituatie, school- of werkgerelateerde problemen, de zoektocht naar identiteit en/of betekenisgeving, het ervaren van onrecht in de wereld en waargenomen onrechtvaardigheid (Sikkens et al., 2017). Dergelijke factoren kunnen bijdragen aan of deel uitmaken van radicaliseringsmechanismen waarbinnen een persoon zich beweegt richting extremistisch gedachtegoed of handelen. Pullfactoren verwijzen naar factoren die personen aantrekken tot de positieve kenmerken en voordelen van een radicale of extremistische groep (Schmid, 2013; Sikkens et al., 2017). De literatuur identificeert onder meer de volgende pullfactoren: de aanwezigheid van radicale groepen of netwerken, de aanwezigheid van radicale groepen op internet, familieleden of vrienden die reeds bij een radicale groep zijn aangesloten en beloningen als status, spanning en sociale binding die tegenover het lidmaatschap van een groep staan (Sikkens et al., 2017). In hoeverre push- en pullfactoren aanwezig zijn bij een persoon en de precieze wijze waarop zij op elkaar inwerken, is een geïndividualiseerd gegeven dat maakt dat ontwikkelingspaden van radicalisering zeer verschillend kunnen verlopen – en niet noodzakelijk resulteren in (gewelddadige) radicalisering.

### Profielkenmerken

Zoals hierboven is aangegeven, zijn ontwikkelingspaden van radicalisering complexe, dynamische en geïndividualiseerde processen (Kleuver, Dijk & Ramirez, 2017; Nickolson et al., 2021; Copeland & Marsden, 2020). Een gevolg daarvan is dat er geen algemeen risicoprofiel opgesteld kan worden van 'de (rechts)extremist'. Het is daarentegen wel mogelijk om een aantal individuele profielkenmerken te identificeren die vaker dan gemiddeld (maar, opnieuw niet noodzakelijk) voorkomen bij rechtsextremistische radicalisering en die vaak aansluiten of samengaan met de hierboven beschreven risicofactoren. Deze paragraaf geeft daarom aandacht aan profielkenmerken waarvan in de literatuur verondersteld wordt dat zij een voedingsbodem kunnen vormen voor ontwikkelingspaden van rechtsextremistische radicalisering. Het is goed om daarbij het volgende hierbij te benadrukken: (a) profielkenmerken mogen

individueel niet worden opgevat als voorspellers voor radicalisering – dit is altijd in cumulatie of interactie met andere kenmerken; (b) persoonsgebonden profielkenmerken zijn niet geïsoleerd van de sociale, culturele en maatschappelijke context waarin een persoon zich beweegt; en (c) het afzonderlijk identificeren van profielkenmerken is enigszins kunstmatig, omdat de verschillende radicaliseringsmechanismen die schuilgaan achter deze kenmerken vaak overeenkomen. In deze literatuurstudie identificeren wij de volgende profielkenmerken van rechtsextremistische radicalisering:

**Geslacht.** Ontvankelijkheid voor (rechts)extremistische narratieven en aansluiting bij extremistische groepen komt relatief vaker voor bij mannen, maar niet uitsluitend (Nickolson et al., 2021; Kimmel, 2018; Dier & Baldwin, 2022; Decker et al., 2020; Küpper & Zick, 2011; Küpper et al., 2021; Bergmann et al., 2019). Uit interviews met meer dan honderd (ex)-extremisten naar hun ervaring met mannelijkheid binnen het rechtsextremisme kwam het narratief naar voren van wat Kimmel (2018) 'gekrenkte rechten' noemt. Dit omvat een gendergebonden gevoel van "recht dat gedwarsboomd wordt door grote economische en politieke verschuivingen, met als gevolg dat ambities verstrikt raken en masculiniteit verloren gaat" (Kimmel, 2018). Inmiddels komt dit op als wervingsgrond bij extreemrechts, wat sterk beïnvloed wordt door de Amerikaanse altright-beweging, incel- en anti-woke denken (interview Wagenaar, 2022). Deze ervaren politiek-economische deprivatie gaat doorgaans gepaard met een persoonlijker gevoel van deprivatie: sociale isolatie of gepest worden, waardoor de behoefte ontstaat voor steun van 'iets' dat groter is dan henzelf (Kimmel, 2018). *"Joining right-wing groups lent a gendered coherence to their sense of emasculation and frustration. Their manhood had been taken from them by unseen conspiratorial forces, and their recruitment was seen as a way to reclaim their manhood and to restore that sense of entitlement"* (Kimmel, 2018). Nadere informatie hierover staat in Paragraaf 3.7 onder het kopje 'Manosfeer'. Ondanks dit profielkenmerk mag de rol van vrouwen in het rechtsextremisme niet over het hoofd gezien worden – niet in aantallen en niet in invloed (Peeters et al., 2022; Daymon & Margolin, 2022; Leidig, 26 oktober 2021; Saarijärvi &

Siafaka, 2021; McClintlock, 2022). Dat vrouwen voor de buitenwereld een soms minder zichtbare rol innemen, kan namelijk een gevolg zijn van de traditionele rolverdelingen binnen het rechtsextremisme. In dat geval wordt van vrouwen verwacht dat zij meer opereren vanuit een achtergrondpositie (Brown, 2019; Daymon & Margolin, 2022).

**Leeftijd.** Radicalisering, waaronder rechtsradicalisering, komt gemiddeld vaker voor bij jongere personen (AIVD, 2023a; Copeland & Marsden, 2020; Nickolson et al., 2021; Van den Berg, 2022; Marwick et al., 2022; Kruglanski, Webber & Koehler, 2019). Het idee dat hieraan ten grondslag ligt vanuit het ontwikkelingsperspectief, is dat personen in de adolescentiefase meer beïnvloedbaar, onzeker en daarmee kwetsbaarder zijn (Schröder et al., 2022). Doorgaans zijn jongeren (12-18 jaar) en jongvolwassenen (19-25 jaar) zoekende naar hun eigen identiteit en daarin gevoeliger voor groepsdruk, waardoor zij relatief vatbaarder zijn voor radicale gedachten en identificatiefiguren (Van den Berg, 2022; Feddes et al., 2015; Schröder et al., 2022; Greve, 2007). Daarnaast zoeken jongeren vaker naar avontuur, sensatie en provocatie (Benslama, 2017; Emmelkamp et al., 2020; McCauley & Moskalenko, 2011; Venhaus, 2010) en vindt politieke socialisatie meestal plaats in de adolescentiefase (Beelmann, 2020; Niemi & Hepburn, 2010; Torney-Purta, 2004; Watts, 1999; Schröder et al., 2022). Dit kan bijdragen aan de vatbaarheid voor radicale ideeën. Enige nuance binnen dit profielkenmerk is op zijn plaats: hoewel de relatief lage leeftijd van radicaliserende personen vaak terugkomt, impliceert dit niet dat rechtsextremistische radicalisering uitsluitend een 'jeugdprobleem' is. Neonazi-organisaties, anti-islambewegingen en burgerwachten laten recent een significante toename zien in de betrokkenheid van volwassenen (Politie's Sikkerhetstjeneste, 2019; Pauwels, 2019).

**Persoonlijkheid.** Op basis van de literatuur kunnen enkele gevolgtrekkingen gedaan worden voor persoonlijkheidskenmerken die relatief vaak voorkomen bij rechtsextremistische radicalisering. Ten aanzien van psychologische kenmerken blijkt dat personen die rechtsextremistisch gedachtengoed onderschrijven, vaker een 'vereenvoudigde' denkwijze hebben en een relatief groot verlangen naar "gehoorzaamheid aan autoriteit, orde, zuiverheid, vertrouwdheid, structuur en een rigide wereldbeeldmentaliteit" (Obaidi, 2020; Alizadeh et al., 2019). Bij rechtsextremisten wordt tevens vaak een rigide denkwijze – ook wel *cognitieve inflexibiliteit* – geïdentificeerd. Hoewel dit strikt genomen geen persoonlijkheidskenmerk is, is het wel een kenmerk waarop individuen kunnen verschillen. Deze inflexibiliteit kenmerkt zich door geslotenheid, simplistisch en zwart-wit percepties van de samenleving, wat bijdraagt aan het gemakkelijker omarmen van rechtse attitudes en extreme ideologieën (Obaidi, 2020; Gambetta & Hertog, 2017; Jost et al., 2003; Zmigrod, Rentfrow & Robbins, 2019; Van Hiel et al., 2016).

**Psychopathologie.** Vatbaarheid voor (rechts)extremistische narratieven en hiernaar handelen kan vergroot worden door psychische problematiek (Weenink, 2015; Krasenberg & Wouterse, 2019; Van den Berg, 2022). Onder meer jongeren met een autisme-spectrum stoornis (ASS) worden herhaaldelijk genoemd in onderzoeken als het gaat over ontvankelijkheid voor (met name online) radicalisering; bijvoorbeeld doordat personen met ASS relatief vaker eenzaam en/of sociaal geïsoleerd zouden zijn en soms een rigide denkwijze hebben (Peeters et al., 2022; Momas, 2021; Woodbury-Smith et al., 2022). Laatstgenoemde ligt in het verlengde van de hierboven genoemde *cognitieve inflexibiliteit*. Ook worden personen met een licht verstandelijke beperking als kwetsbare groep geïdentificeerd, omdat zij relatief gemakkelijk beïnvloedbaar zijn (Platform JEP, 2019). Daarnaast tonen *lone actors* vaak een psychopathologisch ziektebeeld als depressies, angst- en persoonlijkheidsstoornissen (Spaaij, 2012; Corner, Gill, & Mason, 2016). Een nuancering die hier gemaakt moet worden, is dat ziektebeelden (net zoals de andere genoemde profielkenmerken in deze paragraaf) op zichzelf geen voorspellende waarde hoeven te hebben

voor (rechts)radicalisering van individuen, maar op geaggregeerd niveau dikwijls wel samenhangen met risicofactoren of persoonlijkheidskenmerken die maken dat dergelijke ziektebeelden vaker zichtbaar zijn bij ontwikkelingspaden van radicalisering.

**Deprivatie.** Een regelmatig in de literatuur terugkerend profielkenmerk voor rechtsextremistische radicalisering is (relatieve) deprivatie (Schröder et al., 2022; Crenshaw, 2018; Kaldor, 2021; Doosje et al., 2012). Deprivatie is daarin niet uitsluitend een objectieve 'achterstand' ten opzichte van anderen, maar omvat ook subjectieve gevoelens van achterstelling, morele verontwaardiging en woede (Schröder et al., 2022; Böckler & Zick, 2015; Kessler et al., 1999). Er wordt verondersteld dat relatieve deprivatie kan bijdragen aan de verspreiding van rechtsextremistische opvattingen, omdat *outgroups* de schuld kunnen krijgen van gevoelens van deprivatie en ervaren *ingroup*-superioriteit verband lijkt te houden met rechtsextremistisch geweld (Doosje et al., 2012). In een Nederlands onderzoek door Doosje et al. (2012) kwam onder meer naar voren dat ervaren onrechtvaardigheid, ervaren groepsdreiging en relatieve deprivatie determinanten kunnen zijn voor een radicaalrechts *belief system*. Gekeken naar de verschillende hedendaagse narratieven van het rechtsextremisme (zie daarvoor ook: Paragraaf 3.2) blijkt bovendien dat ervaren of gerechtvaardigde grieven, de perceptie van crisis en een strijd om verlossing belangrijke sleutelementen zijn (Sterkenburg, 2019). Tot slot, sociale deprivatie kan in het bijzonder bij jonge personen extra impact hebben, omdat dit hun maatschappelijke participatie kan belemmeren en hen daarmee in kwetsbaardere positie zet (Koehler, 2014).

**Levensgebeurtenissen en familieomstandigheden.** Drastische en kritische 'life events' kunnen een grote impact hebben op het leven van personen (Schröder et al., 2022). Als gevolg hiervan lijken personen ontvankelijker voor radicale boodschappen – zoals hierboven is beschreven bij *Risicofactoren* (Copeland & Marsden, 2020). Kritische familieomstandigheden lijken daarin regelmatig terug te komen bij rechtsextremistische radicalisering (Copeland & Marsden, 2020; Alonso et al., 2008). Voorbeelden hiervan zijn emotionele en/of fysieke mishandeling, schadelijke of afwezige familieleden door overlijden of echtscheiding, jeugdtrauma's en een gebrek aan ouderlijk toezicht dat 'ruimte' creëert voor extremisme. Een ander profielkenmerk dat regelmatig (maar, zeker niet altijd) terugkomt bij (ex)-rechtsextremisten is de aanwezigheid van familieomstandigheden die worden gekenmerkt door racisme, vooroordelen en extreemrechtse denkbeelden (Copeland & Marsden, 2020; Gielen, 2008; Simi, Windisch & Sporer, 2016; Peeters et al., 2022). De invloedrijke rol van ouders wordt daarin breed onderschreven. *"The exact role of parental attitudes in pathways to violence is still not clear, however right-wing extremists commonly state that their families were important in shaping their underlying beliefs and increased their willingness to accept right-wing ideologies"* (Copeland & Marsden, 2020).

**Alternatieve verklaring: isolatie of marginalisatie is een gevolg en niet de oorzaak van radicalisering.** De bovenstaande profielkenmerken komen relatief vaak (en vaak in combinaties) naar voren bij ontwikkelingspaden van rechts-extremistische radicalisering, maar bieden geen volledige verklaring. Tevens bestaan er alternatieve verklaringen voor de relatie tussen profielkenmerken en extremisme. Zo stellen Klandermans en Mayer (2006) dat extremisten veelal volstrekt 'normale' mensen zijn, die sociaal geïntegreerd zijn in de samenleving en verbonden zijn met mainstream groepen en ideeën. Zij concluderen dat isolatie of marginalisatie veeleer een *gevolg* is van hun activisme, en niet zozeer de *oorzaak*. Het bestaan van correlatieve verbanden tussen risicofactoren of risicoprofielen met radicaliseringsprocessen wil dus niet zeggen dat er altijd sprake is van oorzakelijke verbanden. Het is daarom belangrijk in het achterhoofd te houden dat vermeende oorzaken voor ontvankelijkheid voor radicalisering niet altijd radicalisering veroorzaken, maar soms een gevolg hiervan zijn.



### 2.3. Beschermende factoren en weerbaarheid

Er zijn weinig wetenschappelijke studies voorhanden die zich specifiek richten op het identificeren van beschermende factoren tegen gewelddadige rechtsextremistische radicalisering (Marsden & Copeland, 2020; Peeters et al., 2022). De literatuur die wel beschikbaar is, gaat meer in het algemeen over beschermende factoren ten aanzien van gewelddadig extremisme. In dit verband wordt er vaak gesproken over 'het vergroten van de weerbaarheid' tegen extremisme (Nickolson et al., 2021) – waarbij 'weerbaarheid' wordt gedefinieerd als het 'terugveren' na een tegenslag (Sieckelink & Gielen, 2020). De inzet op beschermende factoren en het bevorderen van weerbaarheid op zowel individueel, groeps- als maatschappelijk niveau zijn cruciale elementen in de aanpak en preventie van gewelddadig extremisme.

#### Beschermende factoren en 'weerbaarheidsfactoren'

Lösel, King, Bender en Jugl (2018) hebben een review uitgevoerd van tweeduizend documenten over radicalisering en gewelddadig extremisme – waaronder, maar niet exclusief, rechtsextremisme. Op basis van studies over beschermende factoren en het effect ervan, onderscheiden zij vijf categorieën waarop beschermende factoren zich kunnen manifesteren: individu, familie, school, *peers* en gemeenschap (Lösel et al., 2018; Lösel & Farrington, 2012). Belangrijk om te noemen, zo stellen de auteurs, is dat beschermende factoren van gewelddadig extremisme opvallende parallellen vertonen met die van geweld in zijn algemeenheid. Een voorbeeld hiervan is *zelfcontrole* (categorie: individu) (Pauwels & Svensson, 2017). Deze factor wordt ook in latere studies geïdentificeerd als beschermende factor tegen radicalisering (waaronder: Wolfowicz et al., 2019; Rottweiler & Gill, 2022). Zo zou zelfcontrole bijvoorbeeld de negatieve effecten van groepsdeprivatie als risicofactor voor gewelddadig extremisme reduceren (Rottweiler & Gill, 2022). Andere voorbeelden van beschermende factoren (Lösel et al., 2018) – waaraan ook in bredere geweldsstudies een beschermend effect wordt toegeschreven, zijn: steun door familie; een goede band met de school (Baier et al., 2016); goede schoolprestaties (Baier et al., 2016; Boehnke et al., 1998); contact met non-deviante leeftijdsgenoten en

bindingen met de conventionele samenleving; peer-groep contact met 'buitenlanders' (foreigners, zie Fuchs, 2003); acceptatie van de legitimiteit van de politie (Pauwels & De Waele, 2014); en bereidheid zich aan de wet te houden (Baier et al., 2016). Wat betreft bindingen met een sociale omgeving: deze lijken dus niet alleen cruciaal om jongeren van criminaliteit te weerhouden, maar zij zouden er ook voor kunnen zorgen dat jongeren wegblijven van het extremistische pad waarin geweld een rol kan gaan spelen (Lösel et al., 2018; Wolfowicz et al., 2019; Rottweiler & Gill, 2022; Peeters et al., 2022). Ten aanzien van laatstgenoemde twee beschermende factoren plaatsen Lösel et al. (2018) wel een kanttekening in het geval van extreemrechts. Binnen extreemrechts wordt vaak veel waarde gehecht aan orde en gezag. Tegelijkertijd kan er onvrede zijn over de staat van orde en gezag binnen een bepaald gebied, hetgeen zich kan vertalen in een gevoel dat er (gewelddadige) acties ondernomen moeten worden die het staatsgezag ondermijnen. Op deze manier kan er een innerlijk conflict van waarden ontstaan.

Een andere relevante overzichtsstudie is die van Wolfowicz, Litmanovich, Weisburd en Hasisi (2019). Zij concentreerden zich op het identificeren van risicofactoren en/of beschermende factoren in kwantitatieve radicaliseringsstudies en vonden 57 'geschikte' studies. Hierbij maakten ook deze auteurs geen specifiek onderscheid naar 'typen' radicalisering. In plaats daarvan lieten de auteurs zich leiden door het onderscheid tussen: (a) rechtvaardiging en ondersteuning van radicaal gedrag – attitudes; (b) bereidheid tot radicaal gedrag – intenties; en (c) betrokkenheid bij radicaal gedrag, waaronder terrorisme – gedragingen. Ten aanzien van (a) radicale attitudes bleken de volgende factoren een beschermend effect te hebben: een goede binding met de school; goede schoolprestaties; ouderlijke betrokkenheid; institutioneel vertrouwen; en gehoorzaamheid aan de wet. Ten aanzien van (b) radicale intenties werden er (kleine) beschermende effecten geïdentificeerd voor educatie en vriendschappen met '*out-groups*'. De meest relevante beschermende factoren ten aanzien van (c) radicale gedragingen, waren een goede binding met school; ervaren legitimiteit van en gehoorzaamheid aan de wet (Wolfowicz et al., 2019).

Tot slot identificeren Van Brunt et al. (2017) op basis van een literatuurreview tien beschermende factoren tegen extremisme – zonder daarin een onderscheid te maken voor het ‘type’ extremisme. De door hen geïdentificeerde beschermende factoren zijn als volgt: de inzet op sociale binding; pluralistische inclusiviteit; niet-gewelddadige uitlatingen (creëren van ruimte om zich te uiten in gedachten en emoties); sociale veiligheid; emotionele stabiliteit; professionele of academische betrokkenheid; ‘*global competence*’ (acceptatie en tolerantie van diverse zienswijzen); empathie; gevoel van identiteit en consequenties van handelen doorzien (gevoel van morele betrokkenheid en bewustzijn van hoe bepaalde keuzes resulteren in positieve en negatieve gedragingen).

Nickolson et al. (2021) onderscheiden daarnaast drie soorten ‘weerbaarheidsfactoren’ die de weerbaarheid van personen tegen extremisme (waaronder, maar niet exclusief rechtsextremisme) vergroten. Deze zijn als volgt: (a) cognitieve factoren – dit omvat de manier waarop mensen *denken*; (b) emotionele factoren – dit omvat de manier waarop mensen zich *voelen*; en (c) gedragsfactoren – dit omvat de manier waarop mensen zich *gedragen*. Het meest relevant voor onze literatuurstudie zijn de cognitieve factoren. Nickolson et al. (2021) identificeren op basis van een literatuurreview meerdere componenten die de cognitieve weerbaarheid tegen extremisme van personen kunnen vergroten: meer kennis; kritisch denken; gevoel van identiteit; het opstellen van morele regels; en het *willen* voorkomen van negatieve gevolgen door gedrag. Gekeken naar bijvoorbeeld *meer kennis*, blijkt vooral specifieke kennis zoals ‘kennis van democratie’ bij te dragen aan de weerbaarheid tegen radicale ideeën (Nickolson et al., 2021; Feddes et al., 2019). Meer kennis over democratie is gerelateerd aan meer vertrouwen in de overheid, een positieve houding ten aanzien van democratie en minder steun voor ideologisch-gemotiveerd geweld (Feddes et al., 2019). Gekeken naar *kritisch denken*, een in veel studies terugkerende factor, zijn er twee manieren waarop dit doorgaans gedefinieerd wordt: (1) hoe goed een persoon in staat is om *tegenargumenten* te bedenken tegen radicale boodschappen en (2) in hoeverre een persoon een onderwerp vanuit *verschillende perspectieven* kan bezien – ook wel: *cognitieve flexibiliteit* genoemd. In een

studie van Braddock (2019) kregen deelnemers van een online survey links- en rechtsextremistische propaganda voorgeschoteld. Een deel van de deelnemers werd van te voren gewaarschuwd door een bericht dat hen er op wees dat zij mogelijk propaganda van een politiek extremistische groep te lezen zouden krijgen (met een kans van één op acht). Deelnemers die van te voren gewaarschuwd waren, bleken kritischer op de extremistische propaganda en waren minder geneigd de extremistische groepen die achter de propaganda zaten te steunen – mede doordat zij in staat waren om meer tegenargumenten te bedenken (Braddock, 2019). Dit lijkt aldus het beschermende effect van kritisch denken te bevestigen (Nickolson et al., 2021).

### **Beschermende factoren in de online context**

Nog minder beschikbaar dan studies naar beschermende factoren tegen *rechtsextremistische* radicalisering zijn studies die zich richten op beschermende factoren tegen rechtsextremistische radicalisering in *de online context* (Peeters et al., 2022). De onderstaande bespreking beperkt zich tot de meest relevante uitkomsten van studies. Uit een Duits onderzoek van Nienierza et al. (2019) blijkt bijvoorbeeld dat jongeren online relatief vaak in contact komen met extremistische boodschappen. Daarbij gaf 40 procent van de jongeren aan dat zij wel eens in aanraking kwamen met rechtsextremistische content. De jongeren bij wie dit het geval was, bleken relatief vaak minder vertrouwen te hebben in instituties. Jongeren met meer vertrouwen in instituties kwamen hier minder vaak mee in aanraking. Mogelijk heeft vertrouwen in instituties daarom een beschermend effect op de blootstelling aan online extremistische berichtgeving.

Specifiek gekeken naar het eerdergenoemde beschermende effect van *kritisch denken* lijkt het belang van dit vermogen toe te nemen zodra radicalisering zich verplaatst van het offline naar het online domein (Peeters et al., 2022). Rechtsextremistische narratieven kenmerken zich namelijk door hun dichotome karakter, iets wat zich laat aftekenen langs de lijnen van het 'eigene' en het 'vreemde', de 'wij' en de 'zij', en het 'goede' en het 'slechte' – en wat regelmatig terugkomt in online rechtsextremistische boodschappen. Het vermogen om kritisch te denken kan helpen om hier doorheen te prikken (Peeters et al., 2022). In verschillende studies wordt daarbij onderscheid gemaakt tussen complex denken en kritisch denken. 'Complex' (Liht & Savage, 2013; Stephens et al., 2021) of 'flexibel' (Rottweiler & Gill, 2022) denken is gericht op het ontmaskeren van het zwart-wit denken dat ten grondslag ligt aan rechtsextremistische narratieven. Kritisch denken laat zich afmeten aan de mate waarin jongeren in staat zijn om informatie, berichten, ideeën en inhoud te beoordelen en de bron en inhoud van ideeën in twijfel te trekken en te analyseren (Stephens et al., 2021). Dit sluit aan bij het eerdergenoemde in staat zijn om *tegenargumenten* te formuleren en een onderwerp vanuit *verschillende perspectieven* te kunnen benaderen (Nickolson et al., 2021). Het veronderstelt om deze reden ook een zekere mate van onafhankelijkheid in het denken.

Naast kritisch kunnen denken over dichotome berichtgeving is het relevant om erop te wijzen dat het internet een omgeving is met relatief weinig pedagogisch toezicht; en waarbinnen nepnieuws en desinformatie – zeker in rechtsextremistische kringen – een stevige voet aan de grond heeft (Rogers & Niederer, 2020; Peeters et al., 2022). Het eerder genoemde onderzoek van Nienierza et al. (2019) laat zien dat jongeren moeite hebben met het herkennen van het problematische karakter van extremistische online boodschappen. Onderzoek dat is uitgevoerd in Nederland maakt duidelijk dat jongeren het lastig vinden om in te schatten of informatie betrouwbaar is en, daarmee, nepnieuws te herkennen (Kennisnet, 2017; Van Welzen et al., 2017). Jongeren blijken minder digitaal vaardig en mediawijs dan zij zelf dachten. Ze lijken 'confident' maar niet 'competent' (Peeters et al., 2022).

## 2.4. Radicaliseringsmechanismen

Risicofactoren en profielkenmerken van rechtsextremistische radicalisering die in de voorgaande paragrafen zijn benoemd, maken deel uit van mechanismen die eraan kunnen bijdragen dat een persoon zich beweegt richting gewelddadig of terroristisch gedachtengoed. In de literatuur wordt verondersteld dat het legitimeren van geweld als politiek middel of doel een proces is van verschillende psychologische, sociale, culturele en politieke elementen (Marwick et al., 2022). Dit sluit aan bij het gegeven dat er uiteenlopende ontwikkelingspaden zijn die erin resulteren dat een persoon bereid is extremistisch geweld te gebruiken (Kleuver, Van Dijk & Ramirez, 2017; Marwick et al., 2022; Borum, 2011). In de wetenschappelijke literatuur zijn er verschillende modellen bekend die dergelijke ontwikkelingspaden van radicalisering pogen te beschrijven. Voorbeelden daarvan zijn het *staircase*-model van Moghaddam (2005) dat bestaat uit vijf treden die elk een toenemende bereidheid richting politiek geweld presenteren en het model van McCauley en Moskalenko (2008) waarin radicalisering als *piramide* wordt verondersteld. Dergelijke modellen begrijpen radicalisering vanuit een procesmatige benadering en beschrijven dit proces in zekere zin lineair, terwijl andere auteurs dit liever vermijden (Marwick et al., 2022; Meleagrou-Hitchens & Kaderbhai, 2017).

### Radicaliseringsmechanismen in het online domein

In het kader van de focus van dit onderzoek, te weten *online* radicalisering, heeft deze paragraaf gerichte aandacht voor de mechanismen van online radicaliseringsprocessen. Er bestaat echter weinig onderzoek naar mechanismen van radicalisering in de online context, welke kunnen leiden tot extremistisch geweld, laat staan studies naar mechanismen van online rechtsradicalisering. Wel is recent de volgende relevante publicatie verschenen: "*Mechanisms of online radicalisation: how the internet affects the radicalisation of extreme-right lone actor terrorists*" (Mølmen en Ravndal, 2021). In de woorden van de auteurs:

*“In the absence of an established theoretical model, this article identifies six mechanisms seen as particularly relevant for explaining online radicalization. Having first reviewed a larger set of relevant lone actor terrorists, the study traces these mechanisms in three selected cases where the internet was reportedly used extensively during radicalization. The findings show that the internet primarily facilitated radicalization through information provision, as well as amplifying group polarization and legitimizing extreme ideology and violence through echoing. In all cases, radicalization was also affected considerably by offline push-factors that through their presence made extreme online messages more impactful.”*

Mølmen en Ravndal (2021) identificeren in hun studie zes mechanismen die zich ofwel volledig online, ofwel binnen interacties van online en offline contexten manifesteren. Het gaat hierbij om de volgende mechanismen: (1) compensatie; (2) isolatie; (3) echoing; (4) facilitering; (5) acceleratie; (6) ‘action triggering’. Deze mechanismen kunnen plaatsvinden binnen drie online radicaliseringsfasen. De eerste fase betreft de *pre-radicaliseringsfase* waarbinnen (offline) behoeften en kwetsbaarheden die personen ontvankelijk maken voor alternatieve wereldbeelden een grote rol spelen. In de tweede fase komen personen in aanraking met radicale en extreme online content en worden daardoor beïnvloed. De derde fase omvat de *operationele fase* waarin personen extreme attitudes omzetten in (gewelddadige) actie. De auteurs merken daarentegen op dat deze mechanismen niet allemaal een rol hoeven te spelen binnen een online radicaliseringsproces. Bovendien hoeft het radicaliseringsproces niet altijd in de door Mølmen en Ravndal (2021) beschreven volgorde te verlopen; dit kan per individu verschillen. De zes mechanismen manifesteren zich per fase als volgt:

**Pre-radicaliseringsfase.** Het eerste mechanisme is *compensatie*, waarbij individuen eventuele ‘gebreken’ of kwetsbaarheden die zij ervaren in de offline context online willen compenseren, bijvoorbeeld als gevolg van psychische of sociale problematieken. Personen zoeken online naar een alternatieve sociale

omgeving waarin de drempel voor inclusie in vergelijking met de fysieke wereld lager is. Een volgend mechanisme, *isolatie*, is dan dat deze personen die offline geen (sociale) aansluiting vinden online wel een vorm van sociale verbondenheid vinden en hierin gaan participeren, met als gevolg dat zij sociale vervreemding kunnen ervaren met de fysieke wereld (Mølmen & Ravndal, 2021). Van der Valk en Wagenaar (2010) noemen de *pre-radicaliseringsfase* ook wel de *kennismakingsfase* waarin personen een identiteit zoeken en een groep om bij te horen. Volgens deze auteurs speelt binnen de kennismakingsfase ideologie nog geen grote rol. Het conformeren aan de eisen van de groep en de ideeën die hierbinnen leven, gebeurt volgens hen pas later in de radicaliseringsfase.

**Radicaliseringsfase.** Een mechanisme dat in deze tweede fase kan plaatsvinden, is *isolatie* (Mølmen & Ravndal, 2021). Personen die offline geen (sociale) aansluiting hebben, maar online wel sociale verbondenheid ervaren, kunnen zich als gevolg daarvan (nog verder) terugtrekken uit netwerken en sociale relaties in de fysieke wereld. Zij raken in zekere zin steeds meer ‘ondergedompeld’ in radicale online gemeenschappen, hetgeen bijdraagt aan verdere sociale vervreemding. In dit verband spreken Bjørgo en Carlsson (2005) van ‘bridge burning’, een fenomeen waarbij een persoon publiekelijk extremistisch(er) gaat handelen – soms onder druk van de extremistische groep of juist om loyaliteit aan deze groep te tonen. Gevolg hiervan is dat sociale banden met familie, vrienden en/of collega’s al dan niet onder druk verder verslechteren en het een ‘weg terug’ bemoeilijkt voor een persoon. Het construeren van een groepsidentiteit, waarbij een gevoel van gemeenschap en eenheid ontstaat, is bovendien één van de manieren waarop sociale media worden gebruikt door extreemrechts. Dit wordt versterkt door taalgebruik, symbolen, teksten en merchandise, waardoor er als het ware een subcultuur ontstaat waarvan iemand deel gaat uitmaken. Het gevoel van ‘samen tegen de wereld’ dat vaak wordt gecreëerd, zorgt voor verbondenheid en het gevoel een gezamenlijk doel te hebben (Peeters et al., 2022; NTA, 2018).

Een ander mechanisme in deze fase is *echoing*, wat verwijst naar de manier waarop constante interactie met anderen met dezelfde extreme denkbeelden en overtuigingen de eigen overtuigingen van een persoon kunnen ontwikkelen en versterken – met als mogelijk gevolg zelfs het legitimeren van geweld (Mølmen & Ravndal, 2021). Zogenaamde 'echokamers' die online kunnen ontstaan, zorgen ervoor dat mensen bevestigd worden in hun denkbeelden, radicaler gaan denken en handelen en het reguliere nieuws (nog meer) gaan wantrouwen (Mølmen & Ravndal, 2021; ESS, 2021). Daarnaast wordt extremistisch gedachtegoed binnen deze echokamers genormaliseerd en is het lastig voor individuen om hier tegenin te gaan. Dit kan leiden tot uitsluiting van leden die nuanceren proberen aan te brengen, met wederom als resultaat een online ruimte waarin enkel radicale geluiden te horen zijn (NTA, 2018). Van belang hierbij is dat ook de algoritmen van sociale mediaplatforms kunnen leiden tot steeds extremer wordende denkbeelden (Mølmen & Ravndal, 2021). Verondersteld wordt dat de aanbevelingsalgoritmen op sociale mediaplatforms steeds radicalere content aanbieden om de aandacht van de gebruiker vast te houden (Van Puffelen, 2021); een nadere toelichting hierop staat beschreven in Paragraaf 4.2 en 4.3. De amplificerende werking van deze algoritmen, zo wordt verwacht, kan ervoor zorgen dat een persoon in een zogeheten 'konijnenhol' (*rabbit hole*) belandt, waarbij de persoon steeds extremere content wordt getoond.

Een derde mechanisme in de radicaliseringsfase is *facilitering*. Onder facilitering vallen alle aspecten in de online context die de blootstelling aan en de interactie met extremistische content faciliteren. Internet kan radicalisering faciliteren, doordat het een laagdrempelig middel is voor informatie, communicatie en propaganda van extreme overtuigingen. Personen kunnen snel toegang krijgen tot rechtsextremistische content en daarnaast in contact treden met gelijkgestemden. Bovendien zorgt de mogelijkheid tot anonimiteit voor minder drempels om extreme content te verspreiden of tot zich te nemen. Het laatste mechanisme dat in deze fase kan plaatsvinden, is *acceleratie*. Dit verwijst naar het gegeven dat radicaliseringsprocessen in de online context

sneller plaatsvinden dan traditionele, op fysieke ontmoetingen geënte, radicalisering (Von Behr et al., 2013). In een kort tijdsframe kunnen personen snel en veel informatie tot zich nemen, waardoor de tijd waarin het gedachtegoed of gedragingen extremer kunnen worden aanzienlijk korter is (Mølmen & Ravndal, 2021).

**Operationele fase.** In deze laatste fase worden door Mølmen en Ravndal *facilitering* en '*action triggering*' genoemd als mechanismen. Naast het faciliteren van de ontwikkeling van ideologische overtuigingen in de radicaliseringsfase, kan dit mechanisme ook een 'operationele planning' faciliteren door het verzamelen van informatie (Mølmen, 2021). Denk bijvoorbeeld aan handleidingen voor het gebruik van wapens. Het tweede mechanisme dat kan plaatsvinden, is '*action triggering*'. Dit mechanisme verwijst naar het katalyserende moment waarop personen een online gecreëerde impuls ervaren om over te gaan tot actie of (politiek) geweld (Mølmen, 2021; Mølmen & Ravndal, 2021). Dit kan onder meer gebeuren door het meemaken van een online persoonlijke of sociale gebeurtenis die ervoor zorgt dat een persoon geïnspireerd raakt om daadwerkelijk geweld te plegen. Ook een (gedeelde) 'succesvolle' gewelddadige actie van iemand anders kan een trigger zijn voor iemand om zelf ook over te gaan tot actie. Door de vele gewelddadige content is het mogelijk dat '*action triggering*' zelfs volledig online plaatsvindt. In de praktijk komt rechtsextremistische niet-gewelddadige '*action triggering*' aanzienlijk vaker voor dan rechtsextremistische gewelddadige '*action triggering*'. Oproepen tot democratische vormen van verzet is namelijk één van de manieren waarop sociale mediaplatforms worden ingezet door rechtsextremistische activisten (Roks & Van der Schoot, 2019). Volgers wordt bijvoorbeeld gevraagd niet alleen online te reageren, maar ook offline 'de daad bij het woord te voegen' (Roks & Van der Schoot, 2019). Ook op Nederlandse accounts van sociale mediaplatforms is waar te nemen dat er wordt opgeroepen tot verzet. Ondanks dat de geweldsdreiging over het algemeen beperkt blijft, is het online taalgebruik agressiever en opruiender geworden en is de kans tot het overgaan op geweld groter dan in het verleden (AIVD, 2018).

### **Flitsradicalisering, twijfel en de-escalatie**

Aanvullend op de mechanismen en fasen van online radicalisering, is het goed om stil te staan bij 'flitsradicalisering', 'twijfel' en 'de-escalatie' binnen het rechtsextremisme (Peeters et al., 2022). Zoals eerder genoemd, kan 'online' radicalisering een snel en heftig karakter hebben – we spraken, in navolging van Wagenaar (2021), van 'flitsradicalisering'. Vanwege de anonimiteit die het computerscherm biedt, zijn conversaties binnen chatgroepen veelal extremer dan in de fysieke sfeer, waarbij nieuwkomers in dergelijke groepen zich razend-snel conformeren aan de groepsnormen, daarbij niet beseffend dat sommige uitspraken voortkomen uit overdrijvingen en/of stoerdoenerij. Daarnaast is het internet een schier onuitputtelijke bron van informatie, die ook mensen die geïnteresseerd raken in rechtsextremistische narratieven in staat stelt om in korte tijd heel veel informatie tot zich te nemen. Tegelijkertijd is het belangrijk om te benadrukken dat radicaliseringsprocessen niet hoeven uit te monden in 'action triggering', maar vaak ook onderbroken of beëindigd worden (Kleuver, Van Dijk & Ramirez, 2017). Volgens Wagenaar (in: Peeters et al., 2022) krijgt elk radicaliserend persoon op een gegeven moment te maken met twijfel, hetgeen de-radicalisering tot gevolg kan hebben. Hij onderscheidt drie soorten twijfelers: (a) mensen die onderdeel zijn van een radicaal netwerk en bij wie twijfel zich vertaalt in het uit het netwerk stappen, omdat ze zich gesteund weten in dit proces door bijvoorbeeld vrienden, familieleden of medetwijfelers; (b) mensen die onderdeel uitmaken van een radicaal netwerk zonder dat mensen in hun directe sociale omgeving hiervan op de hoogte zijn, waardoor zij gemakkelijker in staat zijn om uit het netwerk te stappen – zij hebben immers iets om op terug te vallen; en (c) mensen, die ondanks de twijfel, onderdeel blijven uitmaken van het radicale netwerk, mede doordat zij geen alternatief zien door gebrek aan sociale contacten buiten het radicale netwerk. Belangrijk hierbij, zo betoogt Wagenaar, is dat mensen die primair online zijn geradicaliseerd vaak gemakkelijker kunnen uittreden, omdat hun deelname anoniemer is. Het maakt dat het vaak langer duurt voordat geradicaliseerde personen ook voor de (fysieke) buitenwereld bekend komen te staan als 'rechtsextremist' of 'geradicaliseerd'. Juist dit soort 'labels' kunnen een persoon het gevoel geven

dat er geen of nauwelijks meer een weg terug is. Omgedraaid: het ontbreken van die labels vanuit de directe fysieke en sociale omgeving kan uittreding vergemakkelijken.

### **2.5. Verschillen radicaliseringsmechanismen per ideologie?**

In hoeverre is er sprake van verschillen en overeenkomsten tussen (offline en online) mechanismen van rechtsextremistische gewelddadige radicalisering en andere 'typen' radicalisering? Verreweg de meeste literatuur over radicaliseringsprocessen en mechanismen baseert zich op jihadistische radicalisering. Er zijn desondanks een aantal comparatieve studies van rechtsextremistische en andere 'typen' radicalisering beschikbaar die in deze paragraaf worden toegelicht (Marwick et al., 2022; Ebner, 2017; Köhler & Ebner, 2019; Abbas, 2017; Rogers, 2020; Pisiou, 2015; Doosje et al., 2016; Bouhana et al., 2018; Brzuszkiewicz, 2020; Demant et al., 2008; Cermak & Gruenewald, 2015).

#### **Overeenkomsten**

Veruit de meeste studies stellen dat risicofactoren, risicoprofielen en radicaliseringsmechanismen relatief meer overeenkomsten dan verschillen vertonen tussen verschillende 'typen' radicalisering, zoals ook al in de voorgaande paragrafen naar voren is gekomen. In vergelijkende studies van Doosje et al. (2016), Brzuszkiewicz (2020) en Marwick et al. (2022) is er specifiek bekeken welke mechanismen en factoren op individueel-, groeps- en samenlevingsniveau een rol spelen bij radicaliseringsprocessen vanuit verschillende ideologieën. Volgens deze studies blijken er een aantal kenmerken/factoren/mechanismen van radicaliseringsprocessen herhaaldelijk terug te komen bij individuen en groepen met uiteenlopende ideologische overtuigingen (waaronder: links- en rechtsextremisme, religieus gemotiveerd en single-issue extremisme):

- Individuen/groepen met extremistische overtuigingen percipiëren 'serieuze' problemen in de samenleving en zijn zeer ontevreden over de manier waarop huidige instituties omgaan met dit probleem. Hierdoor daalt het vertrouwen in instituties (Doosje et al., 2016).
- Individuen/groepen met extremistische overtuigingen geloven dat hun normen en waarden superieur zijn, waardoor een sterk gevoel van wij-tegen-zij ontstaat (Doosje et al., 2016). Een 'outgroup' wordt gemarkeerd tot vijand (Brzuszkiewicz, 2020). Er worden visies gedeeld van een toekomst waarbij de tot vijand gemaakte groep de samenleving domineert (Marwick et al., 2022).
- Deze retoriek wordt gebruikt om geweld te legitimeren. Er wordt daarbij op een Machiavellistische wijze gedacht (Brzuszkiewicz, 2020). Dit betekent dat geradicaliseerde individuen (gaan) geloven dat het doel alle middelen heiligt. Binnen het ideologisch denkkader wordt geweld gelegitimeerd en aangedragen als oplossing voor het probleem dat wordt waargenomen in de maatschappij (Doosje et al., 2016).
- Brzuszkiewicz (2020) expliciteert een vierde overeenkomst tussen radicaliseringsprocessen met uiteenlopende ideologische overtuigingen. Zij stelt dat zowel binnen rechtsextremistische als jihadistische radicalisering een tweestrijd tussen revolutionaire en traditionele waarden wordt ervaren. De revolutionaire component is vooral belangrijk wanneer de radicale overtuigingen een kleine ideologische minderheid vormen in de samenleving. In deze fase ondersteunen ze allebei een grootschalige transformatie van de samenleving, met als doel de vermeende hedendaagse decadentie te vernietigen. Tijdens de tweede fase, die begint wanneer hun ideologieën aan kracht winnen, gaan ze over tot een herwaardering of heiliging van oude 'tradities' en trachten aan te sluiten bij een verheerlijkt verleden. Dit terugverlangen naar het geïdealiseerde verleden gaat veelal samen met een apocalyptische beeld van de toekomst waarin de vijand zal zegenvieren (Brzuszkiewicz, 2020; Marwick et al., 2022).

- Een volgend belangrijk kenmerk dat jihadisme en rechtsextremisme gemeen hebben, is de cultus van heldendom. Dit gaat gepaard met retoriek over het belang van eer, trots en wraak, waarbij fysieke en mentale kracht centraal staan (Brzuszkiewicz, 2020). Hierbij ligt de focus soms ook op 'hypermasculinity' en 'hypersexuality' (Marwick et al., 2022).
- Andere overeenkomsten tussen jihadistische en rechtsextremistische radicalisering zijn volgens genoemde auteurs een zwart-witte manier van denken en (op strategisch niveau) het aanzetten tot polarisatie en het verspreiden van desinformatie om ideologische standpunten te beargumenteren, te verspreiden en te normaliseren (zie ook: Paragraaf 3.2).

Ebner (2017) stelt daarnaast dat rechtsextremistische en jihadistische narratieven – 'Het Westen is in oorlog met de islam' en 'Moslims zijn in oorlog met het Westen' – elkaar perfect aanvullen, waardoor de twee uitersten retorische bondgenoten worden en er een "spiraal van haat" ontstaat. Doordat rechtsextremisten en islamitische extremisten er vervolgens in slagen om elkaars echokamers binnen te dringen, kan dit bijdragen aan wederzijdse radicalisering. Naast de hierboven beschreven studies zijn er vergelijkende onderzoeken uitgevoerd die zich specifiek focussen op *individuele* kenmerken van personen die rechtsextremistisch en jihadistisch gedachtegoed onderschrijven. In deze studies komt onder meer naar voren dat in Europa jonge geradicaliseerde moslimmannen en jonge rechtsextremistische witte mannen structurele problemen delen die de beide groepen ontvankelijker maakt voor radicalisering. Voorbeelden hiervan zijn ontwrichting en identiteitscrises, een gebrek aan hoop en beperkte educatie- en werkgelegenheidskansen (Abbas, 2017; Marwick et al., 2022). Demant et al. (2008) zien bovendien bij beide ideologieën een zogeheten 'underdogeffect':

*“De repressie van extreemrechts leidt er deels toe dat extreemrechts haar standpunten bevestigd ziet, namelijk dat zij en haar achterban van de ‘gewone man’ niet serieus genomen wordt door de gevestigde orde. Dat verleent extreemrechts overtuigingskracht. Hetzelfde kan gebeuren bij islamitisch radicalisme. Het radicale denkbeeld dat het Westen zich opstelt als de vijand van de moslims en geen moraal heeft, zien radicalen bevestigd in de repressieve, stigmatiserende sfeer ten aanzien van (orthodoxe) moslims (...).” Als gevolg hiervan, kunnen radicaliseringsmechanismen in de hand gewerkt worden.”*

Ook in andere comparatieve studies naar radicaliseringsmechanismen die resulteren in terrorisme wordt geconstateerd dat er vaker sprake is van overeenkomsten dan verschillen (Pisoiu, 2015; Marwick et al., 2022; Bouhana et al., 2018). Echter, bij radicalisering die leidt tot geweld tegen personen of objecten blijken wel deels andere factoren en mechanismen relevant dan bij radicalisering welke (nog) niet heeft geleid tot geweld. Zo blijkt uit een studie naar ontwikkelingspaden richting geweld bij vier Duitse jihadi's en drie rechtsextremisten dat in geen van de casussen sprake was van educatieve of economische deprivatie (Pisoiu, 2015). Wel kwam naar voren dat voor individuen uit de beide groepen het proces naar het gebruik van geweld overeenkomstige aspecten bevat. Deze zijn als volgt: (a) het adopteren van een alternatief wereldbeeld, (b) het ernstig in twijfel trekken van de status quo, (c) een sterk verlangen om veranderingen te beïnvloeden en (d) het onderhouden van persoonlijke relaties met extremistische netwerken (Marwick et al., 2022; Pisoiu, 2015). In een andere comparatieve studie is gekeken naar de kenmerken van *lone actor*-terroristen – zowel rechtsextremistische terroristen als terroristen met andere ideologieën (Bouhana et al., 2018). Hiertoe hebben de onderzoekers data bestudeerd van 182 *lone actors* die gewelddadige aanvallen hebben uitgevoerd in Europa of de Verenigde Staten of zulke aanvallen voorbereidden. Op basis van hun analyses concluderen Bouhana et al. (2018): *“While several differences are noted, perhaps the most surprising finding is the degree of similarity between right-wing extremist lone actors and those adhering to different ideological currents”*.

Zo blijkt dat crimineel en risicogedrag veel voorkomen bij lone actors, ongeacht diens onderliggende ideologie. Daarnaast zijn er overeenkomsten in cognitieve en morele omstandigheden, zoals inflexibiliteit van denken, woede, psychologische angststoornissen en egocentrisme.

### **Verschillen**

De hierboven beschreven studies laten zien dat er veel overeenkomsten zijn in radicaliseringsprocessen, ongeacht het 'type' ideologie. Desalniettemin zien enkele studies wél een aantal verschillen tussen rechtsextremistische radicalisering en andere 'typen' radicalisering. Demant en collega's (2008) beargumenteren bijvoorbeeld dat het islamitisch extremisme is gevormd rond een ideologie “die het onderwerp is van frequente reflectie en discussie”, terwijl een diepgewortelde ideologie bij het rechtsextremisme veelal minder prominent lijkt te zijn in het radicaliseringsproces. In een comparatieve studie van Cermak & Gruenewald (2015) naar gewelddadige extremisten in de Verenigde Staten werd een aantal verschillen geconstateerd in persoonskenmerken tussen extremisten met verschillende ideologische overtuigingen (*left-wing, right-wing en Al Qaeda*). Zo bleken rechtsextremisten vaker een man zijn, vertonen vaker een mentaal ziektebeeld en zijn vaker al eens gearresteerd op het moment dat zij een gewelddadige actie begaan. Deze studie kijkt echter opnieuw naar persoonskenmerken van extremisten, en niet zozeer naar de (verschillen in) ontwikkelingspaden van radicalisering die ten grondslag liggen aan het radicaliseringsproces. Wel is het voorstelbaar dat verschillen in persoonskenmerken ertoe kunnen leiden dat bepaalde ontwikkelingspaden van radicalisering binnen verschillende 'typen' radicalisering anders verlopen. Op basis van de huidige wetenschappelijke literatuur is het moeilijk om hier uitspraken over te doen, aangezien de literatuur voornamelijk overeenkomsten laat zien in de ontwikkelingspaden.



Tot slot, het is relevant om aandacht te geven aan een ander 'type' verschil welke blijkt uit wetenschappelijke literatuur, namelijk hoe de samenleving reageert op rechtsextremistische en jihadistische radicalisering. Marwick en collega's (2022) zien namelijk dat jihadistische aanslagen vaak sterk worden uitvergroot door de aandacht vanuit de media en andere instituties, wat sterk bijdraagt aan een vijandiger beeld van jihadisten. In het geval van rechtsextremistische acties lijkt deze uitvergroting minder het geval: zowel in de hoeveelheid aandacht die eraan gegeven wordt als de manier waarop het gelabeld wordt. Een gevolg van dit verschil is dat opvattingen van jihadisten worden ervaren als 'verder afstaand' van de doorsnee opvattingen in de samenleving – in vergelijking met rechtsextremistische opvattingen en hun afstand tot deze doorsnee opvattingen. Hoewel dit verschil weinig zegt over mechanismen van radicalisering zelf, is het interessant om dit *framing*-gerelateerde kenmerk op het niveau van de samenleving te bezien vanuit het eerdergenoemde 'underdogeffect' van Demant et al. (2008). Een grotere afkeer vanuit de samenleving voor een bepaald 'type' radicalisering (in dit geval het jihadistisch of islamitisch extremisme), alsook de groep of personen die daarmee geassocieerd worden (in dit geval islamitische groepen of personen), kan bijdragen aan een sterker gevoel van deprivatie door deze personen en/of een sterker gemarginaliseerde positie in de samenleving. In een reactie daarop kunnen sommige personen mogelijk ontvankelijker worden voor extremistisch gedachtegoed.

## 2.6. Voorspelt online radicalisering gewelddadig rechtsextremisme?

Deze paragraaf gaat in op de vraag of er een relatie bestaat tussen enerzijds online radicalisering en anderzijds het plegen van incidenten van gewelddadig rechtsextremisme en/of -terrorisme. Er zijn een aantal empirische studies die aanknopingspunten bieden om deze vraag te beantwoorden. Op basis hiervan blijkt dat veruit de meeste personen die (grotendeels) online radicaliseren uiteindelijk niet overgaan tot gewelddadige acties of terroristische daden (Marwick et al., 2022). Ontwikkelingspaden van radicalisering zijn namelijk

uiterst complex, waarbij zowel online als offline factoren een rol spelen en op elkaar inwerken (Gill et al., 2015, 2017; Koehler, 2014; Von Behr et al., 2013). In lijn hiermee constateren verschillende studies dat online radicalisering wel degelijk kan bijdragen aan gewelddadige acties. Om daarvan enkele voorbeelden te noemen: uit een studie van Kenyon et al. (2021: in Binder & Kenyon, 2022) naar terrorisme – niet beperkt tot rechtsterrorisme – wordt duidelijk dat online radicalisering een aanzet kan geven voor gewelddadige acties, zoals tevens blijkt uit de casussen van sommige gewelddadige lone actors (zie daarvoor: hoofdstuk 6). Desalniettemin stellen deze auteurs dat de algemene 'dreiging' van geweld die voortkomt uit online radicalisering slechts beperkt is (Binder & Kenyon, 2022). Ook Mølmen & Ravndal (2021) stellen op basis van verschillende casussen van rechtsextremistische lone actors dat het internet een faciliterende rol heeft gespeeld bij een deel van deze personen, maar dat de precieze relatie tussen het internet en hun offline daden complex blijft.

In een andere relevante studie van Gill et al. (2017) is er specifiek gekeken naar de rol van het internet binnen een steekproef van 223 terroristen uit het Verenigd Koninkrijk die veroordeeld zijn of om het leven kwamen tijdens de uitvoering van hun terroristische handeling. Hoewel deze steekproef zich niet beperkt tot louter rechtsextremisten, zijn hierin een aantal relevante inzichten naar voren gekomen. Zo blijkt uit deze studie dat (a) bij het overgrote deel (61%) van de terroristische daders er bewijs was van online activiteiten welke gerelateerd zijn aan radicalisering dan wel het plannen van een aanval; (b) ruim de helft van de terroristische daders het internet gebruikte om iets te leren van hun beoogde activiteit; en (c) een derde van de daders online materiaal gebruikte om hun aanval voor te bereiden, zoals instructievideo's om bommen te maken, materiaal over toezicht en beveiliging, handleidingen om gif te maken, handboeken voor aanslagplegers, materiaal over terreurtechnieken, instructies hoe men een zelfmoord vest kan maken of een lijk kan wegwerken en zogenaamde trainingshandleidingen voor terroristen (Hassan et al. 2018; Gill et al., 2017). Op basis van deze resultaten zien de auteurs enkele verschillen tussen jihadistische en rechtsextremistische terroristen. Zichtbaar

was namelijk dat rechtsextremistische terroristen significant eerder geneigd waren om (a) online te *leren* en (b) online te *communiceren* (met gelijkgestemden) dan jihadistische terroristen. Als overkoepelende mogelijke verklaring voor deze verschillen veronderstellen Gill et al. (2017) dat rechtsextremistische terroristen mogelijk vaker *lone-actors* zijn in vergelijking met jihadistische terroristen; en daarom (meer) gebaat zijn bij het internet. Aansluitend hierop vinden de auteurs dat *lone acting* terroristen eerder geneigd zijn om het internet te gebruiken dan terroristen die in groepsverband opereren. Het online *leren* bleek voor rechtsextremistische terroristen voornamelijk bedoeld om een terroristisch misdrijf voor te bereiden – zij hadden dan ook een statistisch verhoogde kans om eerder online te leren ter voorbereiding van hun terroristische misdrijf dan andere ‘typen’ terroristen (Gill et al., 2017). Het verschil in online communicatie tussen rechtsextremistische terroristen en jihadistische terroristen bleek grotendeels te wijten aan het feit dat eerstgenoemde een statistisch grotere kans had om extremistische online *forums* te gebruiken dan jihadistische terroristen.

Een laatste relevante studie om te noemen, is die van Hassan et al. (2018). In deze systematische reviewstudie is er aandacht voor de relatie tussen enerzijds blootstelling aan online rechtsextremistische content en anderzijds gewelddadige radicalisering. Evenals Gill et al. (2017) stellen Hassan et al. (2018) dat er aanwijzingen zijn dat het internet en sociale mediakanalen een katalyserend effect hebben op extremistisch geweld, omdat personen die een risico op radicalisering lopen actief materiaal kunnen zoeken om hun belangen te bevredigen, lid kunnen worden van online en offline groepen en informatie kunnen verkrijgen die helpt bij het plegen van aanslagen. De belangrijkste bevindingen zijn: (a) huidige empirische onderzoeken lijken een voorlopig bewijs te geven dat blootstelling aan extremistische content geassocieerd kan worden met extremistische online én offline attitudes, evenals het risico op het plegen van politiek geweld onder rechtsextremistische en radicaal islamitische bewegingen; en (b) ‘actieve’ zoekers van gewelddadig online materiaal lijken eerder betrokken te raken bij politiek geweld dan ‘passieve’ zoekers. De

auteurs stellen op basis van hun resultaten dan ook het volgende: *“Given the current state of evidence, it can be concluded that the roles of the Internet and social media seem to be ones of decision-shaping, which, in association with offline factors, can trigger or facilitate decision-making, particularly so for active seekers of extremist online content”* (Hassan et al., 2018).

Op basis van verschillende empirische studies kan er, ter beantwoording van de in deze paragraaf gestelde onderzoeksvraag, geconcludeerd worden dat aspecten in het online domein van radicalisering kunnen bijdragen aan gewelddadig rechtsextremisme en/of -terrorisme. Echter, in het merendeel van deze studies (maar, in het bijzonder Hassan et al., 2018; Gill et al., 2017 en Von Behr et al., 2013) wordt er benadrukt dat de dreiging van geweld die voortkomt uit *alleen* het online domein relatief beperkt is. Het wordt dan ook sterk aangeraden om de aandacht te richten op de *wisselwerking* tussen online en ‘offline’ factoren, die kunnen resulteren in gewelddadig extremisme (Hassan et al., 2018). Via deze weg kan er een nauwkeuriger inzicht verkregen worden in ontwikkelingspaden van gewelddadige radicalisering en de wijzen waarop het online domein bijdraagt aan bijvoorbeeld keuzes voor politiek geweld of terroristische aanslagen. Een andere kanttekening die in de bredere zin gemaakt kan worden bij de bovenstaande bevindingen, is dat causaliteit een belangrijke rol speelt hierin. Bepaalde correlaties, bijvoorbeeld die tussen ‘actieve’ zoekers van gewelddadig online materiaal en hun betrokkenheid bij politiek geweld in Hassan et al. (2018), kunnen namelijk ook andersom bestaan. Naarmate iemand ‘verder’ in het radicaliseringsproces is, is de kans groter dat deze zoekt naar gewelddadig online materiaal – zonder dat dit materiaal op zichzelf bijdroeg aan die betrokkenheid bij politiek geweld.

## Samenvatting Hoofdstuk 2

In dit hoofdstuk worden onderzoeksvragen 1 en 2 beantwoord:

1. Welk mechanisme of welke mechanismen zijn er in de wetenschappelijke literatuur bekend over radicaliseringsprocessen van individuen en groeperingen in de richting van rechtsextremistisch, gewelddadig gedachtegoed? Wat is hierbij de rol van online content? In hoeverre en op welke aspecten zijn er overeenkomsten óf verschillen in de mechanismen van rechtsextremistische radicalisering ten opzichte van andere vormen van radicalisering?
2. Wat is er bekend over de voorspellende waarde van online radicalisering naar het optreden van gewelddadig rechtsextremisme en/of -terrorisme? In hoeverre is dit uniek voor rechtsextremisme en/of -terrorisme?

De wetenschappelijke literatuur toont een grote veelzijdigheid in het identificeren van risicofactoren die kunnen bijdragen aan het ontstaan van radicaliseringsprocessen. Risicofactoren blijken hierin vaker overeen te komen dan te verschillen tussen verschillende 'typen' radicalisering – zoals links, rechts, religieus, anti-overheid et cetera. Dit hoofdstuk geeft daarom aandacht aan risicofactoren waarvan bekend is dat zij kunnen bijdragen aan de ontvankelijkheid van personen voor extremistisch gedachtegoed, waaronder rechtsextremistische radicalisering. Om een vollediger beeld te geven van de wijze waarop laatstgenoemde tot stand komt, is daaropvolgend beschreven hoe risicofactoren terugkomen in profielkenmerken van rechtsextremistische radicalisering. Er wordt afgesloten met een korte beschrijving van beschermende factoren die op basis van de literatuur geïdentificeerd zijn.

Mølmen en Ravndal (2021) identificeren zes mechanismen die zich ofwel volledig online, ofwel binnen interacties van online en offline contexten manifesteren. Het gaat hierbij om de volgende mechanismen: (1) compensatie; (2) isolatie; (3) echoing; (4) facilitering; (5) acceleratie; (6) 'action triggering'. Deze mechanismen kunnen plaatsvinden binnen drie online radicaliseringsfasen. De eerste fase betreft de pre-radicaliseringsfase waarbinnen behoeften en kwetsbaarheden die personen ontvankelijk maken voor alternatieve wereldbeelden een grote rol spelen. In de tweede fase komen personen in aanraking met radicale en extreme online content en worden daardoor beïnvloed. De derde fase omvat de operationele fase waarin personen extreme attitudes omzetten in (gewelddadige) actie.

Veruit de meeste studies stellen dat risicofactoren, risicoprofielen en radicaliseringsmechanismen relatief meer overeenkomsten dan verschillen vertonen tussen verschillende 'typen' radicalisering. We beschrijven studies naar welke mechanismen en factoren op individueel-, groeps- en samenlevingsniveau een rol spelen bij radicaliseringsprocessen vanuit verschillende ideologieën. Volgens deze studies blijken er een aantal kenmerken/factoren/mechanismen van radicaliseringsprocessen herhaaldelijk terug te komen bij individuen en groepen met uiteenlopende ideologische overtuigingen (waaronder: links- en rechtsextremisme, religieus gemotiveerd en single-issue extremisme). Desalniettemin identificeren we ook enkele studies waar wél een aantal verschillen tussen rechtsextremistische radicalisering en andere 'typen' radicalisering.

Tot slot wordt er ingegaan op de vraag of er een relatie bestaat tussen online radicalisering en gewelddadig (rechts-)extremisme. Op basis van empirische studies blijkt dat de meeste personen die online radicaliseren, niet overgaan tot gewelddadig gedrag. Ontwikkelingspaden van radicalisering zijn uiterst complex, waarbij zowel online als offline factoren een rol spelen en op elkaar in werken. Om de invloed van de online context op extremistisch geweld te kunnen begrijpen dient te worden gekeken naar deze wisselwerking.

# 3 Rechtsextremistische content

Onderzoeksvragen die worden beantwoord in dit hoofdstuk:

3. Wat is er in de wetenschappelijke literatuur bekend over de wijze waarop sociale mediaplatforms en andere online content kan worden geïdentificeerd als rechtsextremistisch? Hoe kan dit worden onderscheiden van andere extremistische content? Is hierbij een verschil tussen de diverse sociale mediaplatforms?

In dit hoofdstuk komt het volgende aan bod:

- Introductie (Paragraaf 3.1).
- Definiëren en identificeren van (rechts)extremistische content (Paragraaf 3.2).
- rechtsextremistische content: online strategieën (Paragraaf 3.3).
- Normalisering, acclimatisering en dehumanisering (Paragraaf 3.4).
- Rechtsextremistische content op mainstream en alternatieve platforms (Paragraaf 3.5).
- Rechtsextremistische content op gameplatforms (Paragraaf 3.6).
- Subculturen die online een gateway kunnen zijn voor het rechtsextremisme (Paragraaf 3.7).

## 3.1. Introductie

Dit hoofdstuk heeft enerzijds aandacht voor de definitie en inhoud van rechtsextremistische content op het internet, anderzijds is er aandacht voor de 'rol' van deze content binnen de online context van het rechtsextremisme en de effecten ervan op haar gebruikers. De opkomst en populariteit van sociale mediaplatforms en online gemeenschappen heeft namelijk bijgedragen aan een reeks frames, berichten en strategieën om rechtsextremistische ideeën wereldwijd te verspreiden (Marwick et al., 2022). Zo is het bijvoorbeeld zichtbaar dat rechtsextremistische activisten online bijdragen aan (a) de negatieve stereotypering van groepen in de samenleving, (b) het toeschrijven van persoonlijke en maatschappelijke problemen aan specifieke groepen, (c) het pleiten voor discriminatie van een bepaalde groep en/of (d) het aansporen, verheerlijken en bagatelliseren van geweld (Roks & Van der Schoot, 2019). Het tegengaan van (rechts)extremistische content op het internet en op sociale mediaplatforms is dan ook van belang voor het kunnen voorkomen van terrorisme en rekrutering binnen deze online context (Alava et al., 2020). Ook ondermijning van de parlementaire democratie door het de-legitimeren van overheid, politiek en instituties, intimidatie en bedreiging van gezagsdragers zijn uitingsvormen van rechtsextremisme in het online domein. Op basis van een verkennend onderzoek naar de aanwezigheid van extreemrechts op sociale mediaplatforms concluderen Roks & Van der Schoot (2019) dat rechtsextremistische groepen sociale mediaplatforms gebruiken om hun ideologie uit te dragen, een groepsidentiteit te construeren en op te roepen tot vormen van verzet. Naast het tot uiting brengen van rechtsextremistische ideeën en identiteiten zijn er op basis van de wetenschappelijke literatuur nog andere doelen te identificeren. Voorbeelden hiervan zijn het delen van rechtsextremistische propaganda, (direct) communiceren met gelijkgestemden, het vormen van virtuele netwerken en gemeenschappen met een collectieve identiteit, rekrutering én de verkoop van producten, zoals merchandise of muziek (Roks & Van der

Schoot, 2019). Gekeken naar de verschillende ‘typen’ sociale mediaplatforms en fora waarop rechtsextremistische content kan worden aangetroffen, is het zichtbaar dat niet alleen websites en fora met rechtsextremistische content, zoals Stormfront, maar ook platforms waarop minder of zelfs geen contentmoderatie plaatsvindt, zoals 4chan, 8kun, Gab en Reddit, en mainstream sociale mediaplatforms als Facebook en TikTok fungeren als podium voor rechtsextremistische overtuigingen (Roks & Van der Schoot, 2019).

### 3.2. Definiëren en identificeren van (rechts)extremistische content

Een exacte formulering van rechtsextremistische content is verre van eenvoudig. Definities van het rechtsextremisme zijn immers uiteenlopend en wisselend per auteur en/of land, met als gevolg dat het afbakenen van content die hier wel en niet onder geschaard wordt moeizaam is. Het online discours van rechtsextremisten is bovendien fluïde en voortdurend in ontwikkeling. In de woorden van Alava et al. (2020): *“Characterising a discourse that is coherent, repetitive and identifiable might be easy, but radical, terrorist discourse is a very complex linguistic and sociolinguistic phenomenon. Moreover, modes of its dissemination and communication are complex.”* Niet alleen de inhoud van online boodschappen is dus lastig in te delen langs scheidslijnen van ‘rechts-extremistisch’ en ‘niet-rechtsextremistisch’, ook het format waarbinnen deze inhoud wordt verspreid, is uiterst veelzijdig. Denk bijvoorbeeld aan tekstberichten, afbeeldingen, symbolen et cetera. Hier komt bij dat (rechts)extremistische boodschappen regelmatig voorzien zijn van retorische middelen als ironie en zogeheten *dog whistles*<sup>5</sup>, waardoor de schadelijkheid ervan ontkend of verborgen kan worden door rechtsextremisten. Ook worden berichten soms verspreid met symbolen of nummercombinaties die verwijzen naar een bepaalde term of gebeurtenis, waardoor het voor buitenstaanders vaak niet

bekend is waarnaar daadwerkelijk verwezen wordt. Het moge dan ook geen verrassing heten dat sociale mediaplatforms, overheden en inlichtingendiensten worstelen met het identificeren, modereren en verwijderen van online rechtsextremistische content.

#### **Inhoudelijke kenmerken en narratieven van het rechtsextremisme**

Het slaan van een brug tussen de definities van het rechtsextremisme en het typeren van (rechts)extremistische content blijkt lastig. In lijn met de definitie van het rechtsextremisme door de NCTV en AIVD (zie daarvoor: Hoofdstuk 1), welke het rechtsextremisme relateert aan denkbeelden gestoeld op vreemdelingenhaat, haat jegens vreemde (cultuur)elementen en ultranationalisme, wordt het rechtsextremisme in de literatuur over het algemeen gedefinieerd aan de hand van vijf kenmerken: nationalisme, racisme, xenofobie, voorkeur voor een sterke staat en antidemocratische attitudes (Sterkenburg, 2019; Mudde, 2000). Deze kenmerken tezamen hoeven echter niet tegelijkertijd aanwezig te zijn bij rechtsextremistische content. Ook vertonen deze kenmerken overlap met *radicaalrechts*. Radicaalrechts kan immers gekenmerkt worden aan de hand van de volgende kenmerken (Sterkenburg, 2019): autoritarianisme (strengere regulering van de samenleving en strenge straffen bij wetsovertredingen), nativisme (verlangen naar een homogene staat zonder *non-native* elementen) en populisme (in het bijzonder wanneer het aankomt op het definiëren van ‘de elite’ als homogene corrupte entiteit). De inhoudelijke kenmerken van radicaalrechts en het rechtsextremisme laten dus overeenkomsten zien, waardoor de inhoud van content niet altijd langs deze kenmerken beoordeeld kan worden. De vraag is dan wanneer de inhoudelijke boodschap van content dusdanig ‘extreem’ en daarmee onwenselijk is (en niet langer radicaal of activistisch) dat het idealiter verwijderd zou moeten worden van een mediaplatform. Het doelmatig verwijderen van online content kan bovendien een terugslag veroorzaken die vaak het onderwerp van discussie is: het kan immers ervaren of afgeschilderd worden als een inperking van de vrijheid van meningsuiting of een vorm van censuur (Barnet, 2021), waardoor nadrukkelijke voorzichtigheid hierin geboden is. Ook kan een te ‘strikte’ vorm van contentmoderatie z’n doel

<sup>5</sup> Meer informatie hierover staat beschreven in Paragraaf 3.3.

voorbij schieten, waardoor content verwijderd wordt die niet noodzakelijk extremistisch is. Een breed gedragen standaarddefinitie van (rechts)extremistische content ontbreekt daarentegen vooralsnog. Hoe (supra)nationale overheden, sociale mediaplatforms en andere relevante instanties hiermee omgaan, blijkt verschillend.

Om toch enige duiding te geven aan het vraagstuk naar een 'definitie' van het rechtsextremistische content, is het zinvol om te kijken naar een zevental narratieven die kenmerkend zijn voor het hedendaagse rechtsextremisme. De frames waarbinnen deze narratieven zowel online als offline uitdragen worden, zijn bedoeld om invloed uit te oefenen op de perceptie van een publiek op politieke en maatschappelijke trends. Wanneer het onder andere gaat over (online) propaganda die verwerkt is in berichtgeving, worden de volgende narratieven onderscheiden door Nikki Sterkenburg (2019):

- *"Wij worden bedreigd"* – dit narratief speelt in op "morele paniek" door te stellen dat 'natives' bedreigd worden. Vanuit een demografisch perspectief beweren rechtsextremisten dat autochtonen een minderheid zullen worden in hun land als een gevolg van dalende geboortecijfers, hogere geboortecijfers onder migrantengroepen en nieuwkomers. Concepten als 'blanke genocide', 'demografische jihad' en 'islamisering' worden hiermee in verband gebracht. Vanuit een cultureel perspectief wordt verondersteld dat 'linkse' onderwerpen als multiculturalisme en diversiteit de westerse cultuur zullen laten verdwijnen.
- *"Er is een samenzwering om ons [het 'blanke ras'] te verzwakken"* – een aanzienlijk deel van rechtsextremistische boodschappen staat in het teken van een vermeende samenzwering door kapitalisten, globalisten, linksen, elitisten, culturele marxisten, joden, zionisten of reptielen die tot doel heeft om het 'blanke ras' te verzwakken. Zo wordt er bijvoorbeeld verondersteld dat er een verborgen 'linkse' agenda uitgevoerd wordt door politici en overheden die onder andere immigratie en islamisering in de hand werken. Aan dit narratief worden talloze complottheorieën verbonden, waaronder die van QAnon (zie: Paragraaf 3.7).
- *"Multiculturalisme zal nooit werken"* – binnen dit narratief worden vaak raciale angsten, anti-immigratiegevoelens en -houdingen verspreid. Rechtsextremisten benadrukken dan sterk de culturele en etnische verschillen tussen het 'blanke ras' en de 'anderen'. Door dit te doen, onderstrepen zij het gegeven dat het *onmogelijk* is om verschillende culturen en etniciteiten vreedzaam naast elkaar te laten leven.
- *"Wij leven niet volgens onze natuur"* – allereerst beweren rechtsextremisten aldus dat multiculturalisme niet 'natuurlijk' is en dat mensen van nature tribaal zijn, ter ondersteuning van hun bewering dat het *niet* immoreel is om mensen te verdelen in 'wij' en 'zij' en daarnaar te leven. Ditzelfde narratief wordt óók gebruikt voor antifeministische en anti-LHBTQIA+ uitspraken. Veel rechtsextremisten stellen namelijk dat biologische verschillen tussen mannen en vrouwen als een gegeven geaccepteerd moeten worden. In lijn hiermee moeten traditionele rolpatronen behouden, hersteld en bevorderd worden. Dit narratief speelt ook in op het gevoel van een mannelijk recht – waarbinnen mannen functioneren als hoofden van grote gezinnen. Om die reden pleiten rechtsextremisten voor een conservatieve samenleving waarbinnen het 'blanke ras' enkel kan blijven voortbestaan als vrouwen het moederschap serieus nemen. Tevens geven rechtsextremisten feministische en LHBTQIA+ gemeenschappen de schuld voor dalende geboortecijfers.
- *"The Great Replacement"* – in lijn met dit narratief wordt eveneens verondersteld dat het 'blanke ras' zal uitsterven als een gevolg van dalende geboortecijfers en gemengde huwelijken. Aanhangers hiervan zijn ervan overtuigd dat 'blanke genocide' plaats zal vinden wanneer de 'blanke' meerderheid een minderheid wordt. Extreme aanhangers stellen zelfs dat een preventieve rassenoorlog *nodig* is om dit te overleven. Terroristen als Anders Breivik en Brenton Tarrant stelden dit narratief dan ook centraal in hun motivatie voor hun daden.
- *"Migranten krijgen de voorkeur boven autochtonen"* – dit narratief stelt het slachtofferschap van autochtonen centraal. Etnische en religieuze minderheden zouden door overheden bevoordeeld worden; overheden zorgen

daarmee niet voor hun 'eigen mensen'. "Volgens dit narratief krijgen minderheden 'alles' zonder daar iets voor te hoeven doen, terwijl autochtonen zich door het dagelijks leven worstelen zonder enige overheidssteun (Sterkenburg, 2019).

- "Verlies van self-government" – een narratief dat vaak terugkomt in rechtsextremistische kringen is het waargenomen verlies van zelfbestuur. Verondersteld wordt dat belangrijke organisaties als de EU, NAVO en de VN, alsook nationale regeringen te machtig zijn geworden. Deze organisaties en instellingen zouden daarom enkel geïnteresseerd zijn in het behouden van hun eigen machtspositie, met als consequentie dat zij besluiten zullen nemen die ongunstig zijn voor burgers. Dit werkt institutioneel wantrouwen in de hand.

### **Sociale mediaplatforms: definitie van extremistische content**

Gekeken naar de wijze waarop verschillende sociale mediaplatforms (rechts) extremistische content definiëren, blijkt dat zij uiteenlopende definities hanteren. Zij zoeken, evenals overheden en inlichtingendiensten, naar praktische handvaten op basis van reeds bestaande definities in academische literatuur. Bovendien blijkt dat platforms over het algemeen geen onderscheid maken in verschillende 'typen' extremisme, maar hun definitie en/of moderatiebeleid enkel richten tot extremistische content in het algemeen. Zo verwijderd Facebook content die "geweld verheerlijkt of het lijden of vernederen van anderen viert" (Meta Transparency Center, webpagina, g.d.(1)). Specifieker gekeken naar de Richtlijnen voor de community van Facebook zijn er een zestal categorieën op basis waarvan Facebook content verwijderd of gebruikers blokkeert – of content voorziet van waarschuwingsberichten: (1) geweld en crimineel gedrag; (2) veiligheid; (3) aanstootgevende inhoud; (4) integriteit en authenticiteit; (5) intellectueel eigendom respecteren; (6) inhoud gerelateerde verzoeken en besluiten. Binnen deze categorieën zijn er een aantal voor deze literatuurstudie relevante definities die Facebook hanteert. Facebook definieert haatdragend taalgebruik als een directe aanval op mensen op basis van wat zij beschermende kenmerken noemen namelijk ras, afkomst, nationaliteit,

beperking, religie, seksuele geaardheid, kaste, sekse, geslacht, genderidentiteit, of ernstige ziekten. Daarnaast definieert Facebook een 'aanval' als gewelddadig of mensonterend taalgebruik, schadelijke stereotypen, beweringen over minderwaardigheid, uitingen van minachting, afschuw of verwerping, scheldwoorden en oproepen tot uitsluiting of segregatie. Daarnaast verbiedt het platform ook het gebruik van schadelijke stereotypen, die worden gedefinieerd als ontmenselijkende vergelijkingen die van oudsher zijn gebruikt om specifieke groepen aan te vallen, te intimideren of uit te sluiten, en die vaak verband houden met offline geweld" (Meta Transparency Center, webpagina, g.d.(2)).

YouTube schrijft in hun gebruikersvoorwaarden dat content die in strijd is met hun beleid/richtlijnen wordt verwijderd. Hierin maken zij onderscheid tussen haat zaaien en intimidatie. Content wordt gecategoriseerd als haat zaaiend wanneer de inhoud oproept tot haat of geweld gericht op groepen op basis van ongeveer dezelfde beschermende kenmerken als die van Facebook, te weten leeftijd, geslacht, ras, kaste, religie, seksuele geaardheid of veteranenstatus (YouTube, webpagina, g.d.). Daarnaast geeft YouTube aan dat hun moderatiebeleid ook gericht is op veelvoorkomende vormen van online haat, zoals het ontmenselijken van leden van deze groepen, deze leden wegzetten als van nature minderwaardig of ziek, het promoten van haat zaaiende ideologieën als het nazisme of van complottheorieën over deze groepen, of het ontkennen dat goed gedocumenteerde gewelddadige gebeurtenissen hebben plaatsgevonden. YouTube beschouwt content als intimiderend wanneer deze gericht is op één persoon met langdurige of kwaadaardige beledigingen op basis van intrinsieke kenmerken, waaronder hun beschermende groepsstatus of fysieke kenmerken. Dit omvat ook schadelijk gedrag, zoals het opzettelijk beledigen/beschamen van minderjarigen, bedreigingen, pesten, *doxxing* of het toejuichen van beledigend fan gedrag.

Op Twitter wordt in hun 'Hateful Conduct' genoemd dat het verboden is voor gebruikers om geweld te promoten en andere mensen rechtsreeks aan te vallen of bedreigen op basis van dezelfde al eerder genoemde beschermende kenmerken (Twitter Help Center, 2022). Twitter verwijderd accounts waarvan het primaire doel is om anderen schade toe te brengen op basis van één van deze factoren. Accounts worden verwijderd op Twitter wanneer er sprake is van gewelddadige bedreigingen; het wensen, hopen of oproepen tot ernstige schade aan een persoon of groep mensen; verwijzingen naar massamoord, gewelddadige gebeurtenissen of specifieke geweldsmiddelen waarbij beschermende groepen het primaire doelwit of slachtoffer zijn geweest; opruiing tegen beschermende categorieën/groepen; herhaalde en/of niet-consensuele beledigingen, scheldwoorden, racistische en seksistische teksten of andere inhoud die iemand vernedert; haatvolle beelden. Relevant is om te vermelden dat Twitter recentelijk verschillende bestuurlijke veranderingen heeft doorgemaakt. De aanstelling van topman Elon Musk brengt bovendien veranderingen ter wege wat betreft de contentregulering op het platform (NOS, 2022a). Zo had Musk als zelfverklaard strijder voor de vrijheid van meningsuiting eerder aangegeven dat Twitter een absolute vrijplaats voor het vrije woord moest worden. Na zijn aantreden hief Musk dan ook de schorsing op van accounts van controversiële personen die bekend staan om het verspreiden van nepnieuws. Inmiddels heeft Musk alweer zijn vertrek aangekondigd als topman bij Twitter wanneer hij een opvolger heeft gevonden, nadat uit een peiling bleek dat Twittergebruikers liever een andere topman zien (NOS, 2022b). Wat de toekomst is voor het platform en hoe de contentmoderatie er daarbij uitziet, is daarom voor nu moeilijk in te schatten.

Het relatief nieuwe mainstreamplatform TikTok – dat in een korte periode wereldwijd razend populair is geworden – heeft in de recente jaren ook een relatief uitgebreid beleid ontwikkeld omtrent ongewenste content op het platform. Zo definieert TikTok haatvolle content om en nabij hetzelfde als de hierboven genoemde platforms, maar TikTok weidt daarin een aparte alinea aan gewelddadig extremisme (TikTok, 2023). Daarbij benoemt het platform dat

het verboden is voor gebruikers om te dreigen of aan te zetten tot geweld, of om gewelddadige extremistische organisaties, personen of handelingen te promoten. Daarnaast gebruikt TikTok de term "georganiseerde haat" voor personen en organisaties die mensen aanvallen op basis van de eerdergenoemde beschermde kenmerken. Onder 'aanvallen' verstaan zij acties die aanzetten tot geweld of haat, die tot doel hebben personen of groepen te ontmenselijken, of die een haatdragende ideologie omhelzen. Onder 'extremistische groepen' vallen volgens het platform; terroristische organisaties, georganiseerde haat-groeperingen en criminele organisaties.

### **Identificeren van rechtsextremistische accounts en online content**

In een onderzoek van Crosset, Tanner en Campana (2018) wordt vastgesteld dat er online niet sprake is van één homogeen "rechtsextremisme", maar eerder van een groot aantal rechtsextremistische thema's dat naast elkaar bestaat op sociale media, waardoor er een grote verscheidenheid aan propaganda ontstaat. Zo stellen zij dat bijvoorbeeld neonazi's niet noodzakelijkerwijs dezelfde doelstellingen en inspiratiebronnen delen als andere rechtsextremistische bewegingen. Desondanks sluiten de door deze groepen gebruikte categorieën elkaar niet uit, en is het mogelijk dat op het kruispunt van discours categorieën zich in verschillende bewegingen bevinden. Of een sociale media-account rechtsextremistisch genoemd kan worden is eenvoudig vast te stellen wanneer het bijvoorbeeld het officiële account is van een rechtsextremistische groep, een oprichter van zo'n groep of een invloedrijke rechtsextremistische propagandist. Dergelijke accounts dragen vaak de naam van de groep of een persoon en gebruiken groepslogo's, afbeeldingen van leiders of bepaalde symbolen die met de beweging worden geassocieerd, waardoor ze gemakkelijk herkenbaar zijn. Er zijn echter talrijke online gebruikers die rechtsextremistische ideeën delen, maar niet noodzakelijk tot een specifieke groep of subcultuur behoren. Crosset et al. (2018) hebben na een uitgebreid onderzoek naar rechtsextremistische ideeën en theorieën een aantal *inhoudelijke* indicatoren vastgesteld die gebruikt kunnen worden om dergelijke gebruikers of accounts te identificeren. Deze indicatoren betreffen uitingen van extremistische vormen van nationalisme of



wit nationalisme; het idee van rassensuperioriteit; de afwijzing van “anderen” (zoals moslims en immigranten), de afwijzing van de moderne samenleving en/of cultuurmarxisme; een kritiek op gelijkheid, pluralisme en feminisme; en, in sommige gevallen, de legitimering van radicale actiemethoden (geweld).

### **Online rechtsextremisme als beeldcultuur**

Naast woordelijke online content wordt door rechtsextremistische groepen steeds meer gebruikgemaakt van beeldcultuur waarbij ideologische standpunten visueel worden afgebeeld. Voorbeelden daarvan zijn het gebruik van symbolen of zogeheten ‘memes’. Deze ‘memes’ zijn afbeeldingen of video’s met een korte tekst of andere beelddelingen met minimale tekst. Deze beelddelingen worden gebruikt om de aandacht te trekken, rechtsextremistische content of ‘verscholen’ boodschappen hierbinnen te normaliseren en personen via dergelijke beelden te (ver)leiden naar rechtsextremistische fuiken. Eventuele ideologische beïnvloeding en vorming van personen kan dan later plaatsvinden. Een duidelijk voorbeeld van een veelgebruikte meme is die van ‘Pepe the Frog’. Dit kikkerachtige karakter werd in 2005 ontworpen en werd in vele memes gebruikt. In 2016 werd Pepe toegeëigend door rechtsextremisten die de kikker gebruikten als afbeelding voor hun eigen nationalistische uitingen (ESS, 2021). Ook emoji’s en andere symbolen spelen binnen de rechtsextremistische internet wereld een grote rol. Zo wordt bijvoorbeeld de clown-emoji gebruikt als verwijzing naar ‘clownworld’ waarmee wordt verwezen naar het idee dat wij leven in een ‘absurde linkse wereld’. Ook worden symbolen door rechtsextremistische activisten gebruikt, zoals het ‘niet gelijk aan’-symbool om te communiceren dat zij het gelijkheidsbeginsel afwijzen (bijvoorbeeld bij gelijkheid tussen rassen of mannen en vrouwen). Sociale mediaplatforms hebben door de vaak impliciete extremistische boodschap die verpakt is in deze meme’s of emoji’s moeite met het modereren van de constante stroom aan dit soort content. Relevant is hierbij dat emoji’s en meme’s deze boodschap vaak impliciet verpakken in een grap, maar op alternatieve sociale mediaplatforms vaak ook expliciet weergeven, wat de normalisering van deze boodschap in de hand kan werken.

## **3.3. Rechtsextremistische content: online strategieën**

Er zijn zeer uiteenlopende manieren waarop rechtsextremistisch gedachtengoed de online leefwereld van personen kan binnendringen, met als gevolg dat online content potentieel een aanzienlijke rol kan spelen in ontwikkelingspaden van rechtsextremistische radicalisering. Een manier om hier duiding aan te geven, is door te kijken naar wijzen waarop online content door rechtsextremisten aangeboden wordt. Vraag is dan welke online strategieën van rechtsextremistische individuen en netwerken kunnen bijdragen aan de blootstelling aan rechtsextremistische content. Julia Ebner (2019 in: Fielitz & Thurston) maakt in dit verband een helpend onderscheid tussen drie ‘typen’ campagnes die door rechtsextremistische netwerken gebruikt worden. Deze zijn als volgt.

### **3.3.1. Radicaliseringsstrategieën voor gelijkgestemden of sympathisanten**

Deze strategieën omvatten het geraffineerde gebruik van ‘microtargeting’ van personen, het ‘kapen’ van jeugdcultuur en het uitbuiten van taboes en onderbelichte grieven, angsten en identiteitscrises, waarmee rechtsextremisten kwetsbare internetgebruikers naar hun netwerken kunnen lokken (Ebner, 2019). Een belangrijke verklarende factor voor personen – in het bijzonder jonge personen – om zich bezig te houden met of aangetrokken voelen tot het rechtsextremisme, is het ‘zoeken’ naar identiteit. Rechtsextremistische manifestaties spelen hier doorgaans handig op in door de vorm aan te nemen van zogeheten ‘*counterculture movements*’ die zich focussen op lifestyle, internet-jeugdcultuur en esthetiek (Peeters et al., 2022; Ebner, 2019). Via deze weg proberen zij jonge personen voor zich te winnen, te rekruteren en te mobiliseren. Online sociale mediaplatforms hebben een cruciale rol in het rekruteren van kwetsbare personen en dienen bij uitstek als ideale plek om nieuwe aanwas te spotten (ESS, 2019). Daarnaast worden personen in de online context niet alleen uitgenodigd om rechtsextremistische content te consumeren, maar worden zij in staat gesteld hier een actieve bijdrage aan te leveren en zelf

content te creëren – bijvoorbeeld door middel van blogs, vlogs, memes of zelfs livestreams waarin ‘maker’ en ‘gebruiker’ direct op elkaar kunnen reageren (Peeters et al., 2022). Dit wordt ook wel ‘*participatory culture*’ genoemd (Peeters et al., 2022). Hierin schuilt volgens Neo (2016) het gevaarlijkste aspect van het internet: de uitgestrekte online mogelijkheden die personen in staat stellen rechtsextremistische ideeën en gedachten te creëren, uit te wisselen met en te verspreiden onder gelijkgestemden – in plaats van beperkt te zijn tot het enkel ‘passief’ tot zich kunnen nemen van informatie. Voor een aantal makers van rechtsextremistische content heeft dit geresulteerd in een enorm bereik en duizenden volgers waarmee zij dagelijks in contact staan en hun ideeën verder verspreiden (Peeters et al., 2022). Zo zijn er bijvoorbeeld Amerikaanse *far-right streamers* actief op platforms als Twitch die op ‘vermakelijke en humoristische wijze’ witsuprematisme en antisemitisme uitdragen richting hun kijkers. Dit biedt vermaak voor de volgers, maar dient tevens als dekmantel door schadelijk en extremistisch gedachtegoed af te doen als een ‘onschuldige’ grap. Dit voorbeeld maakt dan ook deel uit van een breder aspect dat kenmerkend is voor rechtsextremistische content: vermaak.

### **Vermaak door content: het gebruik van humor, ironie en dog whistles**

Rechtsextremistische content dient regelmatig als een vorm van vermaak (ESS, 2021). Zo laat onderzoek van Schumpe et al. (2020) zien dat het zoeken naar sensatie een factor kan zijn om in aanraking te komen met extremistisch gedachtegoed en/of te radicaliseren. Om daarvan een voorbeeld te noemen: vermaak is voor sommigen een motief om zich bezig te houden met complottheorieën (Morosoli et al., 2022; Van Prooijen et al., 2021). Zij ervaren complottheorieën dan als fascinerend en spannend. Als gevolg daarvan kunnen deze personen – gemotiveerd door het vermaak en/of sensatie die het hen biedt – meer extreme content opzoeken of verspreiden. Manieren waarop het ‘vermakelijke’ element tot uiting komt in rechtsextremistische content, is echter erg veelzijdig. Om die reden heeft dit blok specifieke aandacht voor middelen die onder andere – en dus niet uitsluitend – door rechtsextremisten gebruikt worden om content ‘vermakelijk’ te maken voor internetgebruikers: humor, ironie en dog whistles.

Content in rechtsextremistische subculturen kenmerkt zich regelmatig door de overvloedige aanwezigheid van humor en ironie (Fielitz & Ahmed, 2021; Munn, 2019; Van Puffelen, 2021). Dit komt onder meer naar voren in memes, interne referenties en insider jokes – die zowel in tekst als in beeld worden weergegeven. Het gebruik van humor blijkt heden ten dage een centraal ‘wapen’ van rechtsextremisten om open samenlevingen te ondermijnen en de drempel naar geweld te verlagen (Fielitz & Ahmed, 2021). Zichtbaar is namelijk dat er steeds vaker op ‘speelse’ manieren wordt gepoogd om racistische boodschappen online te communiceren. De inzet van ironie hierbinnen heeft een belangrijke rol: het omvat een ‘slimme’ tactiek voor rechtsextremisten om de schadelijkheid van extremistische content te verbergen, ontkrachten of ontkennen. Grenzen tussen ‘kattenkwaad’ en potentieel radicaliserende boodschappen vervagen hierdoor (Fielitz & Ahmed, 2021).

Ironie wordt dan ook ingezet om “het vertrouwen in critici van rechtsextremisme te ondermijnen (‘linkse media reageren altijd zo overtrokken’ – ‘het was maar een grapje’)” en het stelt mensen in staat om daadwerkelijke toewijding aan rechtsextremistische ideeën zowel af te wijzen, als deze tegelijkertijd te omarmen (Van Puffelen, 2021; Wilson, 2017). Munn (2019) schrijft daarover het volgende:

*“By themselves, concepts like the supremacy of the white race or the solution of the ‘Jewish Question’ are too blunt, too forthright. (...) This ideology needs to be repacked in the visual vernacular of the Web: animated GIF’s, dumb memes, and clever references. The idea tumbler of the internet provides the perfect environment for image or language play, for absurd juxtapositions and insider jokes”.*

Naast ironie zijn dog whistles een ander retorisch middel dat gebruikt wordt door rechtsextremisten. Dog whistles zijn ‘verborgen’ politieke boodschappen in tekst of afbeeldingen die bedoeld zijn voor een specifieke groep en vaak enkel door hen begrepen worden. De impliciete betekenis van dog whistles kan altijd ontkend worden, met als gevolg dat gebruikers het norm-overschrijdende karakter ervan kunnen ontkennen en moderatie door sociale mediaplatforms kunnen ontwijken (Van Puffelen, 2021).

### 3.3.2. Manipulerende strategieën om de mainstream te bereiken

Een tweede ‘type’ strategieën dat onderscheiden kan worden, zijn de pogingen door rechtsextremistische netwerken om mainstreamdiscussies te beïnvloeden – met als doel hun eigen boodschap of gedachtengoed te verspreiden onder een breder publiek. Zichtbaar is dan ook dat rechtsextremistische content regelmatig te vinden is op mainstreamplatforms als Facebook, Twitter, YouTube, TikTok en Telegram, en daarmee tevens een proces van normalisering lijkt door te maken (Mudde, 2019; Maly, 2018). Om hierop aan te sluiten, zijn er diverse handleidingen gepubliceerd op rechtsextremistische fora die

uiteenzetten hoe personen overtuigd kunnen worden van rechtsextremistische narratieven – een radicaliseringsproces dat onder andere rechtsextremisten ‘red-pilling’ noemen (Peeters et al., 2022; Sterkenburg, 2019). Dergelijke handleidingen bevatten bijvoorbeeld richtlijnen voor het starten van gesprekken, het opbouwen van vertrouwen, het uitbuiten van gemeenschappelijke grieven en het afstemmen van de taal op de persoon in kwestie (Ebner, 2019). Hiermee kan op een effectieve wijze aangesloten worden op de eerdergenoemde risicofactoren voor rechtsextremistische radicalisering, hetgeen een persoon ontvankelijk maakt voor deze narratieven.

#### **Definitie: Red-pilling**

De term ‘Red-pilling’ is ontleent uit de film The Matrix waarin de hoofdpersoon Neo moet kiezen tussen de rode en de blauwe pil. Het innemen van de rode pil zorgt ervoor dat hij de ‘waarheid’ kan zien en zich zal realiseren dat hij weer uit ‘vrije wil’ kan handelen. De blauwe pil heeft tot gevolg dat Neo doorgaat met het leven zoals het al was – een wereld vol leugens, zo wordt verondersteld in deze film. Rechtsextremisten passen het onderscheid tussen deze pillen toe op de huidige samenleving waarin ‘bluepillers’ – ook wel ‘normies’ – de media en het systeem blindelings geloven, alsook de veronderstelde leugens die zij uitdragen en vertegenwoordigen. De ‘redpillers’ daarentegen zien volgens hen de wereld voor hoe deze ‘echt’ is en hebben het systeem dat bestaat uit allerlei leugens door. Hoewel ‘taking the red pill’ – een zinsdeel dat vaak gebruikt wordt door rechtsextremisten – lijkt te impliceren dat het slechts één moment van ‘bekering’ is, bestaat het daadwerkelijke radicaliseringsproces uit het ‘innemen’ van verschillende ‘red pills’ die bijdragen aan het alternatieve wereldbeeld van een persoon. Zo wordt het ontkennen van de Holocaust gezien als één van de meest ultieme rode pillen (Ebner, 2020).

Als onderdeel van het 'redpilling'-proces wordt er vaak online content verspreid die doorgaans wordt getypeerd als desinformatie en/of nepnieuws. Rechtsextremistische content wordt dan ook in verband gebracht met het ondermijnen van reguliere media en institutioneel vertrouwen (Munn, 2019), hetgeen vormend is voor de counteren van een mediaomgeving van een persoon waarmee een netwerkdynamiek ontstaat die kan bijdragen aan radicaliseringsprocessen. Munn (2019) schrijft het volgende over een dergelijke mediaomgeving:

*"Quietly but profoundly, this environment reengineers "whole realities that the user is then enabled to inhabit" and thus requires a form of re-ontologization (Floridi, 2016). This environment influences the way we "conceive and shape our sense of self" and the "way we understand and rationally organise our experience of the world" (Hernández-Ramirez, 2017). And this environment has the capacity to be transformative, to trigger "deep, structural shifts in the basic premises of our thought" (Gualeni, 2015)."*

Een ander onderdeel van manipulerende strategieën is strategische polarisatie, oftewel opzettelijk tweedracht te zaaien, gemeenschappen te verdelen en binaire wereldbeelden te verspreiden. Doel hiervan is het gematigde midden te dwingen een kant te kiezen om de invloed van politieke randgroepen te vergroten. Dergelijke strategieën starten vaak met rechtsextremistische internetgebruikers die proberen sympathisanten te rekruteren en te mobiliseren op berichtenplatforms als 4chan en 8chan. Vervolgens worden gesprekken voortgezet in versleutelde omgevingen als Telegram en Discord waarbinnen materiaal in de lokale taal verzameld wordt en campagnes gepland worden. Deze worden uiteindelijk gelanceerd op mainstream sociale mediaplatforms als Facebook, Twitter en TikTok. Het doel daarvan is om het online discours vorm te geven, de agenda voor discussies te bepalen en politici onder druk te zetten – met als beoogd resultaat dat 'middengroepen' politieke randgroepen gaan steunen (Ebner, 2019).

### 3.3.3. Intimiderende strategieën om tegenstanders te raken

Deze strategieën zijn erop gericht om kritische en politieke tegenstanders als journalisten, activisten en politici door middel van 'trolling', 'doxxing' en gecoördineerde haatcampagnes lastig te vallen – en bij voorkeur het zwijgen op te leggen of publiekelijk in diskrediet te brengen (Kreissel et al., 2018; Ebner, 2019). 'Trolling' omvat het verschijnsel waarbinnen ongevraagd en in hoge aantallen extreem commentaar geleverd wordt op sociale mediakanalen en andere internetfora om te provoceren, een reactie uit te lokken of haat te zaaien (Peeters et al., 2022). Zo komt het bijvoorbeeld voor dat dit vooraf afgesproken acties zijn in besloten groepen om gezamenlijk honderden reacties te plaatsen op hetzelfde moment. Naast eerdergenoemde tegenstanders, worden ook linkse of progressieve mensen, vrouwen, mensen uit de LHBTQIA+ gemeenschap en/of mensen van kleur het slachtoffer van trolling (NTA, 2018). 'Doxxing' is eveneens een methode die veelvuldig gebruikt wordt door onder andere rechtsextremisten (Ebner, 2019). Hierbij worden privégegevens van een persoon gepubliceerd om deze te beschadigen of te intimideren.

### 3.4. Normalisering, acclimatisering en dehumanisering

Een langdurige blootstelling aan rechtsextremistische content en -strategieën kan personen hiervoor ontvankelijker maken (Munn, 2019; Meloy et al., 2012; Grover & Mark, 2019). In de literatuur worden hiertoe meerdere zogeheten 'cognitieve fasen' geïdentificeerd. Deze fasen overlappen elkaar, zijn niet exclusief of strikt lineair, maar veeleer cyclisch van aard (Oboler, 2008; Dalgaard-Nielsen, 2008; Smith, 2011; Oboler, 2014; Munn, 2019). Het gaat om de volgende fasen:

1. Normalisering.
2. Acclimatisering.
3. Dehumanisering.

**Normalisering.** In deze fase spelen de eerdergenoemde humor en ironie van rechtsextremistische content een belangrijke rol. Vaak wordt dit aangeduid met de term “edgy” humor. Doel hiervan is om internetgebruikers vertrouwd te maken met extremistische ideeën, terwijl deze ideeën een “plausibele ontkenning” blijven behouden. Met andere woorden: door het te koppelen aan ironie vervaagd het onderscheid tussen ernst en satire, waardoor “niet-geïndoctrineerde personen niet kunnen zien of wij een grapje maken of niet” (Munn, 2019). Dit maakt racistische, seksistische of xenofobe boodschappen mogelijk, maar “maakt óók een haastige terugtocht mogelijk wanneer de spreker onder vuur komt te liggen” (Munn, 2019). Rechtsextremistische memes, om een voorbeeld te noemen, maken gebruik van technische functionaliteiten om hun zichtbaarheid en bereik te vergroten (Marwick & Lewis, 2017; Munn, 2019). Dergelijke memes worden gepost, aangepast en opnieuw gepost, en keer op keer worden personen eraan blootgesteld. Dit kan resulteren in een proces van normalisering waarin personen vertrouwd raken met voorheen voor hen schokkende inhoud. Hierin zijn memes op zichzelf relatief onschadelijk, maar de ideologieën die ingebed zijn in deze tekstjes kunnen langzaam deel uit gaan maken van het wereldbeeld van een persoon (Munn, 2019). Bovendien is het zichtbaar dat rechtsextremistische content steeds vaker te vinden is op mainstream sociale media, waardoor dergelijke content een proces van normalisering doormaakt op het samenlevingsniveau (Mudde, 2019; Maly, 2018).

**Acclimatisering.** Deze fase omvat het proces waarbinnen een internetgebruiker gewend raakt en ongevoelig is voor rechtsextremistische inhoud, waaronder racistische en vrouwonvriendelijke inhoud, met als gevolg dat er een *nieuwe cognitieve basis* voor acceptatie gecreëerd wordt (Munn, 2019). Het enorme aanbod van rechtsextremistische content en de snelheid waarmee dit onder de ogen van internetgebruikers komt, zorgt er namelijk voor dat het onmogelijk is om alles cognitief te beoordelen (Berardi, 2008; Munn, 2019). Dit heeft de volgende consequentie volgens Munn (2019):

*“Any ability to hold up each “fact” to scrutiny — to rationally process each trope and ethically weigh each stereotype — quickly becomes overwhelmed. Inundated in this way, the mind responds by becoming desensitized. Content and claims wash over the user. Racism moves from being a noticeable glitch to an environmental default, an accepted background to any online activity (Nakamura, 2013). Empathy dries up and the user becomes psychologically numbed. The mind adjusts to this new normal.”*

**Dehumanisering.** In deze fase zorgt dehumanisering ervoor dat internetgebruikers hele groepen van mensen gaan ervaren als ‘anderen’ – en zich hiervan gaan distantiëren (denk aan ‘indringers’, ‘strijders voor sociale rechtvaardigheid’, ‘culturele marxisten’ of ‘feministen’). Dehumanisering blijkt een belangrijke psychologische voorwaarde voor het gebruik van geweld (Steizinger, 2018). Het idee van dehumanisering is dat rechtsextremisten niet geconfronteerd worden met een ‘mens’ – een wezen dat rechten en vrijheden heeft, maar een ‘ander’ die tot een geheel andere ethische categorie gerekend wordt. Het resultaat is dat ‘echte’ mensen worden verdrongen door fobische ideeën over diens anders-zijn en vanwege eigenschappen als ras, religie, geslacht of levensstijl “hun menselijkheid en daarmee rechten verliezen of verspeeld hebben” (Munn, 2019; Smith, 2011). Dit creëert ruimte voor doelwitten om mishandeld of beheerd te worden. Rechtsextremistische content en taal kunnen bijdragen aan het dehumaniseren van groepen.

### 3.5. Rechtsextremistische content: mainstream en alternatieve platforms

Radicaal- en rechtsextremistische groepen zijn actief op verschillende sociale mediaplatforms, maar hebben nog altijd het grootste bereik via mainstream-platforms als Facebook, Twitter, YouTube en in iets mindere mate Instagram. Deze grote platforms hebben in de afgelopen jaren streng opgestreden tegen kanalen en pagina's die gelinkt zijn aan rechtsextremistische groepen of alternatieve media (Expertise-unit Sociale Stabiliteit, 2021). Alternatieve ofwel 'fringe' platforms zoals Gab, Telegram en 4chan zijn mede daardoor populairder geworden onder rechtsextremistische groepen, omdat hier meer extreme content kan worden gepost. Aangezien het bereik op alternatieve platforms kleiner is, wordt op de grotere platforms meer 'gematigd' materiaal gepost dat aan de voorwaarden van het platform voldoet. Wel is het zichtbaar dat gebruikers via deze 'gematigde' content soms doorgeleid worden naar meer extreme content elders. Alternatieve sociale mediaplatforms bieden daarnaast gelegenheid om contact te leggen met een online internationale gemeenschap van gelijkgestemden.

#### **Mainstream platforms**

Op de diverse mainstream sociale mediaplatforms wordt sinds enkele jaren aan contentmoderatie gedaan, waarbij extreme content of accounts actief worden verwijderd (Conway et al., 2019). In Hoofdstuk 7 gaan we in detail in op deze contentmoderatie door de platforms. Desalniettemin, worden de mainstream platforms verschillend gebruikt wat betreft de verspreiding van extremistische content.

*YouTube:* in de afgelopen jaren heeft YouTube de nodige kanalen verwijderd van het platform, omdat deze in strijd waren met de regels van het platform. Ondanks de aanscherpingen van het moderatiebeleid, blijft YouTube een platform waar relatief veel extremistische content wordt verspreid. Om daarvan een voorbeeld te noemen: YouTube dient onder meer als platform

voor samenzweringstheoretici en randgroepen die overtuigende en boeiende video's maken over problematische onderwerpen voor het daarin geïnteresseerde publiek (Lewis & Marwick, 2017). Het ontwerp en de structuur van YouTube geven ten minste vijf redenen tot bezorgdheid. Ten eerste is YouTube een open platform dat afhankelijk is van door gebruikers gegenereerde inhoud, waardoor mensen met afwijkende of extremistische standpunten rechtstreeks kunnen samenwerken met gevestigde media en informatiebronnen. Ten tweede kunnen de financiële prikkels die YouTube geeft op basis van het aantal kijkers en de kijktijd, makers aanmoedigen om mensen met extreme standpunten aan te spreken en controverses uit te lokken. Ten derde doet het algoritme van YouTube aanbevelingen, deels op basis van eerder gebruikersgedrag. Deze aanbevelingen kunnen het gedrag van gebruikers beïnvloeden, vooral omdat de topaanbeveling automatisch wordt afgespeeld nadat de huidige video is afgelopen. Ten vierde vereist video een beschrijving en duurt het langer om door mensen te worden bekeken dan tekst, waardoor de inhoud moeilijk te modereren is. Tot slot, het gebruik van video kan ertoe leiden dat kijkers parasociale relaties (een eenzijdige relatie met een persoon die je niet daadwerkelijk kent, maar alleen van diens verschijning in video's) aangaan met mensen wiens video's zij vaak bekijken, waardoor de effecten van blootstelling aan die video's kunnen toenemen.

Vooraf het aanbevelingsalgoritme van YouTube is alom bekritiseerd vanwege de bevordering van rechtsextremistische inhoud. Het platform beveelt gebruikers video's aan op basis van wat zij eerder hebben bekeken of opgezocht, en empirisch bewijs laat zien dat video's steeds extremere inhoud kunnen vertonen naarmate de gebruiker verder gaat in zijn zoektocht (Marwick et al., 2022). Dit kan bijdragen aan de verspreiding en normalisering van rechtsextremistische content. Naast video's op YouTube die van zichzelf een rechtsextremistische inhoud hebben, fungeert het platform ook als doorgeefluik voor anti-overheidspropaganda, nepnieuws en complotdenkers (Expertise-unit Sociale Stabiliteit, 2021). Hoe dit aanbevelingsalgoritme op YouTube werkt, wordt verder uitgelegd in Hoofdstuk 4.

*Facebook*: ook dit platform is afgelopen jaren strenger geworden in het handhaven van groepen en individuen die het platform gebruiken voor het verspreiden van *hate speech* en complottheorieën (Expertise-unit Sociale Stabiliteit, 2021). Rechtsextremistische content op Facebook wordt veelal beperkt tot comments bij reguliere nieuwsberichten, hoewel er daarnaast ook sprake blijft van enige activiteit op Facebookgroepen. Op het platform worden in privégroepen memes gedeeld die vervolgens verder worden verspreid via persoonlijke netwerken (Lewis & Marwick, 2017). Het platform is daarnaast ook een centrale plek voor het verspreiden van desinformatie aangezien het een populaire locatie is voor 'fake news' en nieuwsorganisaties gericht op één partij. Daarbij blijkt er verschil te zijn in de manier waarop open en gesloten groepen/pagina's gebruikt worden. Open Facebookgroepen worden soms gebruikt door rechtsextremistische groepen om verschillende activiteiten van activisten weer te geven. Het voordeel daarvan is voor hen dat deze pagina's door iedereen bezocht kunnen worden en daarmee naar verwachting een groter bereik hebben. Gesloten groepen worden vaker gebruikt voor coördinatie, netwerken, het opbouwen van bewegingen en het versterken van groepscohesie (Conway et al., 2019).

*Twitter*: ondanks het verwijderen van verschillende accounts en pagina's staat Twitter onder druk vanwege de grote hoeveelheden rechtsextremistische inhoud die op het platform wordt geproduceerd en verspreid (Expertise-unit Sociale stabiliteit, 2021). Dit komt door de openbare aard van het platform, waardoor meer gebruikers zich kunnen mengen in de rechtsextremistische discourse. De functies van Twitter stellen rechtsextremistische groepen in staat om relatief eenvoudig haatdragende inhoud te verspreiden, terwijl het ook de vorming van rechtsextremistische gemeenschappen faciliteert (O'Callaghan et al., 2013). Berichten (tweets) van leden van deze groepen worden regelmatig gebruikt om gebruikers door te sturen naar inhoud op externe websites, waaronder speciale websites die door bepaalde groepen worden beheerd of platforms voor het delen van inhoud, zoals YouTube.

*TikTok*: sinds enkele jaren behoort TikTok tot de grote mainstream platforms. Momenteel is TikTok één van 's werelds toonaangevende sociale mediaplatforms dat dagelijks door miljoenen mensen wordt gebruikt om video's te ontdekken, delen en te creëren. De video's op het platform geven aanleiding tot talloze virale trends op online platforms, waarbij inhoud ook wordt verspreid naar de andere sociale mediasites. TikTok is echter ook populair bij degenen die de online wereld willen misbruiken voor het maken, posten en promoten van haat en extremisme. Uit een onderzoek van het *Institute for Strategic Dialogue* (ISD) (2021) blijkt dat extremistisch en zelfs terroristisch beeldmateriaal gemakkelijk is te ontdekken op TikTok. Zo identificeert het onderzoek een aanzienlijke hoeveelheid video's die witte superioriteit promoten. De content varieert van aanhang voor de 'Great Replacement'-theorieën tot video's met muziek van *white power bands*. Ook beeldmateriaal van terroristische aanvallen zoals dat van Brenton Tarrant in Christchurch bleek makkelijk te vinden. Accounts met een extremistische aard gebruiken hun profiel om haat te promoten bijvoorbeeld door haat of extremistische referenties te gebruiken in hun profiel (gebruikersnaam, profielfoto etc.). Deze haatcontent gaat verder dan alleen witte superioriteit. Op TikTok wordt bijvoorbeeld content geplaatst waarin sprake is van haatspraak richting moslims, joden, Aziaten, vluchtelingen, vrouwen en de LHBTQIA+ community. Dit soort inhoud op het platform bereikt een aanzienlijk publiek. Het ISD-onderzoek (2021) geeft voorbeelden van haat of extremisme in video's die miljoenen keren zijn bekeken en waarop tienduizenden reacties zijn gegeven. Ten opzichte van de andere grote platforms blijft TikTok een uitzonderlijk geval, omdat het medium in een hele korte periode uitgegroeid is tot één van de grootste huidige sociale mediaplatforms. Het platform heeft hierbij moeilijkheden ervaren in het ontwerpen en uitvoeren van beleid dat gebruikers beschermt en content reguleert, zonder de groei van gebruikers te belemmeren of censuur te riskeren. Desalniettemin is TikTok inmiddels wel gestart met het verbeteren van hun moderatie van extremistische content.

### **TikTok: hashtags en sounds**

Zoals op vele andere sociale mediaplatforms wordt er ook op TikTok gebruikgemaakt van hashtags om inhoud voor gebruikers beter vindbaar te maken. Opvallend is dat uit hetzelfde onderzoek van ISD (2021) blijkt dat content met haatdragende en extremistische inhoud vaak voorzien wordt van algemene, populaire hashtags op TikTok. Voorbeelden daarvan zijn #foryoupage, #foryou en de afkorting hiervan #FYP. Deze hashtags verwijzen naar de 'home' pagina van de gebruiker waar als gevolg van de aanbevelingsalgoritmen op het platform gepersonaliseerde content wordt getoond. Het voordeel van deze algemene en populaire hashtags is dat zij een groter bereik creëren. Het onderzoek laat ook zien dat de makers van extremistische content ook gebruikmaken van hashtags gerelateerd aan algemene huidige politieke discussies en trends op TikTok, zoals #Politiek of #BlackLivesMatters. Door populaire topics te gebruiken beoogt men een groter publiek te kunnen bereiken. Daarnaast wordt er door rechtsextremisten op TikTok óók gebruikgemaakt van hashtags die direct geassocieerd kunnen worden met rechtsextremistische inhoud. Populaire hashtags hierbinnen zijn #MakeEuropeGreatAgain, #Fascism en #TheGreatReplacement.

Audiofragmenten zijn belangrijk voor de gebruikerservaring op het platform TikTok. Fragmenten van muziek of liedjes en geknipte fragmenten van populaire tv-shows of films vormen veelal de 'sounds' op TikTok. Sounds zijn een belangrijk onderdeel van TikTok, waardoor gebruikers door het platform worden geleid en specifieke content te zien krijgen. Sounds worden onder meer gebruikt om trends te creëren rond specifieke clips of stukjes audio. Sounds worden echter ook misbruikt voor het verspreiden van extreme content. Uit het onderzoek van de ISD (2021) blijkt dat geluiden die populair zijn bij veel TikTok-gebruikers (en dus hoog staan in de trends) ook populair zijn onder degenen die beogen om haat en extremisme te versterken.

Een voorbeeld hiervan is het gebruik van het nummer van MGMT 'Little Dark Age' wat in het onderzoek van de ISD (2021) veruit het populairste geluid onder extremistische boodschappen op TikTok is. Het nummer zelf heeft geen extremistische connotaties en was een wereldwijde trend op TikTok in 2020. Vervolgens werd het nummer echter ook gebruikt onder video's waarin bijvoorbeeld foto's werden afgebeeld van prominente fascistische dictators of aan extreem rechts gelinkte symbolen.

### **Alternatieve 'alt-tech' platforms**

Steeds vaker stappen gebruikers over van mainstream sociale mediaplatforms naar de kleinere, alternatieve mediaplatforms (Yonder, 2021). Veel van deze platforms beogen een alternatief te bieden voor de meer populaire technologieën en mainstream sociale media, waardoor deze platforms soms aangeduid worden met de term 'Alt-tech' (Yonder, 2021). Een motivatie voor deze gebruikers kan zijn dat zij de grote platforms te commercieel vinden, maar ook een strenger moderatiebeleid van mainstream sociale media kan gebruikers in de richting duwen van de alternatieve platforms – bijvoorbeeld doordat zij niet langer de content kunnen consumeren die zij wensen of juist op zoek zijn (of doorgeleid worden) naar content die een extremer karakter heeft. De alternatieve of 'niche' platforms kenmerken zich ten opzichte van mainstream platforms door een beperkt of soms zelfs afwezig beleid ten aanzien van contentmoderatie. Sommige van de platforms moedigen gebruikers zelfs aan om het platform te gebruiken vanwege dit gebrek – of, in hun woorden, 'voordeel'<sup>6</sup>. Enerzijds wordt verondersteld dat het beperkte gebruikersbeleid een gevolg is van het kleinere personeelsbestand op alternatieve platforms met daardoor minder technologische mogelijkheden, wat resulteert in minder moderatiemogelijkheden en meer vrijheid voor gebruikers om content te plaatsen (Finnerty &

<sup>6</sup> Bijvoorbeeld, de toenemende intolerantie van mainstream sociale mediaplatforms voor extreemrechtse inhoud en activiteiten heeft ertoe geleid dat extreemrechtse, veelal Amerikaanse, activisten hun eigen platforms oprichtten die dergelijke inhoud en activiteiten verwelkomen en zelfs aanmoedigen.



Grelich, 2020). Anderzijds wordt door verschillende ontwikkelaars van alternatieve platforms juist benadrukt dat het platform staat voor de vrijheid van meningsuiting en expressie, met als gevolg dat contentmoderatie (grotendeels) bewust afwezig is. Content met haatdragende inhoud en desinformatie kan daardoor vaker vrijuit worden verspreid en gedeeld worden onder gebruikers op deze platforms. De wisselwerking tussen mainstream en alternatieve platforms is tweeledig. Gebruikers van mainstream sociale media worden soms doorgeleid naar alt-tech platforms – bijvoorbeeld als rekruteringsstrategie waarin zij door andere gebruikers hiernaar doorverwezen worden; maar, andersom worden extremistische boodschappen die gebruikers aantreffen op deze platforms door hen juist verspreid op mainstream platforms met als doel extremistische content te normaliseren (Zannetthou et al., 2018; Malter, 2017).

Hoewel het aantal alt-tech platforms waarop rechtsextremistische content aanwezig groot is, zijn er enkele platforms die in het bijzonder relevant zijn om in deze literatuurstudie te onderscheiden. Een aantal alt-tech platforms zijn namelijk zo ingericht dat deze rechtsextremistische netwerken faciliteren. Voorbeelden daarvan zijn 4chan, 8kun, BitChute en Parler. Daarnaast zijn er alt-tech platforms die 'ideologisch neutraal' zijn, maar door beperkte moderatie vooralsnog ruimte creëren voor rechtsextremistische content – een voorbeeld daarvan is Discord (Freelon, Marwick & Kreiss, 2020). De website Yonder (2021) biedt een overzichtelijke introductie van een aantal alt-tech platforms ten opzichte van meer mainstream counterparts. Hier wordt verondersteld dat zowel reguliere mainstream platforms als 'niche' alt-tech platforms elk een eigen 'persoonlijkheid' hebben die hen kenmerkt als platform. *"They attract different people, employ different styles of user engagement, and tend to be 'known' for different things. But they have one important trait in common: they are walled gardens of homogenous thinking, tightly ideologized, with high engagement and the ability to easily influence and ignite their network"* (Yonder, 2021). De verschillende alt-tech platforms waarop rechtsextremistische content beschikbaar is, die Yonder (2021) onderscheidt, zijn de volgende :

- *Parler*: dit platform is opgericht in 2018 en wordt vaak getypeerd als het conservatieve alternatief van Twitter. Parler biedt een micro-blogging en sociale netwerkservice, welke is ontstaan als gevolg van internetgebruikers die door het moderatie en 'fact-checkbeleid' van Twitter voorkeur geven aan een platform met minder 'censuur'.
- *Gab*: dit platform is opgericht in 2017. Het platform dient als een 'toevluchtsoord' voor rechtsextremisten die verbannen zijn van mainstream sociale mediaplatforms. Onder andere neonazi's, aanhangers van witsuprematisme en de Alt-Right, rechtse conservatieven, complotdenkers, QAnon-leden en anti-vaxxers zijn aanwezig op dit platform.
- *MeWe*: dit is een minder bekend alt-tech platform dat is opgericht in 2012. Het biedt een alternatief voor Twitter en Facebook aan mensen met extremistische opvattingen.
- *4chan*: dit platform is opgericht in 2003 als onderdeel van de zogeheten 'chans' (te weten: 2chan, 4chan, 8chan/8kun): een online subcultuur van mensen met extremistische opvattingen. Het platform inmiddels voorzien van veel 'verontrustende, gewelddadige en illegale' content die nauwelijks gemodereerd wordt.
- *8kun*: voorheen bekend als 8chan, is opgericht in 2013. Dit platform is gegroeid door gebruikers die 4chan verlieten en op zoek gingen naar nóg extremere content. Het platform dient met name voor rechtsextremisten, waarvan sommigen geweld promoten. Terrorist Branton Tarrant publiceerde zijn manifest op dit platform, evenals andere rechtsextremisten. Verschillende terroristische aanslagen – waaronder die van Tarrant in Nieuw-Zeeland – werden bovendien 'met enthousiasme' live gestreamd op 8kun.
- *BitChute*: opgericht in 2017 biedt dit platform een alternatief voor YouTube. Het is daarmee bedoeld voor video-uploaders die het gebruikersbeleid van YouTube willen vermijden of zelf verbannen zijn van dit kanaal. Onder andere rechtsextremisten en complotdenkers zijn aanwezig op dit platform.

### **Messaging apps en 'dark social'**

Verschillende relatief nieuwe en zeer toegankelijke communicatie-apps zijn een steeds belangrijker onderdeel van de communicatiemogelijkheden van rechtsextremisten. Dergelijke apps zijn softwareprogramma's die zijn ontworpen voor gebruik op mobiele apparaten als telefoons en tablets (Conway et al., 2019). Veel van deze apps vallen in de categorie 'dark social'. Deze categorie verwijst niet zozeer naar de aard van de inhoud, maar naar de beperkte mogelijkheden om de inhoud te kunnen traceren via bijvoorbeeld *messaging apps* en andere vormen van versleutelde chats. Discord en Telegram zijn voorbeelden van dit soort applicaties. Op deze platforms kan er anoniem en in besloten groepen informatie worden verspreid die niet zichtbaar is voor gebruikers buiten die groep. Ook worden deze platforms bijvoorbeeld gebruikt om weblinks te delen naar 'Dark Web'-illegale verkoopsites, waar onder meer wapens verkocht worden. Wat betreft Telegram is het belangrijk om hier een nuance te maken. Dit platform kan zowel als mainstream als alternatief getypeerd worden. De inhoud van de messaging op dit platform is namelijk 'meer' mainstream, terwijl een deel van de aanwezige kanalen/groepen meer naar de alternatieve kant neigen. Desondanks kan er gesteld worden dat Telegram 'meer' alt-tech is, omdat het contactleggen met dit platform ingewikkeld is omwille van het ontbreken van traceerbare contactpersonen. Sinds 2015 wordt er echter wel content gemodereerd op het platform.

### **3.6. Rechtsextremistische content: gameplatforms**

Naast de beschikbaarheid van rechtsextremistische content en de aanwezigheid rechtsextremisten op sociale mediaplatforms, is er een groeiende aandacht voor rechtsextremisme op gameplatforms. Wereldwijde berichten over kwetsbare jongeren die gerekruteerd worden via online games leiden soms tot grote bezorgdheid bij ouders, voogden, professionals en overheden (Young, 2022). Hoewel gameplatforms niet de primaire focus zijn van deze

literatuurstudie, zijn er een aantal noemenswaardige aspecten die in het verband van deze studie relevant zijn. Gegeven dat de link tussen games en extremisme niet geheel nieuw is, is de link tussen gameplatforms en het online rechtsextremisme vandaag de dag niet meer weg te denken (Young, 2022). Zo zijn er terroristische manifesten van rechtsextremisten waarin verwezen wordt naar online games, worden er modificaties aangeboden van populaire games die specifiek aansluiten bij rechtsextremistische waarden ('white power'-spellen) en is er een overvloedige aanwezigheid van rechtsextremisten op populaire platforms als Discord, Steam en Twitch waarop games live worden gestreamd (Lee, 2021; Davey, 2021; Young, 2022). Het is daarom interessant om kort aandacht te hebben voor twee verschillende manieren waarop rechtsextremisten zich infiltreren in deze online gamewereld.

Zowel internet 'waakhonden' als counterterrorisme-experts waarschuwen regelmatig voor ten minste twee functies die online gameplatforms hebben voor rechtsextremisten: (a) het dient als effectief rekruteringsmiddel en (b) het draagt bij aan de verspreiding van rechtsextremistische propaganda onder een breder publiek (Duff, 2019). Wat betreft het gebruik van online gameplatforms en games als plaats om nieuwe aanwas te rekruteren, stelt Duff (2019) het volgende: *"Extremist groups are often trying to find vulnerable youth, and what are vulnerable youth doing? Playing games on the internet"* [daar waar jongeren die gamen natuurlijk niet op voorhand kwetsbaar hoeven te zijn]. Rechtsextremistische ronselaars zijn 'zoekende' naar personen die ontvankelijk zijn voor of geïnteresseerd zijn in bepaalde opvattingen. Zij benaderen deze personen en verwijzen hen bijvoorbeeld door naar andere rechtsextremistische sociale mediakanalen of fora. Dit zijn vaak besloten kanalen of chatrooms waarop weinig of zelfs geen moderatie plaatsvindt. Jonge, soms kwetsbare gamers kunnen zo vrij eenvoudig in aanraking komen met (rechts)extremistische narratieven – met name omdat ronselaars zich gemakkelijk infiltreren in populaire games als Fortnite, Minecraft en Call of Duty (Young, 2022).

*"Some of the themes of games give a starting point for a conversation. If it's a first-person shooter game, it could be a casual conversation about the kind of person you like shooting in the game"* (Duff, 2019). Rekrutering van nieuwe aanwas kan bijvoorbeeld in de 'in-game' chatfunctie, maar het kan ook plaatsvinden op de platforms waarop populaire games gestreamd worden, zoals Discord, Steam en Twitch. Met behulp van online conversaties worden verbindingen gevonden, opgebouwd en versterkt met anderen die eenzelfde mening delen. *"By playing together they can reinforce each other's beliefs, bond over "dark/edgy humour" and use the game to act out violent fantasies* (Young, 2022)." Een andere manier waarop gameplatforms gebruikt worden, is als plaats om rechtsextremistische ideeën en propaganda op te verspreiden onder een breder publiek. Om daarvan een voorbeeld te noemen: verschillende rechtsextremistische groepen gebruiken het gameplatform Steam als sociaal mediakanaal waarop zij hun boodschappen verspreiden. In dat geval heeft het platform met name een sociale functie. Het biedt een relatief veilige ruimte voor mensen om samen te komen, te communiceren, plezier te hebben, maar ook om te discussiëren over onderwerpen die raken aan rechtsextremistische narratieven (Jacob Davey in: Bedingfield, 2021). Laatstgenoemde wordt vergemakkelijkt door het gegeven dat er vaak weinig of geen moderatie aanwezig is op deze platforms.

### 3.7. Subculturen als online 'gateways'

Deze paragraaf geeft aandacht aan een aantal prominente subculturen die met name in het online domein een 'gateway' kunnen vormen voor rechtsextremistische radicalisering. Sommige narratieven die binnen deze subculturen worden verspreid hebben namelijk overeenkomsten met rechtsextremistische narratieven (zie daarvoor: paragraaf 3.2). Het overzicht van deze subculturen is niet uitputtend. De voorbeelden zijn met name bedoeld om te laten zien dat online subculturen die op zichzelf niet georiënteerd zijn op (rechts) extremistisch gedachtegoed – zoals incels, preppers of het paganisme – door de interpretatie van uitgangspunten of ideeën een aanleiding kunnen geven voor personen om zich te interesseren in rechtsextremistische narratieven of geconfronteerd te worden met inhoud die daaraan relateert.

### Conspiriteit en QAnon

Op het eerste gezicht lijken de werelden van het door links gedomineerde New Age (met daarbinnen een positieve focus op het zelf) en de door rechts gedomineerde complottheorieën (met daarbinnen een negatieve focus op de wereldpolitiek) heel tegenstrijdig. Echter, er is een punt waarop deze twee werelden hand in hand lijken te gaan (Ward & Vaos, 2011). Dit 'punt' wordt ook wel 'conspiriteit' genoemd. Conspiriteit is een snelgroeiend online fenomeen waarbinnen een ideologie uitgedrukt wordt die zich voedt met politieke desillusie en de populariteit van alternatieve wereldbeelden (Ward & Vaos, 2011). Het biedt een brede politiek-spirituele filosofie die zich baseert op twee kernovertuigingen: (a) een geheime groep controleert heimelijk de politieke en sociale orde en probeert deze te beheersen – deze overtuiging valt samen met traditionele complottheorieën; en (b) de mensheid ondergaat een 'verschuiving' in het bewustzijn – deze overtuiging is geworteld in de New Age-beweging en refereert naar een wereldbeeld waarin sprake is van een 'ontwaakt nieuw paradigma' (Ward & Vaos, 2011). Binnen conspiriteit heerst het idee dat de best mogelijke strategie om te kunnen omgaan met de dreiging van een totalitaire 'nieuwe wereldorde' is door te handelen in overeenstemming met dit ontwaakte 'nieuwe paradigma' (Ward & Vaos, 2011).

Eén van de bekendste complottheorieën die soms door zowel aanhangers van conspiriteit als door rechtsextremisten onderschreven wordt, is QAnon – daar waar de inhoud van beide fenomenen geen voor de hand liggende combinatie is. Oorspronkelijk is de QAnon-beweging in 2017 ontstaan in Amerikaanse *far right* politieke kringen. Aanhangers hiervan geloven in de aanwezigheid van een onzichtbare en invloedrijke elite met satanistische ideeën "die achter de schermen stiekem de touwtjes in handen heeft" – de zogeheten *deep state* die achter de rug van officiële overheden het wereldtoneel beheerst (Van der Sanden, 2020). Sommige stromingen suggereren dat deze 'elite' uit joden zou bestaan, anderen houden het op, al dan niet figuurlijk, reptielen. De oorsprong van QAnon gaat terug naar een gebruiker met de naam Q die op het eerdergenoemde forum 4chan verschillende cryptische berichten deelde

die verwezen naar deze *deep state*. Om daarvan een voorbeeld te noemen: er werden berichten geplaatst waarin verondersteld werd dat “wereldwijde elites kinderen ontvoeren en vasthouden in ondergrondse gevangenissen om een middel uit hun bloed te halen dat hun leven zou verlengen” (Van der Sanden, 2020). Q beweerde een hooggeplaatste functie te bekleden in de Amerikaanse regering en via deze positie over deze informatie te beschikken. Ondanks de inmiddels sterke indicaties dat Q slechts een hersenspinsel is van een 4Chan-gebruiker die later werd overgenomen door Jim en/of Ron Watkins, gelooft de QAnon-beweging nog altijd dat Q via deze berichten de toekomst voorspelt over de volgende stap die de ‘elite’ zal zetten om de wereld verder te domineren (Pousaz & Roten, 2022). Daarnaast wordt de Amerikaanse oud-president Donald Trump gezien als *ally* die in het geheim een strijd voert tegen de *deep state* (Bouma, 2020). Zichtbaar is dan ook dat Trump zich online regelmatig uitliet met QAnon-propaganda, zoals het retweeten van een tweet waarin Joe Biden een pedofiel genoemd werd.

Hoewel QAnon oorspronkelijk Amerikaanse wortels heeft, is het inmiddels ook overgewaaid naar Nederland. Het lijkt in zekere zin een paraplueterm waaronder steeds meer complotdenkers zich verzamelen (Van der Sanden, 2020). De Coronapandemie was hiervoor de katalysator, zo stellen Van Prooijen en Aupers (in: Van der Sanden, 2020). Deze complottheorie kende in Nederland bovendien ook een offline weerslag en leidde tot tientallen incidenten, waaronder het weigeren van coronavaccinaties, brandstichtingen aan 5G zendmasten tot algemene onrust in het dorp Bodegraven. Allemaal manieren om zich af te zetten tegen de ‘elite’ die volgens hen heerst. Relevant voor deze literatuurstudie is de ‘gateway’ naar extremisme (zowel anti-overheid, rechts en links) die QAnon vormt. Aanhangers worden namelijk voortdurend gestimuleerd om ‘eigen onderzoek’ te doen en raken zodoende soms verzeild in zogeheten online ‘konijnenhollen’ en/of ‘echokamers’ (zie: hoofdstuk 4). Als een gevolg daarvan kunnen hun denkbeelden een steeds extremistischer karakter aannemen.

### **Manosfeer en masculiniteit**

De zogeheten ‘manosfeer’ behelst een gevarieerde verzameling van online ‘mannen-ondersteuningsgemeenschappen’ die zijn ontstaan als reactie op onder meer het feminisme, vrouwelijk empowerment, maatschappelijke veranderingen en de vervreemdende effecten van het neoliberalisme (Rich & Buljaka, 2023). Binnen deze subcultuur wordt aan vrouwen en feministen de schuld gegeven van talloze problemen in de samenleving (zie daarvoor ook: paragraaf 3.2). Sommige van deze gemeenschappen moedigen daarom haat of zelfs geweld jegens vrouwen en feministen aan. Grant en MacDonald (2018) stellen dat de manosfeer of de daaraan verwante nadruk op masculiniteit op zichzelf niet rechtsextremistisch zijn, maar wel overlap kent met diverse extreemrechtse thema’s, met als gevolg dat het een (online) gateway kan zijn naar het rechtsextremisme. Zij stellen vast dat de subcultuur aldus voortkomt als tegenreactie op maatschappelijke veranderingen, zoals (1) de evolutie van de familiestructuur; (2) verhoogde participatie van vrouwen in voortgezet onderwijs en arbeidsmarkt; (3) inspanningen van de overheid om iedereen gelijk te behandelen, welke rechtstreeks tot anti-discriminatie wetten en actief multiculturalisme hebben geleid; (4) de evolutie van een passieve naar een actieve seksualiteit van vrouwen; (5) toenemende acceptatie van en gelijke rechten voor de LHBTQIA+ gemeenschap; (6) meer invloed van vrouwen en mensen van kleur in het culturele veld; en (7) verschuiving naar een postindustriële economie.

Globaal zijn er vier hoofdgroepen te onderscheiden binnen het manosphere-netwerk (Aiston, 2021):

1. Mannenrechtenactivisten die pleiten voor politieke veranderingen die gunstig zijn voor mannen. Hun activisme uit zich in intimidatie tegenover vrouwen.
2. Mannen "die hun eigen weg gaan" en beweren dat vrouwen zó giftig zijn dat mannen hen moeten vermijden.
3. Zogeheten 'pick-up' artists die mannen allerlei verleidingsstrategieën leren, zodat zij succesvoller kunnen zijn in het versieren van vrouwen. Veel van deze strategieën zijn gericht op misbruik van vrouwen, waaronder belediging ('negging') en het negeren van toestemming.
4. Incels – ook wel onvrijwillige celibatairen – die geloven dat zij het recht hebben op een relatie met een vrouw, maar door (onder andere sociale) structuren in de samenleving niet in staat zijn een partner te vinden. Aan personen uit deze subcultuur zijn inmiddels meerdere daden van extremistisch geweld en moord toegeschreven. Om deze reden gaan we verderop uitgebreider in op deze subcultuur van incels.

De manosphere spreekt voornamelijk een mannelijk publiek aan dat zich onder andere geconfronteerd ziet met romantische afwijzing, vervreemding, economisch falen, eenzaamheid en een sombere visie op de toekomst (Rich & Buljaka, 2023). Het grootse probleem ligt volgens hen in wat zij noemen 'mannelijke onrechtiging', die zich richt tot de gevolgen van het feminisme. *"Here it contrasts the growing challenges faced by men with the increasing social, economic and political success experienced by women. This zero-sum claim posits that female empowerment must necessarily equate to male disempowerment, and is evidenced through simplified and pseudoscientific theories of biology and socioeconomics"* (Rich & Buljaka, 2023) – zoals dit ook naar voren komt in rechtsextremistische narratieven. Voor veel jonge mannen begint hun 'kennismaking' met de manosphere daarom niet zozeer met een haat jegens vrouwen of feministen, maar met het verlangen om onzekerheid weg te nemen over hoe de wereld om hen heen werkt (Rich & Buljaka, 2023).

Diverse online influencers exploiteren de onzekerheden en angsten die leven bij dit mannelijk publiek, en dragen hier actief aan bij (Rich & Buljaka, 2023). Bekende voorbeelden daarvan met grote aantallen volgers zijn Andrew Tate, Jordan Peterson en 'alpha'-mannen zoals Elliott Hulse. Andrew Tate beweert bijvoorbeeld dat vrouwen medeschuldig zijn aan seksueel misbruik. Daarop werd hij verbannen van sociale mediaplatforms als YouTube, Facebook, Instagram en TikTok. Het Instagramaccount van Tate had 4,7 miljoen volgers toen het werd verwijderd. In de afgelopen jaren zijn Tate's video's op TikTok miljoenen keren bekeken waarin hij niet alleen vrouwonvriendelijke opvattingen uitdraagt, zoals het goedpraten van geweld tegenover vrouwen, maar ook racistische en anti-LHBTQIA+ uitspraken doet. Tate is in december 2022 in Roemenië gearresteerd op verdenking van onder meer mensenhandel en verkrachting (Bolle, 2022). De gemeenschappen die gevormd worden door de achterban van deze online influencers werken als krachtige echokamers (zie daarvoor: paragraaf 4.3) die bijdragen aan het klimaat van spanning en hysterie. Volgers worden aangemoedigd om mannen belachelijk te maken die niet bereid zijn om de 'feministische samenzweringen' te erkennen die ten grondslag zouden liggen aan de huidige sociale en politieke crisis. Zodra online influencers als Tate eenmaal een omvangrijke achterban hebben opgebouwd, kunnen zij steunen op inkomsten die daaruit genereren. Influencers kunnen die inkomsten vervolgens herinvesteren en een cyclus voortzetten die helpt een nog groter publiek te bereiken (Rich & Buljaka, 2023).

Het alternatieve platform Reddit wordt veelal gezien als belangrijkste thuisbasis van vele manosphere-gemeenschappen (Aiston, 2021). Deze gemeenschappen hebben hun eigen websites of pagina's, maar zijn ook relatief gemakkelijk vindbaar via grote sociale mediaplatforms als Instagram, Facebook en Twitter. Ook op YouTube en TikTok kunnen jongeren geconfronteerd worden met seksistische en antifeministische content. Bovendien is er op TikTok een toename zichtbaar in 'pick-up' artists en mannen "die hun eigen weg gaan". Een voorbeeld van een community in de offline wereld die is voortgekomen uit de manosphere is de gemeenschap van de 'Alfamannen' die ook voet aan grond zette

in het Nederlandstalige deel in Europa. Hierover maakte de Nederlandse documentairemaker Nordin Lasfar in 2020 een documentaire (VPRO Tegenlicht, 2020). Lasfar volgde hierbij mannen die samenkomen in praatgroepen en discussiëren hoe ze hun mannelijkheid en maatschappelijke dominantie kunnen vergroten.<sup>7</sup>

### **Nadruk op masculiniteit en de overlap met rechtsextremistische narratieven**

Jonge mannen die zich aangetrokken voelen tot de manosphere vinden op websites en bij influencers informatie over mannelijkheid (masculiniteit). Hoewel masculiniteit op zichzelf geen politieke kleur draagt, constateren onderzoekers dat masculiene onderwerpen steeds belangrijker zijn geworden binnen rechtsextremistische propaganda (Jones et al., 2019; Lacoviello et al., 2021; Grant & MacDonald, 2018). Deze propaganda suggereert dat mannen steeds vrouwelijker en zwakker worden, hetgeen een zwakke maatschappij zou creëren. Dit zou het resultaat zijn van niet alleen het opkomende feminisme, maar dit wordt ook in verband gebracht met zaken als te weinig lichaamsbeweging en slechte (onmannelijke) voeding. Het vermeende tekort aan mannelijkheid van de "Europese blanke man" zou aan de basis liggen van een groot aantal maatschappelijke problemen. Het verlies of gebrek aan masculiniteit van de hedendaags man zou aanleiding zijn voor verhoogde angstgevoelens bij mannen, depressies, verslavingsproblematieken zoals gaming, alcohol en andere drugs (Borger, 2021). Sommige narratieven binnen de manosphere die raken aan extreemrechts stellen tevens dat door genoemde zwakheden "niet-Europese immigranten de blanke bevolking op termijn zal domineren en uiteindelijk vervangen" (de zogeheten omvolkingstheorie). Het opnieuw opkrikken van 'mannelijkheid' kan de maatschappij terugbrengen naar een

geïdealiseerd 'vroeger', met een door mannen gedomineerde maatschappij en slechts zeer uitzonderlijke migratie van buiten Europa. Om dit ideaal te realiseren, wordt fysieke fitheid en paraatheid gepromoot. Phelan en collega's (2023) onderzochten deze propaganda in meer detail, waarin zij verschillende onderwerpen tegenkwamen, zoals:

- Fysieke fitheid en discipline opbouwen door middel van lichaamstraining, fitness en gevechtssporten. Verschillende organisaties en politieke partijen in Nederland en België propageren bijvoorbeeld om deel te nemen aan gevechtstraining in groepsverband. Zo is MMA (Mixed Martial Arts) op dit ogenblik een populaire gevechtssport om "de jeugd weerbaarder te maken" (Zidan, 2022).
- Zelfstandigheid waarin men niet afhankelijk is van de staat. Denk hierbij aan het recht op bescherming van de eigen familie en een eigen voedselvoorziening door te tuinieren voor het eigen levensonderhoud.
- Het vergroten van weerstand tegen inmenging van staatsstructuren, bijvoorbeeld door middel van wapentraining en survivaltechnieken.
- Het verheerlijken van sterke (historische) mannelijke leidersfiguren in de maatschappij. Dit kunnen bijvoorbeeld belangrijke historische figuren zijn als Plato, Marcus Aurelius, Julius Caesar en Napoleon Bonaparte. Maar, ook invloedrijke en/of spraakmakende hedendaagse (politieke) figuren als Donald Trump, Vladimir Poetin en Andrew Tate.
- Het verwerpen van 'moderne maatschappelijke waarden' en het teruggrijpen naar traditionele (gezins)waarden waarbinnen de 'mannelijke' man een dominante en centrale inneemt in de maatschappij.

<sup>7</sup> De documentairemaker gaat ook naar de The 21 Convention, een manosphere-conferentie onder de titel 'The World's Greatest Convention for men' en praat met deelnemers die hun mannelijkheid willen herwinnen. Een man getuigt hierbij anoniem dat hij na het stuklopen van zijn relatie op zoek ging naar antwoorden. Hij zegt dat zijn eerste zoekvraag op het internet was "Wat willen vrouwen", een zoekterm die we zullen gebruiken in onze trial (zie daarvoor: Paragraaf 5.3).

Phelan et al. (2023) stellen dat door de blootstelling aan zulke narratieven jonge mannen kunnen radicaliseren. In het verlengde hiervan zien we dat ook de positie van de vrouw verduidelijkt wordt in dit soort propaganda. De vrouw vervult een centrale taak in het gezinsleven waarbij haar belangrijkste taken zijn om een "volgende" of onderdanige rol op te nemen in de gezinshiërarchie, zichzelf te goed te verzorgen (weliswaar enkel voor de goedkeuring van haar eigen man) en voor kinderen en het huishouden te zorgen. Deze levenswijze kan op zichzelf een vrije keuze zijn en hoeft voor betrokkenen noch samenleving bedreigend te zijn. Echter, zodra rechtsextremistische narratieven en deze traditionele opvattingen over vrouwenrollen elkaar ontmoeten, kan dit problematisch worden voor individuele rechten van vrouwen en ontspreken tot discriminatie en/of racisme (Sitler-Elbel, 2021). Om daarvan een voorbeeld te noemen: volgens deze levensvisie zouden enkel witte personen geschikt zijn als levenspartners om te kunnen zorgen voor een wit nageslacht.

Het Centre for Analysis of the Radical Right (2019) zoomde in op de opbouw van masculiniteit in rechtsextremistische propaganda en de manier waarop propagatoren hun publiek kunnen duwen richting extremere content. Zij kwamen tot de conclusie dat masculiniteit gerelateerde berichten op zichzelf niet problematisch zijn (het valt veeleer onder de zogeheten 'borderline content', zie daarvoor: hoofdstuk 5). Hier staat tegenover dat deze berichten wel degelijk kunnen leiden tot radicalisering bij een verhoogde of herhaalde blootstelling.

### **Incels**

De term *incels* is een afkorting van *involuntary celibates*. Dit omvat een subcultuur waarbinnen mannelijke leden geloven dat de maatschappij hun verlangen om seks te hebben, te daten of relaties te hebben met vrouwen, verhindert (Lindsay, 2020). Incels geloven namelijk dat er een seksuele hiërarchie bestaat in de westerse samenleving, waardoor zij verdoemd zijn tot een leven van eenzaamheid (Moskalenko et al., 2022a). Zichtbaar is dat er een toename is in radicaal gedachtegoed en terroristische acties door incels. Deze gemeenschap

bevindt zich echter vrijwel volledig online. Het vond zijn origine op een website in de jaren 90, waar mannen die moeite hadden om interacties aan te gaan met vrouwen advies konden vragen aan vrouwen op de site (Beaucamp, 2018). Dit veranderde in 2014: een haatcampagne genaamd Gamergate werd gelanceerd onder mannelijke gamers. Slachtoffers van deze campagne waren met name vrouwelijke videogame designers, die werden bedreigd met verkrachting en de dood (Moskalenko et al., 2022). Vanaf dit moment ontwikkelde zich een incel-cultuur in online domein. Jonge mannen treden toe tot de gemeenschap door "de zwarte pil te slikken" en daarmee tot het inzicht te komen dat zij het slachtoffer waren van feministische waarden en hersenspoeling (Ging, 2019). Incels geloven in een sociale hiërarchie met "Chads" (aantrekkelijke mannen) aan de top, gevolgd door "Normies", de normale meerderheid van de bevolking en op de bodem de incels. De ervaren oneerlijkheid onder incels leidt onder sommige van hen tot het goedpraten van geweld tegen de samenleving, met name verbaal of fysiek geweld tegen feministische vrouwen (Moskalenko et al., 2022a). In 2014 pleegde de 22-jarige 'incel' Elliot Rodger meerdere gewelddadige aanslagen, waarbij hij mensen neerstak, schoot en raakte met zijn auto. Hij bleek diep betrokken bij de incel-gemeenschap en had hier zelfs de status van een soort 'heilige' verkregen door het plaatsen van enkele haatzaaiende YouTube-video's en het uitbrengen van een manifest van ruim 130 pagina's. Hiermee worden andere incels aangemoedigd om gelijksoortige acties uit te voeren (Moskalenko et al., 2022b). Incels vormen daarmee een helder voorbeeld van hoe radicalisering in het online domein plaats kan vinden. Deze gemeenschap heeft niet enkel een extreme ideologie, maar ontwikkelde ook een eigen taal, retoriek en een cultuur van "trolling" en "shitposting", waarbij grappen en werkelijke opinies nauwelijks te onderscheiden zijn (Moskalenko et al., 2022a).

### **(Neo-)paganisme en Vikingcultuur**

Het heidendom of (neo-)paganisme oefent vaak een grote aantrekkingskracht uit op personen die kunnen worden beschouwd als progressief. Dit, omdat deze stroming gelijkheid, animisme, natuur, zelfontplooiing, spiritualiteit en individuele interpretatie centraal stelt. Het heidendom of paganisme is een overkoepele term voor stromingen als het Druidisme, Wicca, Asatru en het Keltische heidendom, welke veelal gebaseerd worden op Keltische en Germaanse tradities, alsook de Oudgriekse en Oudegyptische folklore (Zwissler, 2018). Voor vele paganisten of heidenen is hun levensbeschouwing ontstaan of ontwikkeld door een drang dicht bij de natuur te willen staan en bewuster stil te staan bij onder andere de seizoenen, verbinding en spiritualiteit. Vanwege de nadruk op onderlinge verbondenheid en het belang van stabiele relaties is er sprake van een wederzijdse beïnvloeding tussen het paganisme en bewegingen voor sociale rechtvaardigheid als het feminisme, de rechten voor de LHBTQIA+ gemeenschap en het milieubewustzijn. De nadruk op individuele gezondheid en ontwikkeling binnen het paganisme kan echter óók verenigbaar zijn met consumentisme, culturele toe-eigening en etnisch nationalisme (Zwissler, 2018). Onderdeel van de heidense of paganistische leefstijl is immers de verbinding met het land. Hierbij komen idealen als zelfvoorzienend- en bewust leven regelmatig naar voren. Deze waarden gaan bij sommigen gepaard met traditionele visies op mannelijkheid of een afkeer tegen de moderne wetenschap.

De zogeheten 'Vikingcultuur' is nauw verwant aan het (neo-)paganisme en kent veel gelijkwaardige symbolen. Hoewel de meeste aanhangers van de Vikingcultuur de progressieve waarden van de bredere pagan gemeenschap delen, is er aan de flanken sprake van een aantrekkingskracht van sommige ideeën die verwant zijn aan het rechtsextremistische narratief (De Zutter, 2010). Deze aantrekking ligt onder meer in de 'eigen' voorchristelijke cultuur, vrij van diversiteit en migratie evenals de pseudowetenschappelijke theorie dat Europa oorspronkelijk enkel door 'pure blanken' werd bevolkt en de bijhorende wens om dit te herstellen en behouden (De Wit, 2000). Hiervoor hanteert men soms de termen "het boreale Europa", de "Hyperborea" of de "Ultima Thule", die

een nationalistische en neofascistische connotatie kregen door onder andere Julius Evola en Aleksandre Doegin. Radicale aanhangers van de Noorse tradities keren zich daarom tegen het multiculturalisme en nemen hierbij soms racistische en rechtsextremistische standpunten in, verpakt in een paganistisch Vikingkleed van godensymboliek, seizoensfeesten, runetekens en zelfs muziekconcerten (De Zutter, 2010). Hermannson en collega's (2020) gingen undercover om de Amerikaanse gemeenschap van witte supremacisten te onderzoeken en beschreven het veelvuldig gebruik van Heathen-symbolen en rituelen die rechtsextremisten gebruiken om hun Europese afkomst en blanke identiteitsvorming te benadrukken. Belangrijke voorbeelden hiervan vond hij terug in de aanslag in Charlottesville (2017) en tijdens de Capitoolbestorming in de Verenigde Staten (2020). Onderzoekers Emore & Leader (2020), die zich toeleggen op deze problematiek, benadrukken echter dat dit slechts een klein deel van de overwegend politiek links geïnspireerde Paganistische gemeenschap behelst. Deze gemeenschap is zich ook sterk bewust van deze problematiek en reageerden krachtig door organisaties op te richten die de strijd met haatspraak aangaan, zoals Heathens against Hate en Gods and Radicals. Gemeenschappen die de Noorse spiritualiteit en het (neo-)paganisme praktiseren zijn momenteel erg gepolariseerd. Enkelen omarmen sterk de 'blank Europese heritage' of een 'volksche' component van het paganisme, welke door vele anderen als racistisch wordt aanzien. Het merendeel van de paganisten verwerpen deze ideeën met klem en pleiten sterk voor inclusiviteit, zoals hierboven is beschreven. Wie actief is binnen de Vikingcultuur en/of het paganisme komt vroeg of laat ongetwijfeld midden in deze storm terecht.



### **Preppers community**

De Preppers community is een politiek neutrale gemeenschap van mensen die zich willen voorbereiden op een extreme noodsituatie als gevolg van een ramp of sociale, politiek en/of economische opschudding en ongeregelde heden. Preppers hebben zich vaak medische kennis en zelfverdedigingstechnieken eigen gemaakt en zorgen voor allerhande voorzieningen die hen helpen te overleven en zo min mogelijk afhankelijk te zijn van de staat. Deze voorzieningen kunnen aangebracht zijn in hun eigen huis of op een andere locatie (als een schuilplaats of bunker). Bovendien worden door hen grote voorraden ingeslagen van (langdurig houdbaar, geconserveerd) voedsel, water en andere noodzakelijke goederen. Ook binnen rechtsextremistische kringen geniet dit preppen de nodige aandacht. De aantrekkingskracht ligt hier vooral in het niet langer afhankelijk zijn van de staat – waarin al het vertrouwen is verloren (Luke, 2021). Deze thema's zijn ook terug te vinden in het discours van extremistische organisaties als The Base, The Oath Keepers, The Three Percenters en de Siege-cultuur. In de meest extreme gevallen preppen extreemrechtse individuen ter voorbereiding op het instorten van de maatschappij en vanuit deze chaos een 'blanke etnostaat' te bouwen. Een concreet voorbeeld hiervan is de volgende casus. In 2022 viel de Belgische Federale Politie de woning van de beruchte complotdenker Yannick V. binnen nadat de politie zorgwekkende informatie ontving over een persoon die actief was in het rechtsextremistische milieu en zich mogelijk voorbereidde om het heft in eigen handen te nemen (Lefelon & De Roeck, 2022). Omdat Yannick V. een wapenliefhebber en sportschutter was, opteerde de politie om de speciale eenheden in te schakelen. Deze vrees leek terecht, want Yannick V. verwelkomde de politie met een vuurgevecht waarbij hij om het leven kwam. Uit onderzoek naar het profiel van Yannick V. bleek dat hij zich voorbereidde op het instorten van de maatschappij (Lefelon & De Roeck, 2022) - een gedachtengoed dat aansluit bij dat van de 'Preppers community'.

### **Andere 'gateways' via online influencers**

Naast eerdergenoemde influencers als Andrew Tate en Jordan Peterson zijn er online ook talrijke andere invloedrijke figuren die vaak via grote sociale media-platforms of gameplatforms een ingang kunnen zijn voor (jonge) personen om zich online verder te verdiepen in bepaald gedachtengoed of gemeenschappen.

### Samenvatting Hoofdstuk 3

Dit hoofdstuk beantwoordt onderzoeksvraag 3: Wat is er in de wetenschappelijke literatuur bekend over de wijze waarop sociale mediaplatforms en andere online content kan worden geïdentificeerd als rechtsextremistisch? Hoe kan dit worden onderscheiden van andere extremistische content? Is hierbij een verschil tussen diverse sociale mediaplatforms? Het hoofdstuk gaat in op de afbakening en inhoud van rechtsextremistische content op het internet, en op de 'rol' van deze content binnen de online context van het rechtsextremisme en de effecten ervan op haar gebruikers:

- Definiëren en identificeren van (rechts)extremistische content.
- Rechtsextremistische content: online strategieën.
- Normalisering, acclimatisering en dehumanisering.
- Sociale mediaplatforms waarop rechtsextremistische content aanwezig is.
- Subculturen als 'gateways' naar rechtsextremistische content.

Een exacte formulering van rechtsextremistische content is niet eenvoudig. Definities van het rechtsextremisme zijn immers uiteenlopend en wisselend per auteur en/of land, met als gevolg dat het afbakenen van content die hier wel en niet onder geschaard wordt, moeizaam is. Het online discours van het rechtsextremisme is bovendien fluïde en voortdurend in ontwikkeling. Toch zijn er bepaalde inhoudelijke kenmerken te identificeren die rechtsextremistische content beschrijven, deze worden in dit hoofdstuk toegelicht.

Er zijn zeer uiteenlopende manieren waarop rechtsextremistisch gedachtegoed de online leefwereld van personen kan binnendringen, met als gevolg dat online content een aanzienlijke rol kan spelen in ontwikkelingspaden van rechtsextremistische radicalisering. Een manier om hier duiding aan te geven, is door te kijken naar wijzen waarop online content door rechtsextremisten aangeboden wordt.

Vraag is dan welke online strategieën van rechtsextremistische individuen en netwerken kunnen bijdragen aan de blootstelling aan rechtsextremistische content. Hierop wordt in het hoofdstuk ingegaan:

- Radicaliseringsstrategieën voor gelijkgestemden of sympathisanten.
- Vermaak door content: het gebruik van humor, ironie en dog whistles.
- Manipulerende strategieën om de mainstream te bereiken.

Een langdurige blootstelling aan rechtsextremistische content – en strategieën kan personen hiervoor steeds ontvankelijker maken. In de literatuur worden hiertoe meerdere zogeheten 'cognitieve fasen' geïdentificeerd. Deze fasen overlappen elkaar, zijn niet exclusief of strikt lineair, maar veeleer cyclisch van aard. Het gaat om de volgende fasen: Normalisering; Acclimatisering; Dehumanisering.

Radicaal- en rechtsextremistische groepen zijn actief op verschillende sociale mediaplatforms, maar hebben nog altijd het grootste bereik via mainstreamplatforms zoals Facebook, Twitter, YouTube en in iets mindere mate Instagram. Deze grote platforms hebben in de afgelopen jaren streng opgestreden tegen kanalen en pagina's die gelinkt zijn aan rechtsextremistische groepen of alternatieve media. Alternatieve ofwel 'fringe' platforms zoals Gab, Telegram en 4chan zijn mede daardoor populairder geworden onder rechtsextremistische groepen, omdat hier meer extreme content kan worden gepost. Aangezien het bereik op alternatieve platforms kleiner is, wordt op de grotere platforms meer 'gematigd' materiaal gepost dat aan de voorwaarden van het platform voldoet. Wel is het zichtbaar dat gebruikers via deze 'gematigde' content soms doorgeleid worden naar meer extreme content elders. Online subculturen kunnen fungeren als gateways naar rechtsextremistisch gedachtegoed. Binnen sommige subculturen die niet noodzakelijk extreem-rechts zijn (zoals conspiratorialiteit, incels, preppers en de Vikingcultuur) kunnen personen gemakkelijk met rechtsextremistisch gedachtegoed geconfronteerd worden. Een andere ingang naar extremistische content wordt gevormd door sommige populaire online influencers.

# 4 Aanbevelingsalgoritmen als sociale mediafuik

Onderzoeksvragen die worden beantwoord in dit hoofdstuk:

1. In welke mate zijn er aanwijzingen dat een algoritme gestuurde aanbod van rechtsextremistische content op sociale mediaplatforms kan leiden tot een rechtsextremistische 'sociale media fuik'? Is hierbij een verschil tussen de diverse sociale mediaplatforms?

In dit hoofdstuk komt het volgende aan bod:

- Introductie over de werking van algoritmen (Paragraaf 4.1).
- Aanbevelingsalgoritmen op YouTube en TikTok (Paragraaf 4.2).
- Sociale mediafuiken, filterbubbels, echokamers, amplificatie en empirische studies in het kader van rechtsextremistische content (Paragraaf 4.3).

## 4.1. Introductie over algoritmen

Online platforms willen dat hun gebruikers zo lang mogelijk actief zijn op het platform. Hoe meer tijd een gebruiker op het platform doorbrengt, hoe groter de kans dat een gebruiker op één van de vele advertenties zal klikken en dus inkomsten genereert voor het platform. Ook genereert het platform inkomsten door het tonen van advertenties zelf. Algoritmen zorgen voor een aanbod van content dat beoogt de aandacht van de gebruiker vast te houden. In dit hoofdstuk bespreken we hoe dergelijke algoritmen mogelijkwerwijs kunnen leiden tot een 'sociale mediafuik' voor gebruikers en daarmee kunnen bijdragen aan (rechtsextremistische) radicaliseringsprocessen.

Een 'sociale mediafuik' houdt in dat gebruikers door algoritmen met name content te zien krijgen die aansluit bij hun interesses en surfgedrag, waardoor deze content steeds eenzijdiger van aard kan worden. Voordat we hierop nader in zullen gaan, bespreken we in algemene zin (de werking van) algoritmen op sociale mediaplatforms.

Wat zijn algoritmen op sociale media? Een algoritme is een technische formule waarmee platforms 'berekenen' wat zij een gebruiker willen laten zien. Algoritmen zijn daarmee een tool die platforms in staat stelt invloed uit te oefenen op de verblijfsduur op het platform van haar gebruikers. Door relevante content aan te bieden aan gebruikers, wordt beoogd dat zij geïnteresseerd/betrokken blijven en meer tijd doorbrengen op het platform. Om dit te kunnen doen, verzamelen algoritmisch gedreven sociale media een enorme hoeveelheid data die onder andere betrekking hebben op het online gedrag van gebruikers: het liken van content, het plaatsen van reacties, het volgen van pagina's of personen, het langer stilstaan bij bepaalde berichten (*dwell time*), het wel of niet bekijken van video's et cetera – om hiermee een nauwkeurig beeld te verkrijgen van content waarin een gebruiker geïnteresseerd zou zijn. Met behulp van algoritmische formules wordt dan een nieuw aanbod van relevante content 'berekend'<sup>8</sup>. Reden voor platforms om een algoritmisch aanbevelingssysteem te gebruiken, schuilt in het eigenbelang ervan. Sociale mediaplatforms zijn immers bedrijven die zichzelf moeten voorzien van financiële

8 Tekstfragment uit Kim (2017) over de werking van algoritmen: "After the algorithm creates a pool of potentially interesting content for a user, the algorithm gives each content a rank based on its appeal to the user. The algorithm then populates a user's feed so that content with a higher rank is prioritized and appears near the top of the feed. Pushing the most interesting content to the top of the feed is expected to increase the chance a user will engage with the post. (...) In the end, a feed curated by an algorithm is a culmination of juggling between multiple considerations and predicting the best arrangement of content for attracting user consumption. The way an algorithm comes up with an individual recipe for each user varies by company."

verdiensten om het voortbestaan ervan te verzekeren (Kim, 2017). Gebruikers zo lang mogelijk 'vasthouden' op het platform draagt bij aan deze financiële inkomsten en hangt dan ook samen met het verdienmodel van het platform. Wanneer gebruikers meer tijd besteden op een platform, is de kans bijvoorbeeld groter dat zij klikken op een advertentie die aan hen weergegeven wordt, hetgeen geld oplevert voor het platform. Sociale mediaplatforms zijn daarom niet altijd volledig transparant over de precieze werking van hun algoritmen, omdat zij dan in zekere zin hun businessmodel onthullen (Llansó et al., 2020).

Bovendien zijn platforms voortdurend bezig met het ontwikkelen, verbeteren en aanpassen van hun algoritmen, zodat de werkzaamheid ervan geoptimaliseerd wordt en het hen meer financiële verdiensten oplevert. Een manier waarop dit door vrijwel alle sociale mediaplatforms wordt gedaan, is door verschillende Artificial Intelligence-technieken te implementeren in aanbevelingssystemen (Menczer, 2021). Deze zelflerende netwerken en/of systemen zijn gunstig voor platforms, omdat zij niet alleen het algoritme verbeteren als het gaat over het aanbevelen van relevante content, maar het kan ook bijdragen aan het 'opsporen' en modereren van bepaalde content die in strijd is met de richtlijnen van een platform (Ofcom, 2019; Llansó et al., 2020). Echter, Artificial Intelligence-technieken kennen ook verscheidene nadelen. Zo is het denkbaar dat zelflerende technieken ongewenste gevolgen kunnen hebben die niet volledig te overzien zijn – bijvoorbeeld als het gaat om de verspreiding van extreme content of desinformatie. In de praktijk blijkt namelijk dat controversiële, provocatieve en extremistische content aantrekkelijk wordt bevonden door gebruikers die hier interesse in hebben getoond en daarmee hun betrokkenheid versterkt (Llansó et al., 2020). Het gevolg daarvan is dat algoritmische systemen 'leren' om deze content te prioriteren. Het gevaar dat hierin schuilt, is dat de aanbevolen content steeds extremer van aard kan worden – zonder dat een gebruiker of zelfs de 'maker' van het algoritme zich hier bewust van is of hiertoe aanzet heeft gegeven.

Er wordt, tot slot, gewaarschuwd dat algoritmen op sociale mediaplatforms niet neutraal zijn (Llansó et al., 2020; Stinson, 2022; Algo-transparency, weblink). Om alle data succesvol te kunnen verwerken, moeten algoritmen namelijk generaliserende aannames doen over de wereld. De personen die algoritmen implementeren of ontwikkelen zullen hier expliciet of impliciet vooroordelen over hebben die de werking van het algoritme kunnen kleuren. Tevens blijkt dat sommige algoritmen lijden aan een zogeheten 'cold start'-probleem, populariteitsbias, overspecialisatie en homogenisering bias (Stinson, 2022). De manier waarop algoritmen 'omgaan' met deze omstandigheden, verschilt per platform – afhankelijk van de wijze waarop het algoritme geconfigureerd is. Deze omstandigheden kunnen daarmee resulteren in 'bevooroordeelde' algoritmen die niet altijd wenselijke resultaten opleveren.

#### **Problemen of bias waarmee algoritmen te maken kunnen krijgen (Stinson, 2022):**

- *Cold-start probleem*: wanneer content nieuw beschikbaar is op een platform, zijn er nog geen beoordelingen van gebruikers (*cold start*). Het algoritme kan deze content eigenlijk niet aan iemand aanbevelen, omdat aanbevelingen gebaseerd zijn op eerder gedrag van gebruikers – bijvoorbeeld hun beoordelingen van een video. Sommige platforms kiezen er dan voor om nieuwe(re) items bewust bovenaan weer te geven. Keuzes die hierin gemaakt worden door een platform dragen dan bij aan bevooroordeeldheid van het algoritme.
- *Populariteitbias*: dit omvat het idee dat *populaire* content waarschijnlijk vaker aan gebruikers wordt aanbevolen – daar waar deze aanbevelingen op hun beurt zullen resulteren in meer beoordelingen. Gevolg daarvan is dat populaire content almaar toeneemt in populariteit, waardoor er sprake kan zijn van *over-recommendation* van bepaalde content door algoritmen.

- *Overspecialisatie*: dit fenomeen treedt op wanneer algoritmen content aanbevelen die – gekeken naar variatie – veel beperkter is dan het volledige bereik dat een gebruiker zou ‘willen’. Een dergelijk probleem komt voort uit de exclusieve focus van algoritmen op ‘voorspellingsnauwkeurigheid’ in plaats van gebruikerstevredenheid.
- *Homogeniserende bias*: homogenisering verwijst naar het idee dat content naar verloop van tijd eenzijdiger wordt, omdat content die aan gebruikers aanbevolen wordt door algoritmen steeds minder divers en veelzijdig is. Dit relateert aan de populariteitsbias, omdat de populariteitsbias erin resulteert dat individuele items *over-recommended* worden, terwijl de homogeniserende bias betrekking heeft op het totale contentaanbod

### Drie typen van algoritmen

Algoritmen op sociale mediaplatforms kunnen op verschillende manieren bijdragen aan het genereren van relevante content voor gebruikers – in de vorm van zoeksuggesties, aanbevelingssuggesties en suggesties voor advertenties op het platform. Er kunnen daartoe drie typen van algoritmen onderscheiden worden (Whittaker, 2022);

1. *Zoekalgoritmen*: deze algoritmen betrekken zich op de zoekopdrachten van een gebruiker en het genereren van relevante zoeksuggesties. Een voorbeeld hiervan is het automatisch aanvullen van (het begin van) een zoekopdracht die ingevoerd wordt door een gebruiker.
2. *Aanbevelingsalgoritmen*: deze algoritmen genereren content die naar verwachting relevant is of interessant gevonden wordt door een gebruiker. Een voorbeeld hiervan zijn de ‘aanbevolen video’s’ op YouTube of video’s op de ‘For You-page’ van TikTok.
3. *Ad-tech algoritmen*: deze algoritmen hebben als doel het aanbod van advertenties op sociale mediaplatforms te optimaliseren naar de relevantie ervan voor een gebruiker. Vaak gebeurt dit op basis van data over het online gedrag en demografische kenmerken van een gebruiker.

In dit hoofdstuk zullen we focussen op de rol van aanbevelingsalgoritmen.

## 4.2. Aanbevelingsalgoritmen op YouTube en TikTok

Deze paragraaf gaat in op de werking van aanbevelingsalgoritmen op sociale mediaplatforms, waarbij we willen vermijden om in een al te technisch en complex verhaal te geraken. Omwille van de beknoptheid van de literatuurstudie gaan we niet in op de aanbevelingsalgoritmen van alle sociale mediaplatforms, maar beperken we ons tot YouTube en TikTok. Andere mainstream platforms zoals Twitter, Instagram en Facebook werken uiteraard ook met aanbevelingsalgoritmen, waarvan de uitgangspunten op hoofdlijnen vergelijkbaar zijn. De keuze om te focussen op YouTube en TikTok is dat dit platforms zijn waarnaar relatief veel empirisch onderzoek is uitgevoerd (zie: paragraaf 4.3) en vanwege de potentieel grote impact op gebruikers door de dominantie van video’s boven tekst op de platforms. In paragraaf 4.3 gaan we nader in op de werking van aanbevelingsalgoritmen op de platforms en het ontstaan van filterbubbels. Belangrijk is om de beperking te benoemen dat sociale mediaplatforms niet volledig transparant zijn over hun algoritmen, omdat dit zoals eerder aangegeven in zekere zin hun businessmodel is (Llansó et al., 2020). Daarnaast is het goed om te noemen dat platforms hun aanbevelingsalgoritmen regelmatig door ontwikkelen, waardoor de onderstaande informatie de huidige situatie op YouTube en TikTok beschrijft, maar niet noodzakelijk de toekomstige situatie schetst. Centraal in de onderstaande uitwerking staat de vraag: welke factoren beïnvloeden de aanbevelingsalgoritmen van YouTube en TikTok?

### Aanbevelingsalgoritmen op YouTube

YouTube is een sociale mediaplatform dat video’s aanbiedt. Gewoonlijk zoeken en kijken gebruikers daarop naar video’s die hun interesse hebben. Om contentsuggesties te geven die nauw aansluiten bij deze interesses en gebruikers ‘vast te houden’, moet het platform haar gebruikers leren kennen. Als je een kattenliefhebber bent en je krijgt enkel video’s over honden aanbevolen door het algoritme, dan zul je de applicatie snel sluiten. In vergelijking met verschillende andere sociale mediaplatforms, is YouTube betrekkelijk transparant over de

werking van de aanbevelingsalgoritmen (Chi, 2021). Ten behoeve van deze algoritmen, verzamelt YouTube onder meer data over (a) welke video's een gebruiker wel en niet bekijkt, (b) hoeveel tijd er besteed wordt aan het bekijken van een video, (c) welke video's een gebruiker wel en niet leuk vindt of waaronder een reactie geplaatst wordt en (d) in welke video's een gebruiker niet geïnteresseerd is op basis van feedback door een gebruiker (Chi, 2021). De door het algoritme gesuggereerde content wordt vervolgens gepresenteerd op zes categorieën van het platform: (1) Zoekfunctie; (2) Homepagina; (3) Aanbevolen video's; (4) Trending; (5) Abonnementen; en (6) Meldingen. We zoomen nu verder in op de eerste en derde categorie: de Zoekfunctie en Aanbevolen video's, aangezien deze het meest relevant zijn voor de uitvoering van de trials die aanvullend op de literatuurstudie zijn uitgevoerd. Voor meer informatie hierover, zie hoofdstuk 5. Wanneer een gebruiker een zoekterm of het begin daarvan invoert in de zoekfunctie op YouTube, worden er verschillende zoekresultaten gegenereerd. Twee belangrijke factoren die bepalen welke video's bovenaan worden weergegeven in de rangschikking van deze zoekresultaten, zijn de *zoekwoordrelevantie* en *betrokkenheidstatistieken*. Enerzijds houdt het aanbevelingsalgoritme van YouTube op basis van metadata aldus rekening met hoe 'goed' titels, beschrijvingen van de inhoud en de inhoud zelf van video's overeenkomen met de zoekopdrachten (Chi, 2021). Dit zijn de kenmerken van video's zelf. Anderzijds houdt het aanbevelingsalgoritme rekening met (a) het aantal video's dat een gebruiker van een kanaal bekeken heeft; (b) de laatste keer dat deze gebruiker andere video's heeft bekeken over hetzelfde onderwerp als dat in de video; en (c) welke video's de meeste betrokkenheid (likes, reacties, kijktijd etc.) bij een zoekopdracht hebben opgeleverd. Nadat een gebruiker een video heeft aangeklikt, krijgt deze naast de huidige video Aanbevolen video's voorgesteld op de webpagina. Om hiertoe te komen, analyseert YouTube de activiteitengeschiedenis van een gebruiker, wat resulteert in honderden video's die mogelijk interessant zijn voor deze gebruiker. Deze video's worden vervolgens gerangschikt op basis van (a) hoe goed elke video vergelijkbare gebruikers heeft betrokken en tevreden heeft gesteld; (b) hoe vaak elke gebruiker video's van een kanaal of andere video's over eenzelfde onderwerp

bekijkt; en (c) hoe vaak YouTube elke video reeds getoond heeft aan gebruikers (Chi, 2021). Bovendien constateert het platform dat gebruikers geneigd zijn meer inhoud te bekijken wanneer zij video's voorgesteld krijgen van verschillende kanalen. Andere factoren die het algoritme van YouTube in het algemeen beïnvloeden zijn de kijk- en activiteitengeschiedenis van een gebruiker, de kijktijd van een gebruiker, de snelheid waarmee een video populair raakt, hoe vaak een kanaal nieuwe video's op YouTube plaatst, betrokkenheid door gebruikers en negatieve feedback op video's. Hoewel YouTube aldus enigszins transparant over is over het aanbevelingsalgoritme, is het grootste deel van de precieze werking hiervan vooralsnog niet publiekelijk bekend (Geysers, 2022). Wel toont de informatie dat niet alleen het (eerdere) online gedrag van een gebruiker van invloed is op zijn zoekresultaten en aanbevolen video's, maar ook het online gedrag van andere platformgebruikers.

YouTube heeft overigens een belangrijke ontwikkeling doorgevoerd in het algoritme welke relevant is voor dit onderzoek. In 2016 publiceerde YouTube een paper waarin de zogeheten '*deep learning*'-techniek geadresseerd werd (Covington, Adams & Sargin, 2016). Dit is een Artificial Intelligence-techniek die YouTube in staat stelt om op een dieper niveau door zelflerende machinenetwerken meer gepersonaliseerde aanbevelingen aan gebruikers te doen. Gevolg hiervan is dat deze netwerken de doeltreffendheid, maar ook de complexiteit van het aanbevelingsalgoritme vergroten, waardoor de precieze werking ervan voor 'de buitenwereld' ongrijpbaarder is geworden (Geysers, 2022).

Tot slot, het is relevant te wijzen op de nauwe samenwerking van YouTube met Google als het gaat om het effectief en efficiënt in kaart brengen van de interesses van gebruikers. YouTube is een dochterbedrijf van Google LLC.<sup>9</sup> Bij het aanmaken van een YouTube-account wordt de gebruiker gevraagd dit account te koppelen aan een bestaand Google-account om zoekresultaten en contentsuggesties te personaliseren (Google LLC, g.d., (a)). Via deze weg kan YouTube een realistische reconstructie maken van de gebruiker en diens interessegebieden, omdat Google via haar platforms door middel van cookies, pixel tags, browsergeschiedenis een veelheid aan informatie over gebruikers verzamelt die 'uitgewisseld' wordt met YouTube. Om een voorbeeld te noemen; gebruikers die met behulp van de Google zoekmachine eerder zochten naar antwoorden op maatschappelijke vragen, kunnen vervolgens op YouTube video-suggesties ontvangen die daaraan relateren. Daarnaast verzamelt en gebruikt YouTube informatie over het zoekgedrag van gebruikers via verschillende Googleproducten. Dit gaat van zeer voordehand liggende producten, bijvoorbeeld wat gebruikers zoeken via de Google-zoekmachine en de YouTube applicatie, maar ook minder voordehand liggende producten zoals de Google Play Store, of tijdens het surfen naar websites met een Chromebrowser (Google LLC, g.d. (b)). Deze verzamelde informatie wordt door het algoritme van YouTube gebruikt om gebruikers de meeste geschikte video's voor te schotelen in hun zoekresultaten, suggesties op onder meer de homepage en in-app notificaties. Hiermee tracht YouTube de gebruikerservaring en -betrokkenheid te verhogen en gebruikers dus meer tijd op het platform te laten doorbrengen. Of zoals YouTube het zelf verwoordt; *"To recommend you video's we think you want to watch"*. Echter, dit kan mogelijk ook bijdragen aan verhoogde blootstelling van gebruikers aan extreme content op basis van hun surfgedrag en daarmee aan het ontstaan van filterbubbels.

<sup>9</sup> Google LLC is het bedrijf achter (onder andere) de Google-zoekmachine. Sinds 2015 is Google LLC onderdeel van Alphabet Inc.

De gegevens die verkregen worden via het Google-account van een gebruiker en de data die YouTube vervolgens gebruikt om gepersonaliseerde content aan te bevelen, zijn als volgt:

- *Web & App Activity*: Google registreert de activiteiten van de gebruikers op Google-sites en Google-applicaties, inclusief aanvullende informatie als locatie om betere suggesties te geven en meer gepersonaliseerde ervaringen in Google Maps, Google Search en andere diensten van Google. Deze optie kan worden uitgeschakeld.
- *Locatiegeschiedenis*: indien een gebruiker toestemming geeft zal de locatie van het toestel worden bepaald en gebruikt worden om gepersonaliseerde suggesties te geven. Deze informatie wordt ook verzameld zonder gebruik te maken van Google-diensten.
- *YouTube-geschiedenis*: de kijkgeschiedenis en zoekopdrachten op YouTube worden bewaard als een onderdeel van het Google-account dat wordt gelinkt met YouTube. Deze informatie wordt gebruikt om betere aanbevelingen te doen en video's te hervatten waar de gebruiker deze onderbrak. De gebruiker kan kiezen om deze informatie niet te bewaren.

### **Aanbevelingsalgoritmen op TikTok**

TikTok is een platform waarop korte (muziek)video's gemaakt en gedeeld kunnen worden door gebruikers, welke aan hen gepresenteerd worden op de zogeheten 'For You Page'. TikTok is in de afgelopen jaren uitgegroeid tot één van de meest populaire sociale mediaplatforms wereldwijd. Meerdere factoren hebben hieraan bijgedragen, maar het zeer succesvolle aanbevelingsalgoritme van TikTok lijkt daarin een sleutelrol te hebben (Hern, 2022). Om duiding aan te geven aan het aanbevelingsalgoritme, heeft TikTok een pagina gepubliceerd die beschrijft hoe de content op de For You Page wordt samengesteld (TikTok, 2022). Met behulp van deze aanbevelingssystemen worden de interesses van gebruikers zo nauwkeurig mogelijk in kaart gebracht. Dit systeem beveelt inhoud aan door video's te rangschikken op basis van een combinatie van factoren – dit behelst drie categorieën (TikTok, 2022):

- *Gebruikersinteracties*: deze categorie van informatie is gerelateerd aan de interactie tussen de gebruiker met de content op het platform, zoals video's die een gebruiker leuk vindt of deelt, accounts die gevolgd worden, geplaatste reacties en eigen gemaakte inhoud.
- *Video-informatie*: dit omvat aanvullende informatie van video's, waaronder bijschriften, 'sounds' en hashtags – waarmee video's aan elkaar gelinkt kunnen worden.
- *Apparaat- en accountinstellingen*: TikTok bekijkt ook enkele instellingen van het toestel van de gebruiker, waaronder taalvoorkeur, landinstellingen en het type van het toestel. Overigens geeft TikTok wel aan dat deze parameters minder hard doorwegen in het algoritme.

Deze factoren tezamen worden door het aanbevelingsalgoritme afgewogen en beoordeeld, hetgeen bepaalt hoe de video's gerangschikt worden. Hierin wegen sommige factoren dus zwaarder dan andere. Hoe dit precies in zijn werk gaat, wordt door TikTok niet vrijgegeven. Naar het schijnt, is deze informatie dusdanig verfijnd door de AI-systemen van TikTok, dat zelfs het 'algo-team' van het platform hier slechts beperkt zicht op heeft (Hern, 2022). Wel blijkt dat het aanbevelingsalgoritme voortdurend test en experimenteert met video's op de For You Pages van gebruikers en responses hierop registreert, waardoor het zichzelf in staat stelt te leren. TikTok geeft overigens aan, in tegenstelling tot de wijze waarop YouTube dat doet, dat het aantal volgers van een account en het aantal wel/niet goed presterende video's van een account geen directe factoren zijn in het aanbevelingsalgoritme (TikTok, 2022).

Wanneer een nieuw account wordt aangemaakt, is het natuurlijk niet mogelijk om bovengenoemde informatie onmiddellijk te verzamelen en te analyseren. Daarom wordt de gebruiker gevraagd om enkele interesses te selecteren als huisdieren, reizen of sport (TikTok, 2022). Indien de gebruiker deze informatie niet wenst vrij te geven, zal het aanbevelingssysteem een algemene lijst van populaire video's tonen. In dit geval zullen de eerste handelingen (waaronder het bekijken van video's, likes en reacties) gebruikt worden om de content om de tijdlijn van de gebruiker te personaliseren.

Ondanks het succesvolle aanbevelingsalgoritme, geeft TikTok (2022) aan zich ervan bewust te zijn dat dit algoritme filterbubbels in de hand kan werken. Daarover stelt het platform het volgende: *"One of the inherent challenges with recommendation engines is that they can inadvertently limit your experience – what is sometimes referred to as a filter bubble"*. TikTok claimt aldus deze problematiek in gedachten te houden tijdens de ontwikkelen/aanpassen van aanbevelingssystemen. Om een homogene stroom aan video's te beperken, is het aanbevelingsalgoritme daarom zo geconfigureerd dat (a) repetitieve patronen onderbroken worden (geen video's achtereenvolgend met dezelfde 'sound', inhoud of maker); (b) diversiteit opzettelijk wordt ingebracht (hierdoor krijgt een gebruiker óók video's aanbevolen die niet direct relateren aan diens interesses); en (c) de kijkervaring wordt geoptimaliseerd door bepaalde video's te excluseren uit het aanbevelingssysteem (onder meer in het kader van veiligheid) (TikTok, 2022). Om hiermee in een brug te slaan naar de volgende paragraaf, blijkt in de praktijk dat het TikTok-algoritme wel degelijk – al dan niet 'onbedoeld' – kan bijdragen aan een filterbubbel, of anders gezegd, een sociale mediafuik (Little, 2021).



### Internetgebruikers over de dataverzameling van TikTok

In de afgelopen jaren kwam TikTok in opspraak, omdat het platform de privacy van hun gebruikers (heimelijk) zou schenden. Hierbij zou TikTok op een "agressieve" wijze gebruikersdata verzamelen die volgens experts niet relevant zijn voor de optimalisatie van de applicatie. Op macroniveau waarschuwde Christopher Wray (Directeur FBI) dat de Chinese overheid eenvoudig toegang zou kunnen krijgen tot de data van de miljoenen Europese en Amerikaanse gebruikers en daarmee een veiligheidsrisico zou vormen (Macfarlane, 2022). Onderzoekers van Anonymous en Apple gaan nog een stap verder en claimen dat TikTok bewust zijn gebruikers bespioneert, al dan niet in opdracht van de Chinese overheid. In een video van Anonymous stellen zij het volgende: *"TikTok is essentially malware that is targeting children. Don't use TikTok. Don't let your friends and family use it. Delete TikTok now [...] If you know someone that is using it, explain to them that it is essentially malware operated by the Chinese government running a massive spying operation"* (Ondrak, 2020). Verschillende tech-journalisten deden een diepgaande technische analyse en constateerden dat TikTok minder transparant is dan het claimt te zijn. Zo claimt journalist Matthias Eberl (2019) op basis van zijn onderzoek dat de mobiele TikTok applicatie niet alleen toegang zoekt tot de contacten in het gebruikte telefoontoestel van de gebruiker, maar deze informatie ook actief zou gebruiken om nieuwe gebruikers te rekruteren. Ook zou TikTok browseractiviteiten van gebruikers verzamelen. Wall Street Journal deed onderzoek naar de werking van het TikTok-algoritme, nadat verschillende gebruikers zich luidop afvroegen hoe het algoritme zo accuraat hun interesse kon voorspellen (WSJ, 2021). Hiervoor creëerden de journalisten meer dan 100 geautomatiseerde TikTok accounts en bots die honderdduizenden video's lieten afspelen. Zij claimen in hun onderzoek dat reshares, comments et cetera minder invloed hebben dan TikTok steeds beweert en dat de meest bepalende factor de kijktijd van video's is.

Volgens WSJ (2021) kunnen spectaculaire en ophefmakende video's gebruikers sneller verleiden om een video volledig of zelfs nog een keer opnieuw te bekijken. De tijd die besteed wordt bij een video zou veel sterker doorwegen op het algoritme en zo de gebruiker mee de zogenaamde 'rabbit hole' in kunnen te sleuren.

Op 21 maart 2023 verscheen het bericht dat het kabinet rijksambtenaren verbiedt om TikTok te installeren op hun werktelefoons (NOS, 2023), een besluit waar verschillende andere landen Nederland al in voor gingen. Het gebruik werd per direct al ontraden en gevraagd werd om de mobiele TikTok applicatie te verwijderen. In de nabije toekomst worden telefoons die ambtenaren krijgen van de overheid zodanig ingesteld dat alleen vooraf toegestane apps geïnstalleerd mogen worden via beheerssoftware. "Spionagegevoelige apps" zullen niet worden toegestaan, waaronder op TikTok.

### Alternatieve mediaplatforms

Alternatieve (alt-tech) en fringe mediaplatforms werken vrijwel niet met (aanbevelings)algoritmen. Een reden hiervoor is dat deze relatief kleine(re) platforms vaak onvoldoende financieel budget hebben om de soms ingewikkelde algoritmen te ontwikkelen voor het platform. Ook maken de aanbevelingsalgoritmen op grotere sociale mediaplatforms juist deel uit van de kritiek door alternatieve platforms. Alternatieve platforms maken daarin bewust de keuze om niet met dergelijke algoritmen te werken. Hierbij moet worden aangekend dat op een deel van de alternatieve of fringe platforms veel 'extreme' content aanwezig is (in vergelijking met mainstream sociale mediaplatforms). Gebruikers van deze platforms zullen hierdoor direct/sneller in aanraking komen met extreme content.

### 4.3. Sociale mediafuik en (rechts)extremisme?

#### Sociale mediafuiken (filterbubbels) en echokamers

In de afgelopen jaren is er toenemende zorg over de werking van aanbevelingsalgoritmen op sociale mediaplatforms die mogelijkwijs kan bijdragen aan een zogeheten 'sociale mediafuik' voor gebruikers – met alle gevolgen van dien. Wetenschappers, beleidsmakers en wetgevers verwijzen daarmee naar het fenomeen waarbinnen gebruikers als gevolg van aanbevelingsalgoritmen in een 'fuik' van content terecht kunnen komen, hetgeen ook wel een 'filterbubbel' wordt genoemd (Whittaker et al., 2021). Het idee dat hierachter ligt, is dat – zoals aangegeven is in paragraaf 4.1 – algoritmen content aanbevelen waarvan 'berekend' is dat een gebruiker dit interessant zal vinden. Vaak is deze content daarmee in overeenstemming met en afgestemd op eerder online gedrag van een gebruiker, waardoor voorgestelde content in hetzelfde 'straatje' past als reeds geconsumeerde content. Content die niet relevant wordt geacht voor een gebruiker, wordt door het aanbevelingsalgoritme niet of minder aan hem getoond. Dit kan een eenzijdig contentaanbod opleveren. Als gevolg daarvan ontstaat er een *"personalised universe of information for each user, which may contain increasingly less diverse viewpoints and therefore increase polarisation within society"*, omdat invalshoeken meer verdeeld raken en er minder ruimte is voor andere invalshoeken (Whittaker et al., 2021; Pariser, 2011). Het wegfilteren van andere meningen, attitudes en houdingen veroorzaakt een filterbubbel met eenzijdige content die gebruikers in zekere zin de mogelijkheid ontnemt om een mening of kritische attitude te ontwikkelen op basis van objectieve en veelzijdige informatie (Mediawijsheid, webpagina). Een dergelijke bubbel wordt dan ook gedefinieerd als het verschijnsel waarbinnen aanbevelingsalgoritmen functioneren als "autopropaganda" door onzichtbaar te bepalen welke content een gebruiker wel en niet te zien krijgt, waardoor ideeën worden gepromoot die in overeenstemming zijn met die van een gebruiker en daarmee de *confirmation bias* van die gebruiker versterken (Pariser, 2011; Whittaker et al., 2021). Gekeken naar empirische studies concluderen Whittaker et al. (2021) daarentegen dat er slechts beperkte bewijzen zijn voor de omvang en schadelijkheid

van filterbubbels; iets dat mede veroorzaakt wordt door de verschillende en soms uiteenlopende definities van filterbubbels en daaraan gerelateerde fenomenen (Zuiderveen-Borgesius et al., 2016).

Ter onderscheid van filterbubbels wordt in dit verband de notie van online echokamers onderscheiden. Echokamers worden evenals filterbubbels zowel in de praktijk als in de wetenschappelijke literatuur niet eenduidig gedefinieerd en er is soms sprake van een grijs gebied. In het kader van deze literatuurstudie en de trials is het relevant het verschil tussen deze noties nader te duiden. In lijn met Whittaker et al. (2021) en Zuiderveen-Borgesius et al. (2016) maken we in deze literatuurstudie een onderscheid tussen: (a) gepersonaliseerde content die *self-selected* is door een platformgebruiker, doordat deze er bewust voor kiest om gelijkgestemde meningen en invalshoeken op te zoeken en daarmee terecht kan komen in online groepen die uitsluitend bestaan uit gelijkgestemden – een zogeheten echokamer; en (b) gepersonaliseerde content die *pre-selected* is door het aanbevelingsalgoritme op sociale mediaplatforms zonder dat hieraan een 'bewuste keuze', invoer, kennis of toestemming van een platformgebruiker aan vooraf is gegaan – dit wordt aangeduid als een filterbubbel. Zowel in echokamers als in filterbubbels vindt het eerdergenoemde mechanisme van *echoing* plaats (Mølmen en Ravndal, 2021). Echoing houdt in dat de constante interactie met gelijkgestemde invalshoeken en denkbeelden kan leiden tot het ontwikkelen en versterken van extreme/radicale overtuigingen en het legitimeren van geweld door personen (zie: hoofdstuk 2).

Verschillende gevolgen van filterbubbels op sociale mediaplatforms worden als zorgelijk beschouwd. Daartoe onderscheiden Whittaker en collega's (2021) ten minste vier omstandigheden; (a) algoritmen 'dragen' vooroordelen van hun menselijke ontwerpers, hetgeen verwijst naar het eerdergenoemde gegeven dat algoritmen niet neutraal zijn (Bozdog, 2013) en daarmee een reeds 'gekleurde' filterbubbel kunnen creëren; (b) het onderscheid tussen gebruikerstevredenheid en gebruikersbehoud vertroebelt (Seaver, 2018), waardoor gebruikers mogelijk onjuist worden geïnformeerd over belangrijke sociale gebeurtenissen

(Napoli, 2015); (c) gebruikers zijn zich niet altijd bewust van deze filterbubbels en hoe zij mede daardoor beperkt kunnen worden in hun informatievoorziening (Eslami, 2015; Bucher, 2017); en (d) de 'makers' van aanbevelingsalgoritmen begrijpen niet altijd hoe het algoritme precies in elkaar zit (Napoli, 2014; DeVito, 2016), wat mede wordt versterkt door zelflerende artificiële intelligentie (AI)-technieken. Een gevolg van de toepassing van AI kan zijn dat het tegengaan van filterbubbels bemoeilijkt wordt. Whittaker et al. (2021) relateren deze zorgen aan de context van extremisme en stellen dat *"each of these concerns may cause alarm as they posit a situation in which users may become insulated with confirming information in an opaque system"*.

### **Amplificatie door algoritmen**

Naast het gegeven dat filterbubbels kunnen bijdragen aan het versterken van extreme of radicale invalshoeken, zijn er substantiële zorgen over de amplificerende werking van aanbevelingsalgoritmen op sociale mediaplatforms (Whittaker et al., 2021; Keller, 2021; GIFCT, 2021; H.M. Government, 2019; Raad van de Europese Unie, 2020). In dit verband verwijst amplificatie naar het versterkende effect van aanbevelingsalgoritmen op de content die zij aanbevelen aan gebruikers. Dit effect heeft zowel gunstige als ongunstige uitkomsten (Keller, 2021). Positieve uitkomsten vallen samen met de functionaliteit van algoritmen: het vinden van de juiste informatie op het internet die relevant is voor gebruikers en aansluit bij diens interesses. Negatieve uitkomsten van amplificatie behelzen het idee dat aanbevelingsalgoritmen content kunnen genereren die (steeds) extremer en/of niche van aard is (Keller, 2021). Dit sluit aan bij het eerdergenoemde punt in paragraaf 4.1 over de aantrekkelijkheid van controversiële, provocatieve en extremistische content. Dit verhoogt namelijk de betrokkenheid van gebruikers die hier interesse in hebben getoond, met als gevolg dat algoritmische systemen 'leren' om deze content te prioriteren (Llansó et al., 2020). Door amplificatie wordt een gebruiker niet alleen bevestigd in zijn eigen ideeën (*confirmation bias*), maar krijgen deze ideeën mogelijk een extremer karakter dat gevoed wordt door 'nieuwe' aanbevolen content. Bovendien schuilt hierin het risico op het onbedoeld verspreiden van

misleidende informatie of desinformatie onder een breder publiek, omdat algoritmen niet altijd in staat zijn om deze content te filteren uit het aanbevelingssysteem; iets waar bijvoorbeeld het platform TikTok zich van bewust is en probeert tegen te gaan door aanbevelingsalgoritmen daarop te configureren (TikTok, 2022).

### **Empirische studies**

Om een antwoord te geven op de vraag of en in hoeverre aanbevelingsalgoritmen op sociale mediaplatforms daadwerkelijk (a) sociale mediafinken of filterbubbels creëren die; (b) een amplificerende werking hebben als het gaat om extreme content en daarmee; (c) mogelijk kunnen bijdragen aan rechtsextremistische radicalisering, zijn er verschillende empirische studies die daarin richting geven. De organisatie Global Internet Forum to Counter Terrorism (GIFCT, 2021) heeft een recente reviewstudie uitgevoerd naar elf studies die interacties hebben onderzocht tussen algoritmen en aanbevolen content die relateert aan terrorisme en/of gewelddadig extremisme (jihadistisch extremisme én rechtsextremisme). Op basis hiervan concludeert GIFCT (2021) dat aanbevelingsalgoritmen extremistische content lijken te amplificeren voor zowel het jihadistisch extremisme als het rechtsextremisme. Een aantal van de studies uit de review richten zich specifiek op rechtsextremistische content, het merendeel hiervan is uitgevoerd op YouTube.

De belangrijkste resultaten van deze studies zijn verwerkt in onderstaande tabel (inclusief recentere studies aangevuld door onszelf):

Studie	Platform	Kernbevindingen
O'Callaghan et al. (2015)	YouTube	Conclusie van deze studie is dat aanbevelingsalgoritmen meer rechtsextremistische content aanbevelen aan gebruikers; en dat gebruikers in een ideologische bubbel terecht kunnen komen waarin zij uitgesloten zijn van inhoud die niet aansluit bij hun belief system – i.e. een filterbubbel/sociale mediafuiik.
Ribeiro et al. (2019)	YouTube	In deze studie komt naar voren dat YouTube's aanbevelingsalgoritme rechtsextremistische 'Alt-lite' en 'Intellectual Dark Web' content aanbeveelt aan gebruikers; eenmaal in deze gemeenschappen terechtgekomen, is het tevens gemakkelijk 'Alt-Right' content te vinden via door YouTube aanbevolen kanalen. Deze studie ondersteunt het idee van een 'radicalization pipeline' uit Schmitt et al. (2018).
Ledwich & Zaitsev (2019)	YouTube	Deze studie vindt bewijs voor het gegeven dat YouTube's aanbevelingsalgoritme gebruikers juist actief ontmoedigen om extremistische content te bekijken – waarmee populaire 'radicaliseringsclaims' weerlegd worden. YouTube leidt gebruikers naar reguliere (in plaats van extreme) kanalen.
Reed et al. (2019)	YouTube; Reddit; Gab	YouTube: conclusie is dat extreme en fringe content meer aanbevolen wordt en hoger gerangschikt wordt in aanbevolen content. Reddit en Gab: extreme content wordt niet gepromoot via aanbevelingen die het platform doet.
Gaudette et al. (2020)	Reddit	Deze studie heeft gekeken naar zogeheten upvoting en downvoting algoritmen op Reddit. <sup>10</sup> De conclusie is dat het upvoting algoritme een sleutelrol heeft in het faciliteren van extremistische content.
Whittaker et al. (2021)	YouTube; Reddit; Gab	YouTube: conclusie is dat YouTube meer rechtsextremistische en marginale content promoot/aanbeveelt nadat er interactie heeft plaatsgevonden met rechtsextremistische content door een gebruiker. Reddit en Gab: in tegenstelling tot YouTube worden hier geen bewijzen gevonden voor het aanbevelen van meer extreme content.
Hosseinmardi et al. (2020)	YouTube	Conclusie van deze studie is dat ontwikkelingspaden naar rechtsextremistische content op YouTube erg divers zijn; slechts een deel kan toegeschreven worden aan de werking van aanbevelingsalgoritmen. Eigen voorkeuren van gebruikers lijken een grotere rol te spelen dan amplificatie naar extreme content over een langere periode van tijd – bijvoorbeeld doordat gebruikers offline of via andere online bronnen extremistische content consumeren. In dat geval is er geen sprake van een sociale mediafuiik.
Media Matters (Little, 2021)	TikTok	Deze studie concludeerde dat TikTok's aanbevelingsalgoritme bijdraagt aan het 'pushen' van gebruikers richting extreemrechtse content; enerzijds door aan te bevelen op de For You Page, anderzijds door gebruikers aan te sporen far-right accounts te volgen – hetgeen aanzet lijkt te geven voor een filterbubbel.
Chen et al. (2022)	YouTube	Blootstelling aan alternatieve en extremistische video's is op YouTube sterk geconcentreerd bij slechts een kleine groep mensen. Doorgaans abonneren deze gebruikers zich op dergelijke kanalen en volgen vaker externe links naar deze kanalen – met als gevolg dat YouTube's algoritme hen vaker extreme content aanbeveelt. Gebruikers die daarentegen niet geabonneerd zijn op extremistische kanalen lijken zelden alternatieve en extremistische video's of kanalen aanbevelen te krijgen, waardoor deze studie geen bewijs vindt voor een filterbubbel of sociale mediafuiik.
Brown et al. (2022)	YouTube	In tegenstelling tot veel voorgaande studies die gebruikmaakten van anonieme gegevensverzameling die geen rekening houden met de gepersonaliseerde aard van aanbevelingsalgoritmen, is deze studie uitgevoerd met een steekproef van echte personen. De studie vindt echter geen overtuigend bewijs voor het ontstaan van rabbit holes die gebruikers naar extreme content leiden. Zichtbaar is zelfs dat het aanbevelingsalgoritme van YouTube gebruikers over een lange tijd juist minder ideologische content aanbeveelt en hen naar 'gemiddelde' ruimten brengt in plaats van naar extremistische randen.

<sup>10</sup> Gebruikers op Reddit kunnen berichten of links upvoten (verwachting dat anderen de inhoud hiervan ook leuk zullen vinden) en downvoten (verwachting dat anderen de inhoud hiervan niet leuk zullen vinden). Berichten die veel 'ge-upvoted' worden, worden meer bovenaan op de site weergegeven aan gebruikers bij een topic.

Hoewel de resultaten van de studies niet geheel eenduidig zijn, laat het merendeel hiervan zien dat aanbevelingsalgoritmen op sociale mediaplatforms gebruikers aanbevelingen geven in de richting van meer extreme content. Op deze bevindingen zijn echter een aantal kritische kanttekeningen te plaatsen:

- Een belangrijke nuance vanuit de bevinding van Hosseinmardi et al. (2020) is dat dat ontwikkelingspaden naar rechtsextremistische content op YouTube erg divers zijn; en dat *slechts een deel hiervan* toegeschreven kan worden aan aanbevelingen door aanbevelingsalgoritmen. Er wordt door hen namelijk geen bewijs gevonden voor meer extreme content (amplificatie) over een langere periode van tijd – eigen voorkeuren van gebruikers lijken hierin een grotere rol te spelen: onder meer doordat gebruikers offline of via andere online bronnen extremistische content consumeren.
- Uit de studie van Chen et al. (2022) komt bovendien de nuance naar voren dat de blootstelling aan alternatieve en extremistische video's op YouTube sterk geconcentreerd is bij slechts een kleine groep mensen. Doorgaans abonneren deze gebruikers zich op dergelijke kanalen en volgen vaker externe links naar deze kanalen – met als gevolg dat YouTube's aanbevelingsalgoritme hen vaker extreme content aanbeveelt. In vergelijking met het abonneren op kanalen en het volgen van externe links blijkt het relatieve belang van algoritmen in het aanbevelen van extreme content beperkt. Tevens blijkt dat gebruikers die niet geabonneerd zijn op extremistische kanalen zelden alternatieve en extremistische video's of kanalen aanbevolen krijgen, waardoor deze studie geen bewijs vindt voor een filterbubbel of sociale mediafuik (Chen et al, 2022). Anders gezegd, gebruikers die er al voor kozen om extreme video's te kijken, krijgen mogelijk vaker extreme content aanbevolen door het aanbevelingsalgoritme, maar gebruikers die hier geen specifieke interesse in tonen, zullen hier nauwelijks of niet mee geconfronteerd worden.
- Tot slot, Broderick (2022) stelt dat mainstream sociale mediaplatforms veeleer fungeren als 'gateway' of doorgang naar alternatieve platforms als 4chan, Gab of 8kun waarop geen algoritmisch gedreven aanbod van content plaatsvindt. Het consumeren van extremistische content op

dergelijke platforms kan dan bijdragen aan radicalisering, zonder dat aanbevelingsalgoritmen of sociale mediafuiken daarin een rol hoeven te spelen.

Er zijn ook enkele methodische nuances die gemaakt moeten worden bij het bovenstaande overzicht van studies:

- Allereerst is het belangrijk te noemen dat deze studies – afgezien van Chen et al. (2022) en Brown et al. (2022) – zijn uitgevoerd voor of ten tijde van relevante aanscherpingen door YouTube om de verspreiding van extremistische content tegen te gaan (GIFCT, 2021). De effecten van deze aanpassingen komen daarom mogelijk niet (volledig) terug in de uitkomsten eerder uitgevoerde studies. Het ontbreken van overtuigend bewijs voor filterbubbels of mediafuiken in de recentere studies (zoals Hosseinmardi et al. (2020); Chen et al. (2022); Brown et al. (2022)) kan er daarom op wijzen dat de aanpassingen in algoritmen of verbeterde contentmoderatie door platforms tot gevolg heeft dat filterbubbels minder 'ontstaan' als gevolg van het aanbevelingsalgoritme.
- Een andere nuancering is dat er rekening gehouden dient te worden met het 'type' platform: mainstream of alt-tech. Zoals in Hoofdstuk 3 is beschreven, is het moderatiebeleid van content op alternatieve platforms namelijk beperkter dan op mainstream mediaplatforms. Als gevolg hiervan blijft extremistische content meer aanwezig op het platform: het wordt immers niet gemodereerd. Dit suggereert dat de eigen keuze door een gebruiker om te kiezen voor een dergelijk platform met extremistische content een belangrijkere rol heeft dan de aanbevelingsalgoritmen in het aanbevelingssysteem (GIFCT, 2021), zoals door verschillende auteurs is aangegeven. Bovendien blijkt dat passieve consumptie van en betrokkenheid bij online content eveneens van invloed is op gebruikers, waardoor het risico op radicalisering voornamelijk blijft bestaan (Rowa, 2020).
- Een laatste methodologische kanttekening is dat de genoemde studies veelal gebruikmaken van de Application Programming Interface (API) van YouTube. Dit bootst de gepersonaliseerde interactie tussen een gebruiker

en het platform (zoals zoekgedrag, liken en resharing door gebruikers) niet goed na (Whittaker et al., 2022; Brown et al. (2022)). Iedere persoon is immers uniek, met eigen interesses en (online) zoekgedrag. Het onderzoek naar de relatie tussen aanbevelingsalgoritmen en het bestaan van sociale mediafuiken/filterbubbels is een onderzoeksveld in ontwikkeling waarin nog methodische verbeteringen mogelijk zijn. Het meer rekening houden met gepersonaliseerd zoekgedrag is één van deze mogelijke verbeteringen. In Hoofdstuk 5 komen we hier verder over te spreken, waarin we de trials op basis van zoekgedrag beschrijven die wij hebben uitgevoerd op YouTube en TikTok.

#### Samenvatting Hoofdstuk 4

Dit hoofdstuk beantwoordt onderzoeksvraag 4: In welke mate zijn er aanwijzingen dat het algoritme gestuurde aanbod van rechtsextremistische content op sociale mediaplatforms kan leiden tot een rechtsextremistische 'sociale mediafuike'? Is hierbij een verschil tussen de diverse sociale mediaplatforms? In dit hoofdstuk komt het volgende aan bod:

- Introductie over de werking van algoritmen.
- Aanbevelingsalgoritmen op YouTube en TikTok.
- Sociale mediafuiken, filterbubbels, echokamers, amplificatie en empirische studies in het kader van rechtsextremistische content.

De resultaten van empirische studies zijn niet geheel eenduidig. Over het algemeen laten vooral de eerder uitgevoerde studies zien dat aanbevelingsalgoritmen op YouTube en TikTok gebruikers aanbevelingen geven in de richting van meer extreme content. Vermoedens van eventuele filterbubbels of mediafuiken werden daarmee ondersteund. Latere studies laten echter zien dat ontwikkelingspaden naar rechtsextremistische content op sociale media platforms zoals YouTube en TikTok erg divers zijn; en dat slechts een deel hiervan toegeschreven kan worden aan aanbevelingen door algoritmen.

Gedrag / eigen keuzen van gebruikers zijn zeker zo belangrijk. Zo lijkt blootstelling aan alternatieve en extremistische video's op YouTube sterk geconcentreerd bij slechts een kleine groep mensen. Gebruikers die er al voor kozen om extreme video's te kijken, krijgen mogelijk vaker extreme content aanbevolen door het algoritme, maar gebruikers die hier geen specifieke interesse in tonen, zullen hier nauwelijks of niet mee geconfronteerd worden. Resultaten van latere (recentere) studies laten tevens zien dat van sociale mediafuiken naar extremistische content op platforms minder sprake lijkt te zijn. Interpretatie van deze recente onderzoeksresultaten suggereert dat recente aanscherpingen in het (moderatie)beleid van extremistische content door sociale mediaplatforms hun vruchten lijken af te werpen. Op alternatieve platforms als Gab is dat niet het geval omdat dergelijke algoritmen daar niet of minder aanwezig zijn. Tot slot blijkt uit de review van literatuur dat mainstream sociale mediaplatforms kunnen fungeren als 'gateway' naar alternatieve platforms als 4chan, Gab of 8kun waarop geen algoritmisch gedreven aanbod van content plaatsvindt. Deze extremistische websites kunnen evenzo bijdragen aan radicalisering – zonder dat algoritmen of een sociale mediafuike daarin een rol spelen.

# 5 Trials op YouTube en TikTok

Onderzoeksvragen die worden beantwoord in dit hoofdstuk:

1. In hoeverre leiden aanbevelingsalgoritmen op sociale platforms, op basis van gepersonaliseerd zoekgedrag door gebruikers (persona's / invoer van zoekwoorden) tot suggesties voor extreme content? In hoeverre is er sprake van een relatie tussen het zoekgedrag van gebruikers op Google en aanbevelingen naar meer extreme content die gebruikers krijgen op YouTube? In hoeverre verschilt dit met de aanbevelingen op TikTok; een platform dat niet beheerd wordt door Google?

In dit hoofdstuk komt het volgende aan bod:

- Introductie (Paragraaf 5.1).
- Uitvoering en resultaten trials op YouTube (Paragraaf 5.2).
- Uitvoering en resultaten trials op TikTok (Paragraaf 5.3)

## 5.1. Introductie

In hoofdstuk 4 hebben we geconcludeerd dat eerdere empirische studies naar de interactie tussen aanbevelingsalgoritmen en extremistische content, hoewel niet eenduidig, over het algemeen laten zien dat algoritmen op YouTube en TikTok gebruikers aanbevelingen geven in de richting van meer extreme content. Met als belangrijke nuance dat ontwikkelingspaden naar rechtsextremistische content erg divers zijn en dat slechts een deel hiervan toegeschreven kan worden aan aanbevelingen door algoritmen. Resultaten van recentere studies laten echter zien dat van sociale mediafijnen naar extremistische content op deze platforms inmiddels minder sprake lijkt te zijn. Deze recente onderzoeksresultaten suggereren dat recente aanscherpingen door sociale mediaplatforms in het (moderatie)beleid van extremistische content effect lijken te hebben. Een methodologische kanttekening bij de in hoofdstuk 4 besproken empirische studies, is dat in deze studies veelal gebruik wordt gemaakt van de Application Programming Interface (API) van YouTube<sup>11</sup>. Deze manier van gegevens verzamelen bootst relaties tussen gebruikers en het platform niet goed na, zoals Whittaker et al. (2022) en Brown et al. (2022) constateren, omdat de gepersonaliseerde aard van de interactie tussen een gebruiker en content (zoals zoekgedrag, liken en resharing door gebruikers) niet wordt meegenomen. Deze studies houden onvoldoende rekening met de unieke interesses en actief (online) zoekgedrag van personen ('personalisatie') welke belangrijke aspecten zijn voor het kunnen ontstaan van filterbubbels.

<sup>11</sup> Een API (Application Programming Interface) is een set van regels waarmee verschillende softwareapplicaties met elkaar kunnen communiceren en gegevens uitwisselen. Het stelt ontwikkelaars in staat om gemakkelijk toegang te krijgen tot de functionaliteit van andere applicaties, waardoor het mogelijk is om bijvoorbeeld gegevens van de ene app naar de andere te verzenden.

Om beter rekening te houden met de gepersonaliseerde aard van aanbevelingsalgoritmen hebben we ter aanvulling op de literatuurstudie een empirische studie ('trials') opgezet en uitgevoerd op twee grote mainstream sociale mediaplatforms: YouTube en TikTok. De 'trials' beogen tevens een aanvulling te zijn op de studie van Chen et al. (2022) waarin de onderzoekers aantoonde dat eigen keuzen van gebruikers op YouTube een grotere rol speelt bij de blootstelling aan extremistische content dan aanbevelingsalgoritmen. Deze studie vond geen bewijs voor het bestaan van een sociale mediafuik.

In de trials onderzoeken we in hoeverre algoritmen op basis van gepersonaliseerd zoekgedrag door gebruikers leiden tot suggesties voor extremere content. In de trials zal specifiek worden gekeken naar aanbevelingen voor video's op YouTube en TikTok op basis van de invoer van zoekwoorden. *Nota bene*: bestaand onderzoek naar filterbubbels heeft zich vooral gericht op 'doorklikken' op video's en het bestuderen van content welke vervolgens door algoritmen worden aanbevolen. In de trials onderzoeken we welke content door algoritmen worden aanbevolen op basis van de invoer van zoekwoorden door gebruikers. Om ethische woorden zien we in de trials af van likes, reshares en clicks op content. Om in de trials meer gepersonaliseerd zoekgedrag te kunnen simuleren hebben we twee zogeheten 'persona's' opgezet. Een 'persona' is een begrip afkomstig uit de marketing en kan worden gedefinieerd als een fictief persoon dat geldt als een verpersoonlijking van een 'abstracte' doelgroep. Persona's worden gemodelleerd naar de doelgroepen erachter: het zijn typische representaties ervan. Ten behoeve van de trials representeren de persona's personen met online zoekgedrag binnen twee van de subculturen die zijn beschreven in paragraaf 3.7, te weten een persona dat aansluit bij de manosphere (waarbinnen een sterke nadruk ligt op masculiniteit) en een persona dat past binnen de Vikingcultuur. Paragraaf 5.2 gaat hier nader op in.

### Onderzoeksvragen ten aanzien van de trials

De eerste onderzoeksvraag voor de uitvoering van de trials luidt als volgt:

- 1a. In hoeverre leiden aanbevelingsalgoritmen op YouTube, op basis van gepersonaliseerd zoekgedrag door gebruikers (persona's / invoer van zoekwoorden) tot suggesties voor extreme content?
- 1b. In hoeverre leiden aanbevelingsalgoritmen op TikTok, op basis van gepersonaliseerd zoekgedrag door gebruikers (persona's / invoer van zoekwoorden) tot suggesties voor extreme content?

Een tweede factor die we in de trials meenemen, is de mogelijke invloed die zoekgedrag op zoekmachine Google heeft op de aanbevelingsalgoritmen op YouTube (onderzoeksvraag 1b). YouTube is een website van YouTube LLC, een dochteronderneming van Google LLC (dat vervolgens weer onderdeel is van Alphabet Inc.). Zoals toegelicht is in Paragraaf 4.2 wordt een gebruiker bij het aanmaken van een YouTube-account aangespoord om dit nieuwe account te koppelen aan een bestaand Google-account om zoekresultaten en suggesties te personaliseren. Omdat deze link tussen zoekresultaten op Google en aanbevelingen op YouTube in de eerder besproken empirische studies nog niet is gelegd, willen we dit aanvullend onderzoeken in de trials. De tweede onderzoeksvraag in de trials luidt daarom als volgt:

- 2a. In hoeverre is er sprake van een relatie tussen het zoekgedrag van gebruikers op Google en aanbevelingen naar meer extreme content die gebruikers krijgen op YouTube?
- 2b. In hoeverre verschilt dit met de aanbevelingen op TikTok; een platform dat niet beheerd wordt door Google?



## Keuze voor YouTube en TikTok

Gekozen is om de trials uit te voeren op de mainstream sociale mediaplatforms YouTube en TikTok. De keuze om te focussen op deze twee platforms is dat dit platforms zijn waarnaar relatief veel eerder empirisch onderzoek is uitgevoerd (zie: Paragraaf 4.3) en vanwege de potentieel grote impact op gebruikers door de dominantie op de platforms van video's boven tekst.

## Persona's en 'prompting'

In de trials beogen wij om het surfgedrag van gebruikers natuurlijker na te bootsen, in vergelijking met de API-techniek. Via deze weg kan er meer rekening gehouden worden met de gepersonaliseerde aard van aanbevelingsalgoritmen en eventuele filterbubbels die daaruit kunnen ontstaan. Daarom hebben we voorafgaand aan de trials profielen opgesteld van fictieve personen (persona's). De profielen van deze persona's leiden we af van subculturen die op zichzelf niet radicaal of extremistisch zijn, maar waarbinnen uit andere onderzoeken bekend is dat personen wel eerder in aanraking kunnen komen met rechtsextremistische narratieven door bijvoorbeeld overlap in ideologische uitgangspunten (o.a. Grant & MacDonald, 2018; Zwissler, 2018). Deze subculturen kunnen dan fungeren als zogeheten 'gateway' naar meer rechtsextremistische content. Dit is toegelicht in Paragraaf 3.7. In de trials zullen zoekopdrachten worden toegepast die passen bij het interessegebied van het persona, dat wil zeggen het zoeken op onderwerpen die gangbaar zijn binnen de betreffende subculturen. Hoewel we hiermee een natuurlijker surfgedrag simuleren dan bij API's, blijft ook dit tot op zekere hoogte kunstmatig aangezien we geen content liken of delen (sharen). Dit doen we om ethische redenen; we willen immers niet stimuleren dat de trials in het verband van dit onderzoek zelf bijdragen aan algoritmische amplificatie van geplaatste content op sociale mediaplatforms. Om die reden spreken we liever van 'prompting': een term die staat voor het 'uitlokken' van algoritmen door zoekgedrag (in dit geval het invoeren van zoekresultaten op YouTube en TikTok). In de trials is gebruikgemaakt van twee persona's:

1. Het persona dat is gebruikt in de trials op YouTube, is een persoon die zich interesseert in de Vikingcultuur binnen het (neo-) paganisme/heidendom. Zoals aangegeven in paragraaf 3.7 vallen onder het neo-heidendom religies die de natuur als heilig en beziel ervaren, waaronder Wicca en de Noorse religies Odinisme en Asatru. De Vikingcultuur is hieraan nauw verwant. De neo-paganistische gemeenschap is grotendeels progressief, maar oefent aan de randen ook aantrekkingskracht uit op sommige rechtsextremistische groepen (De Zutter, 2010). Deze aantrekking ligt vooral in de 'eigen' voorchristelijke cultuur, vrij van diversiteit en migratie (De Wit, 2000). De focus ligt hierbij onder andere op de pseudowetenschappelijke theorie dat Europa oorspronkelijk enkel door 'pure blanken' werd bevolkt en de bijhorende wens om dit te herstellen en behouden. Sommige aanhangers van Noorse tradities keren zich tegen het multiculturalisme en nemen hierbij soms rechtsextremistische standpunten in (De Zutter, 2010).
2. Het tweede persona dat wij meenamen in de trials, is een persoon die informatie zoekt over mannelijkheid en masculiniteit. Zoals toegelicht is in paragraaf 3.7 kunnen personen hierbij in aanraking komen met websites en online influencers die raken aan de 'manosfeer'-subcultuur. Deze manosfeer behelst een gevarieerde verzameling van online 'mannen-ondersteuningsgemeenschappen' die zijn ontstaan als reactie op onder meer het feminisme, vrouwelijk empowerment, maatschappelijke veranderingen en de vervreemdende effecten van het neoliberalisme (Rich & Buljaka, 2023). Binnen deze subcultuur wordt aan vrouwen en feministen de schuld gegeven van talloze problemen in de samenleving (zie daarvoor ook: Paragraaf 3.2). De manosfeer spreekt daarom voornamelijk een mannelijk publiek aan dat zich onder andere geconfronteerd ziet met romantische afwijzing, vervreemding, economisch falen, eenzaamheid en een sombere visie op de toekomst (Rich & Buljaka, 2023). Het opnieuw opkrikken van 'mannelijkheid' kan de maatschappij terugbrengen naar een geïdealiseerd 'vroeger', met een door mannen gedomineerde maatschappij en slechts zeer uitzonderlijke migratie van buiten Europa. Om dit ideaal te realiseren, wordt onder meer fysieke fitheid en paraatheid gepromoot.

Veel manosphere-gemeenschappen hebben hun eigen websites, maar zijn ook relatief gemakkelijk vindbaar via grote sociale mediaplatforms als Instagram, Facebook en Twitter. Ook op YouTube en TikTok kunnen jongeren geconfronteerd worden met seksistische en antifeministische content. Hoewel masculiniteit op zichzelf geen politieke kleur draagt, constateren onderzoekers dat masculiene onderwerpen steeds belangrijker zijn geworden binnen rechtsextremistische propaganda (Jones et al., 2019; Lacoviello et al., 2021; Grant & MacDonald, 2018); en kan aldus bijdragen aan radicalisering bij (met name) jongemannen in het geval van herhaaldelijke blootstelling aan dergelijke online content. Dit persona is onderzocht op TikTok.

In de trials gaan deze persona's online en de onderzoekers trachten vervolgens een zo natuurlijk mogelijk surfgedrag te simuleren op basis van zoekwoorden die typerend zijn voor de betreffende subculturen. Vervolgens wordt gekeken hoe dit online gedrag van invloed is op welke soort berichten getoond worden op de tijdlijnen in YouTube en op TikTok. Paragrafen 5.2 en 5.3 tonen overzichten van toegepaste zoekwoorden voor de beide persona's.

### **Contentanalyse van suggesties en video's**

In de trials bekijken de onderzoekers welke suggesties door Google, YouTube en TikTok worden teruggegeven op basis van de ingevoerde zoekopdrachten. Daarnaast worden er relevante video's afgespeeld en bekeken. Voor de door het platform aanbevolen suggesties wordt allereerst vastgesteld of deze wel of geen inhoudelijke relevantie hebben voor het interessegebied (bijbehorende subcultuur en de thema's van de ingevoerde zoekopdrachten) van het desbetreffende persona. Relevante video's worden door de onderzoekers vervolgens bekeken en beoordeeld op inhoud: neutraal, borderline of extremistisch. De in de volgende alinea gegeven definities waren leidend bij de beoordeling of content als neutraal óf borderline óf extreem moesten worden beoordeeld. De beoordeling hiervan is tot stand gekomen op basis van intersubjectiviteit, dat wil zeggen overeenstemming tussen ten minste twee beoordelaars.

### *Borderline content*

Sociale mediaplatforms als YouTube en TikTok hebben beleid om extremistische – en daarmee vaak illegale – content van het platform te verwijderen, dit is aangescherpt rond 2019-2020 (zie: paragraaf 7.2 voor meer informatie hierover). Desalniettemin blijft er een grijs gebied van content over: content die ervaren kan worden als onwenselijk, maar die niet op grond van wettelijke kaders verwijderd kan worden. In dit verband wordt er gesproken van *borderline content* (Heldt, 2020). Naast wettelijke kaders bepalen ook sociale normen (impliciet en cultureel verondersteld) wat er wel en niet gezegd kan worden. Dit maakt de beoordeling van content tot op zekere hoogte subjectief en daarmee relatief (Heldt, 2020). De notie van *borderline* verwijst in feite naar content die niet te classificeren is als neutraal, maar ook niet als extremistisch, illegaal en/of niet-toegestane content. YouTube definieert *borderline content* als *"content die in de buurt komt van het schenden van de community-richtlijnen, maar de grens hiervan niet helemaal overschrijdt"*. Ook op Europees niveau wordt geworsteld met het afbakenen/definiëren van *borderline content*. De EU Counter-Terrorisme Coordinator (2020) geeft de volgende definitie: *"'Borderline content' is toegestane content, omdat het voldoet aan de servicevoorwaarden van de sociale media platforms, maar komt in de buurt van verboden content. Borderline-inhoud zou legale schadelijke inhoud omvatten. Vanuit juridisch oogpunt kan 'content op het randje' worden opgevat als legale content waarvan de wettigheid onduidelijk of dubbelzinnig blijft (bijv. geweld versus aanzetten tot geweld"*. Relevant om hier te noemen, is dat (met name grote) sociale mediaplatforms voortdurend bezig zijn met het ontwikkelen, vernieuwen en verbeteren van hun aanpak van *borderline content* op het platform. In de afgelopen jaren zijn dan ook verschillende ontwikkelingen doorgevoerd in het moderatiebeleid op deze platforms (zie ook paragraaf 7.2).

### Beperkingen van dit onderzoek

Binnen dit onderzoek was het mogelijk om twee trials op te zetten op basis van één persona/subcultuur per sociaal mediaplatform. Het opzetten van trials en het bekijken en beoordelen van content van een groot aantal video's door de onderzoekers vraagt namelijk een relatief grote tijdsinvestering. De resultaten van de trials moeten daarom worden gezien als indicatief voor in ieder geval de toegepaste persona's/subculturen. In hoeverre de gevonden resultaten generaliseerbaar zijn naar andere persona's / subculturen moet worden uitgewezen door eventuele aanvullende trials in de toekomst.

## 5.2. Opzet en uitvoering van de trial op YouTube

De trial op YouTube voerden we uit op computers die nog niet voor andere doeleinden zijn gebruikt, waardoor deze geen technische markers als cookies bevatten. Beïnvloeding van de toekomstige resultaten in de trials wordt zo geminimaliseerd. Er is een VPN-verbinding gebruikt die als locatie Nederland aangeeft. Het verloop van de trial en de stappen die hierin zijn genomen op YouTube worden schematisch weergegeven in het onderstaande flowchart schema. Onder deze flowchart zijn de genomen stappen en de bevindingen op basis hiervan in de trials in chronologische volgorde weergegeven.

### Vorbereiding van de trial

1. *Registratie Google account.* We maken een nieuw e-mail-account aan via de 'Gmail sign up' pagina. Hiertoe gebruiken we een fictieve, nog niet bestaande identiteit (voornaam, achternaam) en controleren we dat er daadwerkelijk niemand met deze voor- en achternaam gevonden kan worden via zoekmachines en sociale media platforms. Dit maakt de identiteit geschikt voor deze taak zonder eventuele persoonsverwisseling of inbreuk op privacy.
2. *Registratie YouTube account.* We maken een nieuw account aan op het platform YouTube, waarbij we gebruik maken van dezelfde identiteitsgegevens als bij de registratie van de Google account. Wanneer we de account activeren verifiëren we ook alle instellingen. Bij alle keuzen die we als gebruiker kunnen maken, kiezen we steeds voor de meest gepersonaliseerde optie<sup>12</sup>. Zodra het Google/Gmail account is aangemaakt kunnen we ook de instellingen aanpassen die Google toelaten meer of minder informatie te delen met sociale media platforms zoals YouTube. De minst privacy vriendelijke opties zijn in de meeste gevallen reeds als standaard-opties aangevinkt. In alle andere gevallen kiezen we voor de instellingen die het delen van communicatie toelaten en dus meer gepersonaliseerde resultaten zullen geven. De geselecteerde opties geven YouTube dus de vrijheid om resultaten te personaliseren volgens het profiel en de activiteiten van de gebruiker.

<sup>12</sup> Hierbij activeren we de volgende instellingen:

- Choose personalisation settings: "Express personalisation". Use personalisation settings that deliver tailored content and ads.
- Confirm personalisation settings and cookies. These settings will use your data to give you a personalised experience across services and more control over the ads that you see:
- Web & App Activity: This setting will be turned on for things like faster searching, more relevant results and more helpful app and content recommendations.
- YouTube History. This setting will be turned on for things like better video recommendations and to help remember from where you left off.
- Ad personalisation. This setting will be turned on to make the ads that you see more useful to you.

3. *Koppelen Google en YouTube account.* We koppelen het nieuw aangemaakte Google-account met het nieuw aangemaakte YouTube account.
4. *Activatie en login gekoppelde YouTube account.* De activatie van het gekoppelde YouTube account voeren we uit via de Chrome browser, zodat we de eventuele effecten kunnen onderzoeken van de Google activiteiten op de aanbevelingen in YouTube.

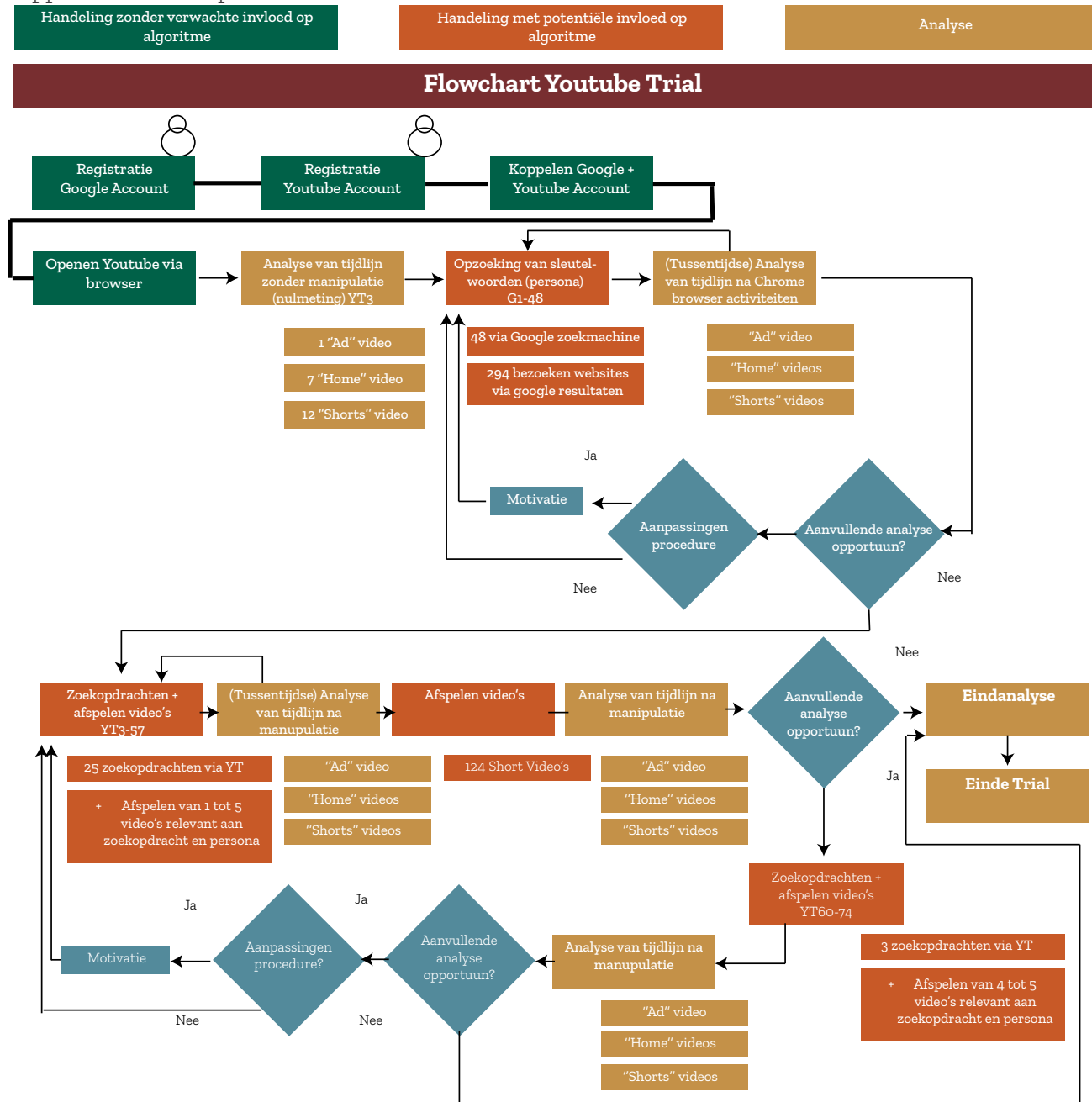
### **Nulmeting op basis van de videosuggesties op YouTube**

5. *Analyse van suggesties op startpagina YouTube zonder manipulatie door activiteiten op Google (nulmeting).* Zodra we het YouTube account hebben geactiveerd toont de startpagina van YouTube suggesties voor video's. De aanbevolen video's op de startpagina van YouTube analyseren we kwalitatief als nulmeting voor de mogelijke beïnvloeding hiervan door Google activiteiten<sup>1314</sup>. Nota bene: bij het activeren van het YouTube account hebben we via het Google account nog geen activiteiten ondernomen zoals Google opzoeken, Chrome browser activiteiten etc.

<sup>13</sup> Meer informatie over de uitvoering van trials is desgewenst te verkrijgen bij TextGain, via [info@textgain.com](mailto:info@textgain.com).

<sup>14</sup> De startpagina is als volgend opgebouwd: (a) 1 advertentie; (b) 7 "homepage" video's; (c) 12 "shorts" video's; (d) 6 "recommended movies by YouTube Movies"; en (e) homepage video's volgens het "infinite scrolling" design. Infinite scrolling is een online design benadering waarbij steeds nieuwe content wordt geladen wanneer een gebruiker bij het tijdelijke einde van de pagina komt. Hierdoor ontstaat een theoretisch oneindige stroom van nieuwe content. Bij de uitvoering van de homepage analyses in de trial beperken we ons tot de video's die relevant zijn binnen het kader van dit onderzoek, te weten de "7 Homepage" video's (b), 12 "Shorts" video's (c), en de eerste 23 video's uit de homepage video's (e).

Figuur 1. Flowchart van de stappen uit de trial op YouTube.



## Resultaten van deze nulmeting

- Bij het eerste bezoek aan de YouTube homepage worden video's gesuggereerd die beantwoorden aan het interesseveld van het brede publiek, rekening houdende met de informatie die door de gebruiker wordt vrijgegeven. In dit geval werd enkel de locatie (Rotterdam, Nederland) vrijgegeven. Dit wordt meteen gereflecteerd in de suggesties op de homepage door o.a. Nederlandstalige video's met hoogtepunten uit grote Europese voetbalcompetities (Feyenoord) of Nederlandstalige video's met een groot aantal views die maatschappelijk relevant zijn voor Nederland.
- Een mogelijke invloed van de locatie stellen we niet vast bij de Shorts video suggesties. Hier zien we vooral Engelstalige video's met een hoog aantal views (5 tot 91 miljoen) over een breed spectrum van onderwerpen.
- De homepage video's die worden weergegeven onder de Shorts video's vertonen daarentegen we hetzelfde patroon als de andere homepage video's die hierboven werden beschreven, dus suggesties op basis van locatie.

## Manipulatie via Chrome/Google

6. *Opzoeken van eerste set zoektermen via Chrome/Google en bekijken aanbevolen websites.* We onderzoeken de invloed van Chrome browser activiteiten en Google zoekopdrachten op de suggesties van YouTube 's homepage. Hiertoe voeren we, met gebruik van de Chromebrowser, in de zoekbalk van Google zoekwoorden in die passen binnen de interesse van het persona, namelijk de interesse in de Vikingcultuur (Odinisme, Wodanisme, Asatru e.d. en mannelijke strijders). De toegepaste zoekwoorden hebben betrekking op gangbare thema's en termen binnen de Vikingcultuur om een gepersonaliseerd zoekgedrag te simuleren. De opzoeken betreffen:

- Algemene opzoeken over de stromingen en basisbegrippen van deze subcultuur.<sup>15</sup>
- Specifieke sleutelbegrippen en -figuren (inclusief Goden) uit deze subcultuur.<sup>16</sup>
- Gebruikte symboliek binnen de subcultuur zoals runen en mythische figuren.<sup>17</sup>

Voor deze zoekopdrachten beperken we ons steeds tot Nederlandstalige sleutelwoorden, zoals "Odinisme". Enkel wanneer we geen relevante Nederlandstalige resultaten krijgen van de zoekmachine verfijnen we de zoekopdracht door middel van een vraagstelling zoals "Wat is Odinisme". Voor alle zoekwoorden die we hebben ingevoerd in Google selecteren we de Google-resultaten die aansluiten bij het interesseveld van het persona. We klikken op de websites die de Google-resultaten suggereren en bezoeken deze, evenals relevante subpagina's van deze websites. Daarna herhalen we de zoekopdracht via de "images" zoekmachine van Google.<sup>18</sup> In totaal zijn er 48 zoekopdrachten uitgevoerd in de Google zoekmachine. Via deze zoekopdrachten bezochten we in totaal 294 relevante websites die beantwoorden aan de interesse van het persona.

15 Bijvoorbeeld: Odinisme, Wodanisme, Wotanisme, Asatru, Heathenry, Edda.

16 Asgaard, Yggdrasil, Ragnarok, Walhalla, Odin, Thor.

17 Sigrune, Eihwaz, Algiz, Othala, Odal, Tyr, Tiwaz, Lebenrunne, Zonnerad, Triskelion, Vegvisir, Ægishjalmur, Fenris.

18 Het aantal en de selectie van de te bezoeken websites per ingevoerde zoekterm zijn afhankelijk van:

- Het aantal relevante resultaten. Sommige zoekopdrachten geven meer relevante resultaten dan andere zoekopdrachten. Hierdoor worden voor de ene zoekopdracht 5 websites bezocht terwijl voor andere slechts 2 relevante resultaten beschikbaar waren.
- De relevantie van de resultaten. Enkel websites binnen het interesseveld van het persona worden bezocht.
- De taal van de relevante resultaten. Er werd steeds prioriteit gegeven aan Nederlandstalige websites. Bij afwezigheid werd prioriteit gegeven aan Engelstalige websites.
- Bekijken van websites tijdens vorige zoekopdrachten. Sommige websites werden meermaals weergegeven in de zoekresultaten. Er werd steeds prioriteit gegeven aan websites die nog niet werden bezocht.

## Tussentijdse meting effect van Chrome/Google manipulatie op videosuggesties YouTube

7. *Analyse van suggesties op startpagina YouTube ná de Chrome/Google manipulatie.* Na het invoeren van de zoekwoorden op Chrome/Google en het bezoeken van aanbevolen websites op Chrome, analyseren we wederom welke video's op de startpagina van YouTube worden aanbevolen. Doel is vast te stellen wat de invloed is van de Chrome browser activiteiten en Google zoekopdrachten op de suggesties van YouTube 's homepagina. We openen opnieuw de YouTube homepagina, zonder enige voorafgaande manipulatie op YouTube. De selectie en analyse van de aanbevolen video's doen we vervolgens op dezelfde wijze als in stap 5<sup>19</sup>. Na de 48 zoekopdrachten via de Chrome browser in de Google zoekmachine en het bezoeken van 294 relevante websites, zonder een enkele activiteit op het platform YouTube, blijkt dat 28 procent van de suggesties op de YouTube homepagina rechtstreeks gerelateerd zijn aan de online zoekopdrachten op Google. In de Shorts suggesties zien we de eerste gerelateerde video op de zesde plaats. Het onderwerp van deze video is Fenris, een wolf uit de Noorse mythologie. Onder de Shorts start de oneindige lijst van suggesties. Hierbij zien we bij de eerste twee suggesties onmiddellijk video's die gerelateerd zijn aan onze online zoekopdrachten. Deze betreffen een video over Odin en een video over de hamer van de god Thor.

19 Te weten, de "7 Homepage" video's (b), 12 "Shorts" video's (c), en de eerste 23 video's uit de homepage video's beschreven als (e).

## Invloed van het bekijken van aanbevolen video's op nieuwe videosuggesties van YouTube

8. *Opzoeken van tweede set zoektermen en bekijken aanbevolen video's.* We willen nu vaststellen wat de invloed is van het zoekgedrag op YouTube zelf op de video's die worden gesuggereerd op YouTube 's homepagina. Allereerst wisten we alle gegevens die invloed kunnen hebben op de resultaten, opdat er geen invloed is door eerdere opzoeken in de Chrome Browser. We voeren via de zoekbalk in YouTube daartoe zoekwoorden in. Dit zijn wederom zoektermen op gangbare thema's en termen binnen de Vikingcultuur, opdat we een gepersonaliseerd zoekgedrag door het persona simuleren. Het betreft (wederom) algemene opzoeken over de stromingen en basisbegrippen van de Vikingcultuur; specifieke sleutelbegrippen en -figuren (inclusief Goden) uit de subcultuur<sup>20</sup>; en gebruikte symboliek binnen de subcultuur zoals runen en mythische figuren<sup>21</sup>. In totaal zijn er 25 zoekopdrachten uitgevoerd in zoekbalk van YouTube. Voor alle ingevoerde zoekopdrachten bekijken we een selectie van video's die door YouTube worden gesuggereerd.<sup>22</sup>
9. *Analyse van nieuwe suggesties op startpagina YouTube na het bekijken van aanbevolen video's.* Nadat we in stap 8 de relevante video's hebben bekeken die door de zoekopdrachten werden gesuggereerd, analyseren we in stap 9 opnieuw welke video's worden aanbevolen op de YouTube homepagina.

20 Asgaard, Yggdrasil, Ragnarok, Walhalla, Odin, Thor.

21 Sigrune, Eihwaz, Algiz, Othala, Odal, Tyr, Tiwaz, Lebenrune, Zonnerad, Triskelion, Vegvisir, Ægishjalmur, Fenris.

22 Het aantal - en de selectie van de te bekijken video's per ingevoerde zoekterm zijn afhankelijk van:

- Het aantal relevante resultaten. Sommige zoekopdrachten geven meer relevante resultaten dan andere zoekopdrachten. Hierdoor worden voor de ene zoekopdracht 5 video's bezocht terwijl voor andere slechts 1 relevant resultaat beschikbaar was.
- De relevantie van de resultaten. Enkel video's binnen het interesseveld van het persona worden bezocht.
- De taal van de relevante resultaten. Er werd steeds prioriteit gegeven aan Nederlandstalige video's. Bij afwezigheid werd prioriteit gegeven aan Engelstalige video's.
- Bekijken van video's tijdens vorige zoekopdrachten. Sommige video's werden meermaals weergegeven in de zoekresultaten. Er werd steeds prioriteit gegeven aan video's die nog niet werden bekeken.

De analyse van de aanbevolen video's doen we op dezelfde wijze als in stappen 5 en 7 (zie aldaar). Doel hiervan is om vast te stellen of het bekijken van de aanbevolen video's (in stap 8) resulteert in een toename van het aantal aanbevolen video's dat relevant is voor het persona. Met 'relevant' bedoelen we dat de content van de aanbevolen video's gerelateerd is aan de ingevoerde zoekopdracht én afgespeelde video's. Dit wordt in de analyse bevestigd; we constateren dat het opzoeken en vervolgens afspelen van content (video's) die beantwoordt aan het interesseveld van het persona ervoor zorgt dat er steeds meer relevante content op de YouTube homepage verschijnt. Ten opzichte van de eerder geconstateerde invloed van online Chrome/Google zoekopdrachten buiten het platform, is de invloed op de relevantie van de door het algoritme gesuggereerde video's bijna dubbel zo sterk.

10. *Afspelen van YouTube Shorts en nieuwe analyse van de suggesties op startpagina YouTube.* Vervolgens gaan we op zoek naar de invloed van het bekijken van de *Shorts*<sup>23</sup> suggesties die worden weergegeven via de YouTube homepage. Doel hiervan is wederom om vast te stellen of het bekijken van de aanbevolen video's resulteert in een toename van het aantal video's dat relevant is voor het persona. We starten dit deel van de trial met de video die als eerst in de lijst met suggesties staat gerangschikt. Daarna bekijken we de 124 volgende *Shorts* video's die door het algoritme aan de gebruiker worden voorgesteld. Indien de video beantwoordt aan het interesseveld van het persona bekijken we de "Short" volledig. Indien dit niet het geval is, scrollen we onmiddellijk verder naar de volgende video. Nadat we deze *Shorts* hebben bekeken, doen we een nieuwe analyse van de YouTube homepage. Hierbij stellen we vast dat sprake is van een status quo of lichte stijging van relevante video's in alle verschillende secties van de homepage.

### **Confrontatie van het persona met het debat over 'racisme in de Vikingcultuur'**

In Paragraaf 3.7 hebben we beschreven dat bij een klein deel van de aanhangers van de Noorse traditie binnen de Paganistische gemeenschap sprake is van racistische narratieven over witte superioriteit en haatspraak naar andere culturen en religies. Het merendeel van de Paganisten verwerpen deze ideeën met klem en pleiten sterk voor inclusiviteit. Geïnteresseerde personen die online informatie opzoeken over de Vikingcultuur zullen vroeg of laat ongetwijfeld met deze problematiek en discussies in aanraking komen. Deze confrontatie van het persona met dit debat is de volgende stap in de trial.

11. *Zoekopdracht 'racisme in paganisme' en bekijken aanbevolen video's.* We willen vaststellen wat de invloed is van het zoeken van informatie over de hierboven genoemde discussie op de content (video's) die wordt gesuggereerd op YouTube's homepage. We voeren daartoe via de zoekbalk in YouTube een zoekopdracht in die het persona meer informatie kan geven over de discussie. We gebruiken hiervoor de algemene zoekopdracht "Racisme in Paganisme". Bewust laten we het persona niet actief zoeken naar informatie over radicale content, hoewel dit in de praktijk naar verwachting wel zal voorkomen. Wij willen echter onderzoeken of aan personen die algemene zoekopdrachten invoeren door het algoritme vervolgens wel of niet radicalere content wordt gesuggereerd.

<sup>23</sup> YouTube Shorts zijn video's van maximaal 60 seconden die te bekijken zijn via kanalen en de Shortspagina op de homepage.



### Resultaten van de opzoeking:

De resultaten die het YouTube algoritme zijn te onderscheiden in de volgende drie categorieën:

1. Video's van officiële media die berichten over de discussie 24.
2. Video's van 'Pagan-influencers' of experts die de discussie aankaarten 25.
3. Borderline content.

Indien we de video's van officiële media en experts nader bekijken dan blijkt dat deze soms meer, soms minder kritisch staan tegenover het maatschappelijk probleem dat ze willen toelichten. Het persona wordt zowel geconfronteerd met kritiek op racisme in de Vikingcultuur als met personen uit de subcultuur zelf die aan het woord worden gelaten die het racisme bagatelliseren. Zo wordt tijdens de zeven minuten durende reportage van Sacramento Bee hoofdzakelijk leden van de extremistische organisatie Asatru Folk Assembly (Southern Poverty Law Center, g.d.) aan het woord gelaten en slechts heel beperkt kritische stemmen van experts. Ook de video van Vice is niet terughoudend in zijn berichtgeving en zet de spotlights op Paul Waggener, zijn organisaties Operation Werewolf en Wolves of Vinland en Lana Lokteff van Red Ice TV door, onder andere, hun namen voortdurend op het scherm te laten verschijnen. De experts zijn wel kritischer over de activiteiten van deze individuen en organisaties.

24 Resultaten onder eerste negen gesuggereerde video's: (1) Al Jazeera English; (4) Sacramento Bee; (5) Vice; (9) Rare Earth.

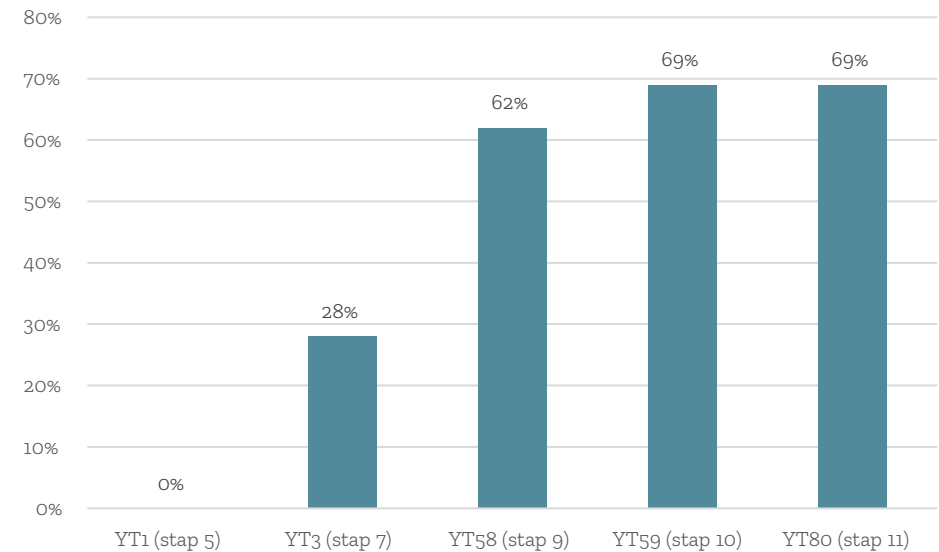
25 Resultaten onder eerste negen gesuggereerde video's: (3) The Norse Witch; (6) Lora O'Brien; (7) The Big Boss; (8) University of Colorado Boulder.

Om de zoektocht van het persona verder te zetten voeren we twee zoekopdrachten uit naar onderwerpen uit de video van Vice. Als eerste laten we het persona op zoek gaan naar informatie over de organisatie Wolves of Vinland (Burley, 2020) die in desbetreffende video wordt genoemd. Voor deze zoekopdracht geeft YouTube een resultaat waarvan acht video's (van de eerste negen suggesties die worden gegeven) rechtstreeks gelinkt kunnen worden aan de zoekopdracht. Na één zoekopdracht en het bekijken van relevante video's zien we vervolgens meteen in drie van de eerste vijf suggesties op de YouTube homepagina content die we kunnen categoriseren als borderline content<sup>26</sup>. Daarnaast vinden we in de YouTube homepagina resultaten na de Shorts nog video's terug van kanalen zoals "Norse Magic and Beliefs", "The Desert Heathen" en opnieuw Paul Waggener die op openbare fora reeds in opspraak kwamen omdat zij racisme niet wilden veroordelen. Vervolgens laten we het persona op zoek gaan naar een tweede persoon die in de eerste set video's werd genoemd: Lana Lokteff (Mattheis, 2018). Lokteff levert online bijdragen om de White Nationalist Movement toegankelijk te maken voor een breder publiek (Bowman & Stewart, 2017; Mattheis, 2018). Voor deze zoekopdracht geeft YouTube als resultaat suggesties voor video's waarin experts de rol van vrouwen binnen de alt-right gemeenschap bespreken. Een video waarin Lokteff wordt geïnterviewd of zelf het woord neemt vinden we pas terug onder de Shorts suggesties. We bekijken vervolgens vijf video's die relevant zijn voor de zoekopdracht en analyseren vervolgens opnieuw de YouTube homepagina. Na deze zoekopdracht en het bekijken van de relevante video's constateren we een daling van het aantal gesuggereerde video's die we kunnen categoriseren als borderline content.

26 Video's van het kanaal Operation Werewolf; het Kanaal Paul Waggener; het kanaal The Wisdom of Odin (zie o.a.: Burley, 2020).

### Grafische weergave resultaten YouTube trial per fase

De stijging van het *percentage relevante resultaten (video's)* dat in de tijdlijn van het persona verschijnt in de verschillende fasen van de trial, kan ook grafisch worden weergegeven. De term 'relevante video's' verwijst in de getoonde grafiek naar video's die inhoudelijk relevant zijn voor het persona (hier: een persoon die zich interesseert in de Vikingcultuur binnen het (neo-)paganisme/heidendom). Opzoeken in de Chrome Browser resulteren in 28 procent video's die inhoudelijk relevant zijn voor het persona (hebben inhoudelijk betrekking op de Vikingcultuur), de overige suggesties voor video's zijn niet relevant voor het persona. Na opschoning opdat de opzoeken in de Chrome Browser geen invloed meer hebben, worden zoekopdrachten herhaald in de zoekbalk van YouTube. Dit resulteert in 62 procent relevante video's. Dit percentage stijgt door naar 69 procent na het afspelen van de shorts. Tot slot worden zoekopdrachten ingevoerd die rechtstreeks raken aan de gepolariseerde discussies over racisme binnen de Vikingcultuur. Deze zoekopdrachten leiden niet tot een hoger percentage relevante video's. Opvallend in de grafiek is de afvlakking van het percentage relevante video's na YT58. Het YouTube algoritme toont een maximum van 69 procent relevante video's in de trials.



Figuur 2. Percentage van relevante video's voor het persona op YouTube als functie van trialfase.

#### Legenda van de grafiek

YT1 = nulmeting voorafgaand aan enige manipulatie.

YT3 = percentage relevante video's op homepagina YouTube na zoekopdrachten via Chrome Browser.

YT58 = percentage relevante video's op homepagina YouTube na zoekopdrachten in YouTube zoekbalk.

YT59 = percentage relevante video's op homepagina YouTube na afspelen "Shorts".

YT80 = percentage relevante video's op homepagina YouTube na opzoeken gepolariseerde discussie

### *Ontwikkeling aandeel borderline content*

Hoe ontwikkelt zich nu het aandeel suggesties voor video's met extreme content of borderline content? Zoals eerder is toegelicht, suggereren de YouTube aanbevelingsalgoritmen video's met neutrale content of borderline content omtrent de Vikingcultuur, maar geen video's met extremistische content. Nadat we allereerst algemene zoekopdrachten hebben ingevoerd, suggereert YouTube video's over de Vikingcultuur die bijna allemaal inhoudelijk neutraal zijn. Ook nadat we de video's hebben bekeken, verschijnen opnieuw suggesties op de YouTube-homepagina die als neutraal zijn te beoordelen. Hetzelfde beeld zien we nadat we zoekdrachten hebben ingevoerd en video's bekeken omtrent specifieke thema's en symbolen binnen de Vikingcultuur. Ook dan suggereert YouTube video's over de Vikingcultuur die grotendeels inhoudelijk neutraal zijn. Echter, zodra we zoekopdrachten invoeren die rechtstreeks raken aan de gepolariseerde discussies over racisme binnen de Pagan/Vikingcultuur, constateren we dat er video's met borderline content verschijnen op de homepagina van YouTube (zie hiervan de beschrijving na stap 11). Nadat aanvankelijk het aantal gesuggereerde video's met borderline content toeneemt tot ongeveer een derde van alle aanbevolen content, stabiliseert het aandeel video's met borderline en daalt zelfs licht als nog nieuwe zoekopdrachten worden gegeven.

### **Samenvatting en duiding van de bevindingen op de YouTube-trial**

#### *Samenvatting*

In de trial onderzochten we of en in hoeverre aanbevelingsalgoritmen op YouTube op basis van gepersonaliseerd zoekgedrag door gebruikers (persona's / invoer van zoekwoorden) leiden tot suggesties voor extremere content. Ook zijn we nagegaan of er sprake is van een relatie tussen het zoekgedrag van gebruikers op Google en aanbevelingen naar meer extreme content die gebruikers krijgen op YouTube. In de trial is beoogd om een zoekgedrag te simuleren door middel van een persona die een persoon representeert die geïnteresseerd is in de Vikingcultuur en Noorse tradities. Nadat we de nodige voorzorgen namen om beïnvloeding van historisch online zoekgedrag uit te sluiten, zochten we via de Chromebrowser en Google zoekmachine sleutelwoorden op die aansluiten

aan het interesseveld van het persona. Hierbij deden we tientallen opzoeken via de Google zoekmachine en bezochten bijna 300 webpagina's gerelateerd aan het Paganisme, Odinisme, Asutru of Heathenry. Daarna bekeken we of activiteiten buiten de YouTube applicatie een invloed hebben op de suggesties die door YouTube aan de gebruiker worden gedaan.

1. We concluderen dat, enkel door browser activiteiten, en dus zonder enige activiteit binnen de YouTube applicatie, deze suggesties hierop sterk worden afgestemd.

Vervolgens herhaalden we de zoekopdrachten op YouTube en analyseerden opnieuw welke video's door het algoritme worden gesuggereerd aan de gebruiker;

2. Hierbij stelden we vast dat de invloed van activiteiten op het YouTube platform een logischerwijze sterke invloed hebben op deze suggesties. Deze invloed was nog, sterker dan de invloed van de browser activiteiten op Chrome/Google.

In een volgende fase van de trial confronteerden we het persona met de gepolariseerde discussies omtrent racisme binnen een deel van de (neo-)paganistische- en Vikingcultuur;

3. We stelden vast dat de aanbevelingsalgoritmen geen radicale content suggereren, en dat media en experts in meer of mindere mate kritisch berichten over deze discussie. Maar we constateerden tegelijkertijd ook dat vooraanstaande makers van borderline content een podium krijgen en dat video's waarin deze personen figureren een hoge aanbeveling krijgen op de YouTube homepagina.

### Duiding van de resultaten

De resultaten van de trial geven ondersteuning voor de notie dat YouTube lessen heeft geleerd uit de ophef die ontstond in 2017, waarbij werd gesteld het YouTube-algoritme de gebruikers naar extremere content zou leiden (Ledwich & Zaitsev, 2019). Hierop voerde YouTube aanpassingen door om aanbevelingen naar extreme content te voorkomen, contentmoderatie te versterken en problematische content creators van het platform te verwijderen (Munger & Phillips, 2020). De bevindingen van de trials suggereren echter ook dat borderline content op YouTube al dan niet gedeeltelijk de plaats lijkt te hebben ingenomen van de extremistische content die voor 2017 welig tierde. Wanneer een gebruiker extremere zoektermen gaat invoeren, komt deze op een gegeven moment terecht in een grote *pool* (of 'kuil') van borderline content in plaats van een fuik met daarin content die toeneemt in radicaliteit. Dergelijke pools met borderline content lijken 'culturen' te worden binnen het mainstream discours, maar ook in het politieke domein. Het risico is dan ook aanwezig dat zo'n cultuur het karakter krijgt van een zelfstandig functionerende ecologie die zichzelf versterkt; dat wil zeggen dat de aanwezige borderline content steeds normaler wordt gevonden ('mainstreaming' en normalisering).

Het resultaat dat borderline content op YouTube de plaats lijkt te hebben ingenomen van de extremistische content is opmerkelijk aangezien YouTube in 2022 aankondigde het fenomeen van borderline content aan te pakken door deze te downgraden in de suggesties. In de woorden van YouTube Chief Product Officer Neal Mohan: *"borderline content concerns videos that don't quite cross the line of our policies for removal but that we don't necessarily want to recommend to people"* (Bell, 2022). We concluderen hierom op basis van onze trial dat de strategie om borderline content lagere aanbevelingen te geven kennelijk verbeteringen behoeft. Borderline content tast de grenzen van de contentmoderatie af die door het gebruik van gecodeerde taal, symbolen en ambigue taalgebruik steeds meer expertise vereist om deze correct te identificeren. Deze trial toont aan dat het gebrek aan deze expertise op dit ogenblik een zwakke schakel is in de strijd tegen borderline content op YouTube. Een effectieve aanpak van

borderline content staat of valt met de correcte identificatie ervan. In dit proces kan een belangrijke rol worden weggelegd voor de partners uit het YouTube Trusted Flagger-programma (Google LLC (webpagina, g.d. (c)) evenals andere Trusted Flaggers (zie paragraaf 7.4 voor een discussie over deze zogeheten Trusted Flaggers).

Een belangrijke bevinding van de trial is dat in de praktijk relatief radicalere (borderline) content vaker wordt verkregen door de invoer door de gebruiker zelf van radicalere zoekopdrachten en niet zozeer door de algoritmische aanbevelingssystemen van de platforms. Dit kan opzichzelf worden geduid als een positieve uitkomst: om radicalere content te 'zien', moest hierop expliciet gezocht worden door de gebruiker. Het effect van contentmoderatie van radicale content op het platform is dus zichtbaar en ondersteunt de notie dat contentmoderatie belangrijk en nodig is. Een mogelijke aanbeveling richting sociale mediaplatforms is: houd rekening met zoekopdrachten van gebruikers (bijvoorbeeld in een bepaald land of gebied). Dit omvat vragen als welke woorden worden waar ingevoerd door gebruikers en hoe kun je hierop anticiperen als mediaplatform? De nationaal coördinator (zie: Hoofdstuk 7) kan hierover het gesprek aangaan met het platform. Hier is mogelijk winst te behalen: niet alleen aandacht hebben voor zoekresultaten, maar juist voor de zoektermen.

Een andere opvallende bevinding in de trials is dat de meest radicale (borderline) content op YouTube te vinden was in de *commentsecties* van video's (over artikelen) van nieuwsmidia. Dergelijke berichten lijken daarmee een aantrekkingskracht te hebben op personen bij wie dit bepaalde reacties uitlokt. Zo worden veel mediaberichten waarin enkel nieuws weergegeven wordt kritisch, afwijzend of woedend becommentarieerd of worden deze berichten opnieuw gedeeld door YouTube-gebruikers (bijvoorbeeld om andere reacties uit te lokken of de eigen boodschap te verspreiden). In een onderzoek door Whitney Phillips (2018) komt eveneens naar voren dat reguliere nieuwsmidia online al dan niet onbewust kunnen bijdragen aan een verdere verspreiding van (rechts)

extremistische politieke standpunten. Amerikaanse nieuwsmedia werden namelijk "gekaapt" om vervolgens berichten van haatgroepen te versterken. Phillips (2018) stelt daarom voor om "betere praktijken" te handhaven die bedoeld zijn om manipulatie en schade als gevolg van dergelijke nieuwsberichtgeving te verminderen. Met betrekking tot de Nederlandse context zou het een suggestie kunnen zijn dat de nationaal coördinator hierover het gesprek aangaat met nationale of internationale nieuwsmedia. Doel is dan om deze media en journalisten bewust(er) te maken van dit soort online dynamieken. In zekere zin doet het daarmee een beroep op hun mediawijsheid. Echter, dit blijft meer een constatering dan een aanbeveling. Het is namelijk lastig om deze spiraal aan te pakken vanuit een beleidsmatig oogpunt. Wel kunnen platforms (beter) letten op de commentaren onder berichtgeving en weblinks naar extremere content actief verwijderen. Ook kunnen nieuwsmedia die gebruikmaken van sociale mediaplatforms als YouTube ervoor kiezen om een (strenger) eigen moderatiebeleid te hanteren omtrent hun *commentsectie*, wat hen de mogelijkheid geeft om onwenselijke of schadelijke reacties te verwijderen.

### 5.3. Opzet en uitvoering van de trial op TikTok

#### Opzet van de trial op TikTok

Ook de trial op TikTok voerden we uit op computers die nog niet voor andere doeleinden zijn gebruikt, waardoor deze geen technische markers als cookies bevatten. Beïnvloeding van de toekomstige resultaten in de trials wordt zo geminimaliseerd. Er is een VPN-verbinding gebruikt die als locatie Nederland aangeeft. Het verloop van de trial en de stappen die hierin zijn genomen op TikTok worden schematisch weergegeven in het onderstaande flowchart-schema. Onder deze flowchart zijn de genomen stappen en de bevindingen op basis hiervan in de trials in chronologische volgorde weergegeven.

#### Vorbereiding van de trial

1. *Registratie Google account.* We maken een nieuw e-mail-account aan via de 'Gmail sign up' pagina. Hiertoe gebruiken we een fictieve, nog niet bestaande identiteit (voornaam, achternaam) en controleren we dat er daadwerkelijk niemand met deze voor- en achternaam gevonden kan worden via zoekmachines en sociale media platforms. Dit maakt de identiteit geschikt voor deze taak zonder eventuele persoonsverwisseling of inbreuk op privacy.
2. *Registratie TikTok account.* We maken een nieuw account aan op het platform TikTok, waarbij we gebruik maken van dezelfde identiteitsgegevens als bij de registratie van de Google account. Wanneer we de account activeren verifiëren we ook alle instellingen. Bij alle keuzen die we als gebruiker kunnen maken, kiezen we net zoals in de YouTube trial steeds voor de instellingen die het delen van communicatie toelaten en dus meer gepersonaliseerde resultaten zullen geven. De geselecteerde opties geven TikTok dus de vrijheid om resultaten te personaliseren volgens het profiel en de activiteiten van de gebruiker.
3. *Koppelen Google en TikTok account.* We koppelen het nieuw aangemaakte Google-account met het nieuw aangemaakte TikTok account.
4. *Activatie en login gekoppelde TikTok account.* De activatie van het gekoppelde TikTok account voeren we uit via de Chrome browser, zodat we de eventuele effecten kunnen onderzoeken van de Google activiteiten op de aanbevelingen in TikTok.

## Nulmeting op basis van de videosuggesties op TikTok

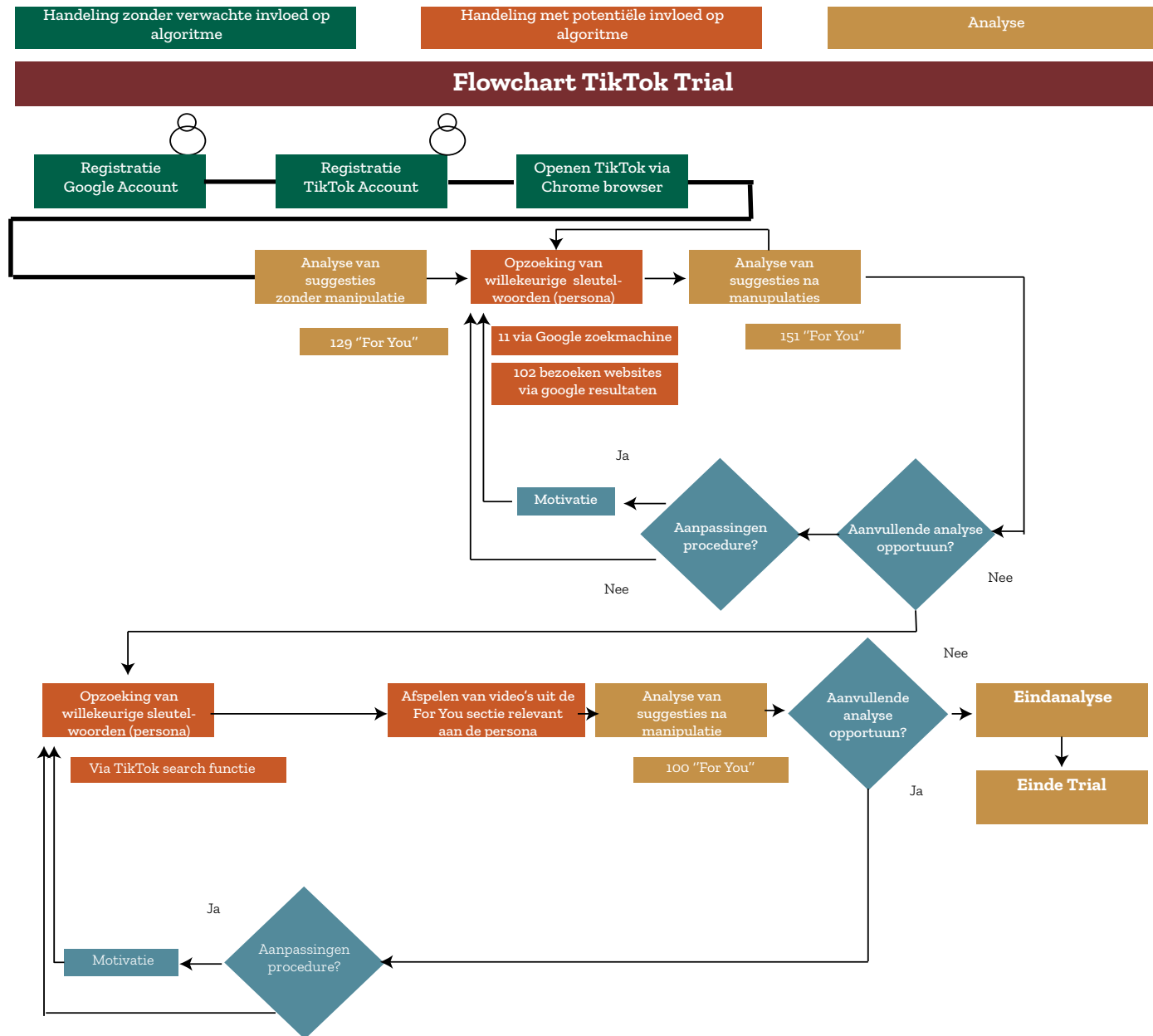
5. *Analyse van suggesties op startpagina TikTok zonder manipulatie door activiteiten op Google* (nulmeting). Zodra we het TikTok account hebben geactiveerd toont de startpagina van TikTok suggesties voor video's. De aanbevolen video's op de startpagina van TikTok analyseren we kwalitatief als nulmeting voor de mogelijke beïnvloeding hiervan door Google activiteiten<sup>27</sup>. Nota bene: bij het activeren van het TikTok account hebben we via het Google account nog geen activiteiten ondernomen zoals Google opzoeken, Chrome browser activiteiten etc.

## Resultaten van deze nulmeting

Bij het eerste bezoek aan de TikTok tijdlijn worden video's gesuggereerd die beantwoorden aan het interesseveld van het brede publiek, rekening houdende met de informatie die door de gebruiker wordt vrijgegeven. In dit geval werd enkel de locatie (Rotterdam, Nederland) vrijgegeven. Dit wordt meteen gereflecteerd in de suggesties op de homepagina door bijvoorbeeld hoofdzakelijk Nederlandstalige video's te tonen en video's van accounts met een zeer groot aantal volgers. We bekeken tijdens deze nulmeting 129 video's uit de For You tijdlijn. Geen enkele van deze video's zijn al relevant voor het vooropgestelde interesseveld (masculiniteit) van het persona.

<sup>27</sup> Deze startpagina van TikTok is zeer eenvoudig opgebouwd en toont afhankelijk van de schermgrootte van de gebruiker een aantal video's in de For You tijdlijn. Deze tijdlijn tracht de gebruiker een theoretisch oneindige reeks van video's te tonen. Wanneer de gebruiker geïnteresseerd is in de video zal de gebruiker deze bekijken, wanneer deze niet geïnteresseerd is kan de gebruiker swipen (mobiele telefoon, tablet, etc.) of klikken (desktop, laptop, etc.) om meteen naar de volgende video te gaan. Aangezien we tijdens deze trial willen vermijden om extremistische accounts of accounts met borderline content een groter bereik te geven beperken we ons tot de For You-tijdlijn.

Figuur 3. Flowchart van de stappen uit de trial op TikTok.



## Manipulatie via Chrome / Google

6. *Opzoeken van eerste set zoektermen via Chrome/Google en bekijken aanbevolen websites.* We onderzoeken de invloed van Chrome browser activiteiten en Google zoekopdrachten op de suggesties van TikTok 's homepagina. Hiertoe voeren we, met gebruik van de Chrome browser, in de zoekbalk van Google zoekopdrachten in die passen bij de interesse in masculiniteit van het persona. De toegepaste zoekopdrachten hebben betrekking op gangbare thema's binnen de subcultuur om daarmee een gepersonaliseerd zoekgedrag te simuleren. De eerste zoekopdracht die we hebben ingevoerd was dezelfde vraag als in de documentaire van Lasfar: "Wat willen vrouwen?" (zie: Paragraaf 3.7). Vervolgens selecteren we uit de "All" zoekresultaten van Google enkele resultaten die aansluiten bij het interesseveld van het persona. We klikken op de websites die de Google-resultaten suggereren en bezoeken deze, evenals relevante subpagina's van deze websites.<sup>28</sup> Als resultaat op deze zoekopdracht worden we onmiddellijk geconfronteerd met een gemeenschap van "professionele vrouwenexperts" die mannen bijstaan om hun mannelijkheid te vergroten en succesvoller te worden bij vrouwen. Op deze websites worden specifieke termen aan ons geïntroduceerd, zoals de alpha, beta of sigma man en belangrijke principes, zoals discipline en fysieke paraatheid om succes bij vrouwen te vergroten. Op deze websites staat de status van de man centraal en het succes van de man afgemeten door middel van de kwantiteit aan vrouwen die hij kan versieren. Naast deze eerste zoekopdracht hebben we nog tien

<sup>28</sup> Het aantal en de selectie van de te bezoeken websites per ingevoerde zoekterm zijn afhankelijk van:

- Het aantal relevante resultaten. Sommige zoekopdrachten geven meer relevante resultaten dan andere zoekopdrachten. Hierdoor worden voor de ene zoekopdracht 5 websites bezocht terwijl voor andere slechts 2 relevante resultaten beschikbaar waren.
- De relevantie van de resultaten. Enkel websites binnen het interesseveld van het persona worden bezocht.
- De taal van de relevante resultaten. Er werd steeds prioriteit gegeven aan Nederlandstalige websites. Bij afwezigheid werd prioriteit gegeven aan Engelstalige websites.
- Bekijken van websites tijdens vorige zoekopdrachten. Sommige websites werden meermaals weergegeven in de zoekresultaten. Er werd steeds prioriteit gegeven aan websites die nog niet werden bezocht.

andere zoekopdrachten in Google ingevoerd. In totaal zijn dus elf zoekopdrachten uitgevoerd via een Chrome browser in de Google zoekmachine.<sup>29</sup> Via deze zoekopdrachten bezochten we in totaal 102 relevante websites die beantwoorden aan de interesse van het persona.

## Tussenmeting effect van Chrome/Google manipulatie op videosuggesties TikTok

7. *Analyse van suggesties op startpagina TikTok ná de Chrome/Google manipulatie ('resultaten eenmeting').* Na het invoeren van de zoekopdrachten op Chrome/Google en het bezoeken van aanbevolen websites op Chrome, analyseren we wederom welke video's op de startpagina van TikTok worden aanbevolen. Doel is vast te stellen wat de invloed is van de Chrome browser activiteiten en Google zoekopdrachten op de suggesties van TikTok 's homepagina. We openen opnieuw de TikTok homepagina, zonder enige voorafgaande manipulatie op TikTok. We beoordelen de eerste 151 zoekresultaten (video's) uit de For You tijdlijn. Het blijkt dat géén van de aanbevolen video's op de TikTok homepagina gerelateerd is aan de online zoekopdrachten op Chrome/Google.

## Invloed van het bekijken van aanbevolen video's op nieuwe videosuggesties van TikTok

8. *Opzoeken van tweede set zoekopdrachten en bekijken aanbevolen video's.* We willen nu vaststellen wat de invloed is van het zoekgedrag op TikTok zelf op de video's die worden gesuggereerd op TikTok 's homepagina. We voeren via de zoekbalk in TikTok daartoe de volgende zoekopdrachten in: "Wat willen vrouwen" en "Hoe word ik een echte man". Na een tussentijdse analyse van de zoekresultaten voeren we vervolgens nog de volgende zoekopdrachten in de zoekbalk van TikTok: "Wat is een alpha man" en "Wat is een

<sup>29</sup> De 11 zoekopdracht waren: Wat willen vrouwen - Hoe word ik een echte man - wat is een alpha man - wat is een beta man - de workout van een echte man - wie is andrew tate - wat is hustler university - wie is jordan peterson - wie is jack donovan



beta man". Voor alle ingevoerde zoekopdrachten bekijken we een selectie van video's die door TikTok worden gesuggereerd.<sup>30</sup>

9. *Analyse van nieuwe suggesties op startpagina TikTok na het bekijken van aanbevolen video's.* Nadat we in stap 8 de relevante video's hebben bekeken die door de zoekopdrachten werden gesuggereerd, analyseren we in stap 9 opnieuw welke video's worden aanbevolen op de TikTok homepage. Doel hiervan is om vast te stellen of het bekijken van de aanbevolen video's resulteert in een toename van het aantal video's dat relevant is voor het persona. Dit wordt in de analyse bevestigd; we constateren dat het zoeken en afspelen van content (video's) die beantwoordt aan het interesseveld van het persona ervoor zorgt dat er steeds meer relevante content op de TikTok homepage verschijnt. Dit is conform de verwachte werking van het TikTok algoritme, dat een belangrijke waarde toekent aan het kijkgedrag van de gebruiker. Het bekijken van een video wordt gezien als interesse en het weg scrollen als geen interesse. Het gevolg hiervan is dat er dus meer video's in de For You tijdlijn verschijnen over onderwerpen die door de handelingen van een gebruiker als interessant worden gecategoriseerd.

### **Confrontatie van het persona met prominente influencers binnen de masculiene subcultuur**

In Paragraaf 3.7 hebben we toegelicht dat de zogeheten manosphere en de nadruk op masculiniteit een online 'gateway' kunnen zijn voor confrontatie van gebruikers met rechtsextremistische content. Geïnteresseerde personen die online informatie opzoeken over masculiniteit, lopen een reële kans om met deze content in aanraking te komen via prominente influencers binnen de

<sup>30</sup> Het aantal - en de selectie van de te bekijken video's per ingevoerde zoekopdracht zijn afhankelijk van a. het aantal resultaten met relevantie voor het persona. Sommige zoekopdrachten geven meer relevante resultaten dan andere zoekopdrachten. Hierdoor worden voor de ene zoekopdracht 5 video's bezocht terwijl voor andere slechts 2 relevante resultaten beschikbaar waren; b. de relevantie van de resultaten. Enkel video's binnen het interesseveld van het persona worden bezocht; c. de taal van de relevante resultaten. Er werd steeds prioriteit gegeven aan Nederlandstalige video's. Bij afwezigheid werd prioriteit gegeven aan Engelstalige video's.

masculiene subcultuur zoals Andrew Tate (en de door hem opgerichte Hustler University), Jordan Peterson en Jack Donovan. Deze influencers zijn bekend bij een breed publiek mede doordat zij volop aandacht in de media krijgen. Het overkoepelende narratief van deze influencers is het 'empoweren' van mannen om 'hun lot in eigen handen te nemen door een sterke masculiene en dominante levenshouding aan te nemen die gekenmerkt is door fysieke paraatheid, een luxueuze levensstijl, discipline, zelfzekerheid en dominantie'. De content die zij op sociale mediaplatforms plaatsen flirt steeds met de grenzen van de netiquette van sociale netwerken om haatspraak tegen te gaan en kan worden beschouwd als borderline content. Voorbeelden hiervan zijn de omstreden uitspraken door Tate (NOS, 2022c) of Peterson (Schleicher, 2018). De confrontatie van het persona met deze prominente influencers binnen de masculiene subcultuur is daarom de volgende stap in de trial.

10. *Zoekopdrachten en bekijken aanbevolen video's.* We willen vaststellen wat de invloed is op de content (video's) die wordt gesuggereerd op TikTok 's homepage door het zoeken van informatie via de zoekbalk in TikTok over de genoemde prominente influencers binnen de masculine subcultuur en de binnen deze context veel gebruikte slogan 'Reclaim masculinity' (Sieckelink et al., 2019).

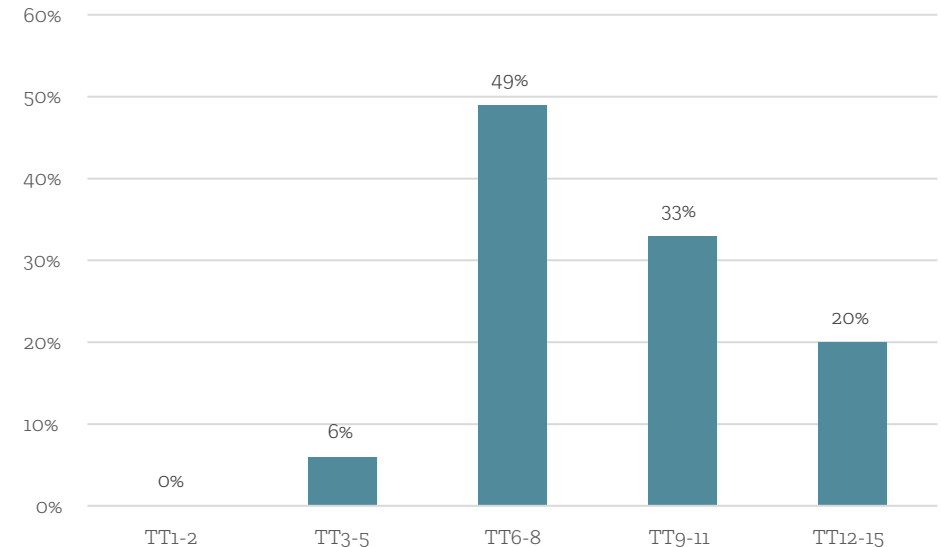
### **Resultaten opzoekingen**

Na het invoeren van zoekopdrachten in TikTok naar de drie influencers en de slogan is het aantal relevante, onderwerp-gerelateerde video's op TikTok 's homepage aanzienlijk hoger dan bij de eerdere algemenere zoekopdrachten ("Wat willen vrouwen" en "Hoe word ik een echte man"). We bekijken ruim 600 aanbevolen video's. Veel van de content op deze video's kan worden gekarakteriseerd als borderline en zijn direct gerelateerd aan de zoekopdrachten. Nieuwe masculiene subthema's die we die niet rechtstreeks introduceren via zoekopdrachten verschijnen slechts sporadisch in de TikTok For You tijdlijn. Tevens constateren we dat bij geen van de aanbevolen video's de content is te duiden als extremistische content die op grond van wettelijke kaders verwijderd kan worden. We merken tevens op dat tussen de "relevante video's" ook een

zeer beperkt aantal de-radicaliseringsboodschappen verschijnen die door het algoritme naar voren geschoven worden omdat ze gelijke hashtags of persoonsnamen bevatten die betrekking hebben op eerdere TikTok zoekopdrachten.

### Grafische weergave resultaten TikTok trial per fase

Het percentage relevante resultaten (video's) dat in de tijdlijn van het persona verschijnt in de verschillende fasen van de trial, kan ook grafisch worden weergegeven. De term 'relevante video's' verwijst in de getoonde grafiek naar video's die inhoudelijk relevant zijn voor het persona (een persoon die zich interesseert in masculiniteit). In de nulmeting en na de trial op Google/Chrome Browser verschijnen nog geen relevante video's. De algemenere zoekopdrachten ("Wat willen vrouwen" en "Hoe word ik een echte man") op het TikTok platform resulteren in 6 procent relevante video's. Opvallend in de grafiek is de sterke stijging van het percentage relevante video's na het invoeren van specifiekere zoekopdrachten die relevant zijn voor het persona. Het invoeren van zoekopdrachten over prominente influencers binnen de masculiene subcultuur en de binnen deze context veel gebruikte slogan 'Reclaim Masculinity' resulteert in hogere percentages relevante content, namelijk 33 respectievelijk 20 procent. De relatieve daling vanaf TT9 in de grafiek heeft te maken met de specificiteit van de zoekopdracht, en niet met de extremiteit van de inhoud van de zoekopdracht. De zoekopdrachten 'Wat is een alpha man' en 'Wat is een beta man' (TT6-8) zijn heel specifiek en leiden in bijna de helft tot het verschijnen van relevante video's op de For You pagina. Dit betekent dat gebruikers bij actief zoeken op specifieke content de hoogste percentages relevante video's voor het persona terugkrijgen.



Figuur 4. Percentage relevante video's voor het persona op TikTok als functie van trialfase.

#### Legenda van de grafieken

TT1 = nulmeting voorafgaand aan enige manipulatie.

TT2 = percentage relevante video's op For You pagina van TikTok na zoekopdrachten via Chrome Browser.

TT3-5 = percentage relevante video's op For You pagina van TikTok na zoekopdrachten via TikTok zoekbalk: 'Wat willen vrouwen' en 'Hoe wordt ik een echte man'.

TT6-8 = percentage relevante video's op For You pagina van TikTok na zoekopdrachten via TikTok zoekbalk: 'Wat is een alpha man' en 'Wat is een beta man'.

TT9-11 = percentage relevante video's op For You pagina van TikTok na zoekopdrachten via TikTok zoekbalk: 'Reclaim masculinity' en 'Wie is Andrew Tate'.

TT12-15 = percentage relevante video's op For You pagina van TikTok na zoekopdrachten via TikTok zoekbalk: 'Wat is Hustler University', 'Wie is Jordan Peterson' en 'Wie is Jack Donovan'.

### *Ontwikkeling aandeel borderline content*

Hoe ontwikkelt zich nu het aandeel suggesties voor video's met extreme content of borderline content? Zoals eerder al is toegelicht, suggereren de TikTok aanbevelingsalgoritmen video's met neutrale content of borderline content omtrent masculiniteit, maar geen video's met extremistische content. Nadat we als eerste algemene zoekopdrachten hebben ingevoerd omtrent masculiniteit (TT3-5), is het aantal suggesties van video's met borderline content aanvankelijk nog zeer beperkt. Nadat we zoekdrachten invoerden omtrent specifieke thema's binnen de masculiniteit (TT6-8), neemt het aandeel suggesties voor borderline video's toe tot ongeveer een kwart – ten opzichte van het aandeel neutrale video's is dit nog steeds een minderheid. Echter, wanneer zoekopdrachten worden uitgevoerd die rechtstreeks raken aan prominente influencers binnen de masculiene subcultuur en de binnen deze context veel gebruikte slogan 'Reclaim Masculinity' dan stijgt het aandeel gesuggereerde video's met borderline content aanzienlijk. Een derde tot de helft kan worden gekenschetst als borderline content en is direct gerelateerd aan de zoekopdrachten. Zoals eerder gesteld verschijnen nieuwe masculiene subthema's die we die niet rechtstreeks introduceren via zoekopdrachten slechts sporadisch in de TikTok For You tijdlijn.

### **Samenvatting en duiding van de bevindingen op de TikTok-trial**

#### *Samenvatting*

In deze trial onderzochten we of en in hoeverre algoritmen op TikTok aanbevelingen doen naar meer extreme content, binnen een gepersonaliseerde context van zoekgedrag (persona) op het platform. In de trial is beoogd om een gepersonaliseerd zoekgedrag te simuleren door middel van een persona met interesse in thema's die worden gerepresenteerd in de manosphere-subcultuur. Om een eventuele invloed vast te stellen van zoekgedrag op Chrome/Google op de aanbevelingen voor video's op TikTok voerden we allereerst op de Chromebrowser en Google zoekmachine relevante zoektermen in die aansluiten aan het interesseveld van het persona. In totaal zijn elf zoekopdrachten uitgevoerd en bezochten we in totaal 102 relevante websites. Daarna

bekeken we of deze invloed hebben op de suggesties die door TikTok aan de gebruiker worden gedaan via de For You pagina. Dit blijkt niet het geval.

1. Resultaat: we hebben geen invloed gevonden van de Chrome/Google activiteiten op de TikTok suggesties.

Vervolgens herhaalden we de zoekopdrachten op het TikTok platform en analyseerden opnieuw welke video's door het algoritme worden gesuggereerd aan de gebruiker.

2. Hierbij stelden we vast dat na deze zoekopdrachten in de TikTok applicatie er steeds meer relevante video's op de For You pagina verschenen.

In een laatste fase van de trial confronteerden we het persona via zoekopdrachten met de content die wordt geproduceerd door of over prominente influencers binnen de masculiene subcultuur. We bekeken 652 video's die via de tijdlijn van TikTok werden gesuggereerd.

3. We stelden vast dat het aanbevelingssysteem geen radicale content suggereert, maar ook dat vooraanstaande makers van borderline content een podium krijgen. Veel van de content op deze video's kan worden gekarakteriseerd als borderline en zijn direct gerelateerd aan de zoekopdrachten.
4. Na het invoeren van zoekopdrachten in TikTok naar de drie influencers en de slogan neemt het percentage relevante, onderwerp-gerelateerde video's op TikTok 's homepagina enorm toe. Hierbij stellen we vast dat hoe langer de gebruiker op het platform vertoeft, hoe accurater het algoritme kan voorspellen welke video's de gebruiker leuk vindt en dus volledig zal bekijken.

### Duiding van de resultaten

Uit de trial blijkt dat het algoritme van TikTok na de opzoekingen geen suggesties voor video's met extreme content teruggeeft. Ook verschijnen suggesties voor nieuwe masculiene onderwerpen (indien we deze niet zelf introduceren via zoekopdrachten) slechts zeer sporadisch in de TikTok For You tijdlijn. De resultaten van de trial geven ondersteuning voor de notie dat TikTok, net zoals YouTube, heeft geleerd van kritiek waarbij werd gesteld het TikTok algoritme de gebruikers steeds naar extremere content zou leiden. De bevindingen van de trial laten echter ook zien dat, net zoals bij YouTube, borderline content op TikTok grotendeels de plaats lijkt te hebben ingenomen van de extremistische content op het platform. Desalniettemin lijkt het aanbod van borderline content wel te worden beperkt door de inspanningen van het platform. Zo constateerden we dat onderwerpen die in het verleden in de media als problematisch werden geïdentificeerd, minder vaak terugkwamen in de resultaten nadat ze op TikTok werden opgezocht. In januari 2023 meldde TikTok dat het borderline content beter kan detecteren (Schneider, 2023).

## 5.4. Handelingsperspectieven platforms

Op basis van de bevindingen van de trials kunnen we enkele handelingsperspectieven formuleren voor de sociale mediaplatforms. Een belangrijke bevinding in de trials is dat radicalere/borderline content vaker wordt verkregen als gevolg van de invoer *door de gebruiker zelf* van radicalere zoekopdrachten en niet primair door de algoritmische aanbevelingssystemen van de platforms. Om in aanraking te komen met radicalere content, moest hierop dus expliciet gezocht worden door de gebruiker. Deze uitkomst is in lijn met de nuances die we eerder plaatsten in paragraaf 4.3, namelijk dat ontwikkelingspaden naar rechtsextremistische radicalisering grotendeels aan andere factoren kunnen worden toegeschreven dan aan de werking van aanbevelingsalgoritmen. Een aanbeveling naar de platforms is dat zij zich dus ook moeten focussen op het gedrag van gebruikers zelf. Een dergelijke aanpak gebeurt bijvoorbeeld reeds bij personen die bijvoorbeeld zoeken op informatie omtrent zelfmoord: men

krijgt dan informatie over een hulplijn. Borderline content lijkt op de platforms grotendeels de plaats te hebben ingenomen van extremistische content. Effecten van contentmoderatie van radicale content op de platforms zijn zichtbaar. Dit ondersteunt de notie dat contentmoderatie belangrijk en nodig is. De bevindingen in de trials suggereren echter ook dat er zeker nog ruimte is voor verdere verbetering om borderline content accurater te detecteren en te weren uit de tijdlijn van de gebruikers. Een effectieve aanpak van borderline content staat of valt met de correcte identificatie ervan, wat vraagt om praktisch handzame definities/labels voor content. Hierin kan een belangrijke rol worden gespeeld door zogeheten Trusted Flaggers. Zij zijn immers toegerust om te beoordelen wat de risico's zijn van borderline content (zie paragraaf 7.4 voor een discussie over *trusted flaggers*). Een andere mogelijke aanbeveling voor de platforms is: houd rekening met zoekopdrachten van gebruikers in een bepaald land of gebied; welke woorden worden waar ingevoerd door gebruikers; hoe kun je hierop anticiperen als mediaplatform? Hier is mogelijk winst te behalen door niet alleen aandacht te hebben voor zoekresultaten, maar juist voor de zoektermen van gebruikers. Deze aanbeveling geldt niet alleen voor de genoemde platforms, maar ook voor zoekwoorden op Google.

## Samenvatting hoofdstuk 5

Om beter rekening te kunnen houden met de gepersonaliseerde aard van algoritmen hebben we ter aanvulling op de literatuurstudie een empirische studie opgezet en uitgevoerd op twee grote sociale media platforms (YouTube en TikTok). In deze 'trials' onderzochten we of aanbevelingsalgoritmen aan gebruikers wel of niet steeds extremere content tonen op basis van gepersonaliseerd zoekgedrag (persona's / invoer van zoekwoorden) (onderzoeksvraag 5). Om in deze trials een meer gepersonaliseerd zoekgedrag te kunnen simuleren hebben we zogeheten 'persona's' opgezet. Persona's worden gemodelleerd naar de doelgroepen erachter, het zijn typische representaties ervan. Ten behoeve van de trials representeren de persona's personen met online zoekgedrag binnen twee van de subculturen, te weten de 'manosfeer/masculiniteit' en de 'Vikingcultuur'. Een tweede factor (naast gepersonaliseerd zoekgedrag) die we in de trials meenemen is de mogelijke invloed van eerder (zoek)gedrag op Chrome/Google op de aanbevelingsalgoritmen van dochteronderneming YouTube en (als referentie) TikTok.

### *Resultaten trials YouTube*

1. De trials laten zien dat eerdere activiteiten op Chrome/Google er toe leiden dat aanbevelingen voor content op YouTube hierop sterk worden afgestemd.
2. De invloed van activiteiten op het YouTube platform hebben een sterkere invloed op de aanbevelingen dan de browser activiteiten op Chrome/Google.
3. In een volgende fase van de trial confronteerden we het persona met de gepolariseerde discussies omtrent racisme binnen een deel van de Vikingcultuur. We stelden vast dat het aanbevelingssysteem geen radicale content suggereert. Maar we constateerden ook dat vooraanstaande makers van borderline content een podium krijgen en dat video's waarin deze personen figureren een hoge aanbeveling krijgen op de YouTube homepage.

### *Resultaten trials TikTok*

4. De trials laten zien dat eerdere activiteiten op Chrome/Google (in tegenstelling tot YouTube) geen invloed hebben op aanbevelingen voor content op TikTok.
5. Naarmate er meer zoekopdrachten op TikTok worden ingevoerd, neemt het aandeel relevante video's toe.
6. In de laatste fase van de trial confronteerden we het persona via zoekopdrachten met de content die wordt geproduceerd door of over prominente influencers binnen de masculiene subcultuur. Uit de trial blijkt dat het algoritme van TikTok na de opzoeken geen suggesties voor video's met extreme content teruggeeft. Ook verschijnen suggesties voor nieuwe masculiene onderwerpen (indien we deze niet zelf introduceren via zoekopdrachten) slechts zeer sporadisch in de TikTok For You tijdlijn. Maar we constateerden ook dat vooraanstaande makers van borderline content een podium krijgen. Veel van de content op deze video's kan worden gekarakteriseerd als borderline en zijn direct gerelateerd aan de zoekopdrachten.

### *Discussie van de resultaten van de trials en handelingsperspectieven*

De resultaten van de trial geven ondersteuning voor de notie dat zowel TikTok als YouTube hebben geleerd van kritiek waarbij werd gesteld dat aanbevelingsalgoritmen op de platforms de gebruikers naar steeds extremere content zou leiden. Om in aanraking te komen met radicalere content, moest hierop expliciet gezocht worden door de gebruiker. Deze uitkomst is in lijn met de nuances in paragraaf 4.3, namelijk dat ontwikkelingspaden naar rechts-extremistische radicalisering grotendeels aan andere factoren kunnen worden toegeschreven dan aan de werking van aanbevelingsalgoritmen. Aanbeveling naar de platforms is dat zij dus moet focussen op ook het gedrag van gebruikers zelf.

De bevindingen van de trials zijn in lijn met de vooronderstelling dat op de beide platforms borderline content grotendeels de plaats lijkt te hebben ingenomen van extremistische content. Hoewel beide platforms maatregelen hebben genomen om behalve extremistische content ook het fenomeen van borderline content aan te pakken, suggereren de bevindingen in de trial dat er zeker nog ruimte is voor verdere verbetering om borderline content accurater te detecteren en te weren uit de tijdlijn van de gebruiker. Borderline content tast de grenzen van de contentmoderatie af die door het gebruik van gecodeerde taal, symbolen en ambigue taalgebruik steeds meer expertise vereist om deze correct te identificeren. Het gebrek aan deze expertise is op dit ogenblik een zwakke schakel in de strijd tegen borderline content op de platforms. Een effectieve aanpak van borderline content staat of valt met de correcte identificatie ervan. In dit proces kan een rol worden weggelegd voor Trusted Flaggers. Ook geldt als mogelijke aanbeveling voor de platforms: houdt rekening met zoekopdrachten van gebruikers in een bepaald land of gebied; welke woorden worden waar ingevoerd door gebruikers; hoe kun je hierop anticiperen als mediaplatform?

Hier is mogelijk winst te behalen: niet alleen aandacht hebben voor zoekresultaten, maar juist voor de zoektermen. Deze aanbeveling geldt niet alleen voor de platforms, maar ook voor zoekwoorden op Google.

Een andere opvallende bevinding in de trials op YouTube was dat de meest radicale content te vinden was onder (in de commentsecties) van artikelen in de nieuwsmedia. Dergelijke berichten kunnen aantrekkingskracht hebben op personen of bepaalde reacties van hen uitlokken. Er worden veel media-berichten becommentarieerd door YouTube-gebruikers; opnieuw geplaatst; boze of opruiende comments onder video's waarin enkel nieuws wordt weergegeven. Suggestie zou kunnen zijn dat de nationaal coördinator hierover het gesprek aangaat met nationale of internationale nieuwsmedia. Doel is dan om media bewuster te maken van dit soort online dynamieken; het doet in zekere zin een beroep op hun mediawijsheid. Echter, dit blijft meer een constatering dan een aanbeveling. Het is namelijk lastig om deze spiraal aan te pakken vanuit een beleidsmatig oogpunt. Wel kunnen platforms (beter) letten op de commentaren onder berichtgeving; verwijderen van weblinks naar extremere content.

# 6 Online en offline radicalisering?

Onderzoeksvragen die worden beantwoord in dit hoofdstuk:

1. Zijn er gevallen bekend waarbij de rechtsextremistische 'sociale mediafuik' een bijdrage heeft geleverd aan radicalisering en uiteindelijke rekrutering van bezoekers van sociale mediaplatforms?
2. Op welke wijze droeg de 'sociale mediafuik' dan bij aan radicalisering en eventuele rekrutering?
3. Zijn er gevallen bekend waarbij naast online radicalisering sprake was van actieve offline rekrutering door rechtsextremisten/-terroristen?

In dit hoofdstuk komt het volgende aan bod:

- Introductie (Paragraaf 6.1).
- Online invloeden bij bekende rechtsextremistische radicaliseringscasussen (Paragraaf 6.2).
- De wisselwerking tussen offline en online domein in radicaliseringsprocessen (Paragraaf 6.3).

## 6.1. Introductie

Dit hoofdstuk beschrijft een aantal bekende radicaliseringscasussen waarin de online context een rol speelt. Tevens wordt er aandacht gegeven aan de veronderstelde wisselwerking tussen het online en offline domein bij (rechts-extremistische) radicalisering. In de afgelopen decennia is er namelijk een forse toename zichtbaar in extreemrechtse activiteiten die de online en offline context overbruggen. Hoewel er verwacht wordt dat deze online context een belangrijke rol speelt in deze activiteiten, wordt er vooralsnog verondersteld dat de offline context cruciaal is. Zoals eerder in deze literatuurstudie is beschreven, wijst de wetenschappelijke literatuur regelmatig op de verwevenheid van het online en offline domein binnen radicalisering (Gill et al., 2015, 2017; Koehler, 2014; Von Behr et al., 2013). Actueel onderzoek wijst immers uit dat deze contexten, zoals genoemd bij de uitgangspunten in Hoofdstuk 1, niet als twee zelfstandige domeinen begrepen kunnen worden (Gill et al., 2015; Mattheis, 2019; Whittaker & Herath, 2019). Het betekent dat de twee domeinen veeleer in het verlengde van elkaar liggen; zij vormen een zeker 'continuüm'. Er is daarentegen (nog) geen duidelijk patroon te onderscheiden in hoe dit 'continuüm' eruitziet voor rechtsextremistische radicalisering, omdat er betrekkelijk weinig bekend is over de precieze wijze waarop deze contexten met elkaar interacteren binnen ontwikkelingspaden van (gewelddadige) radicalisering (Von Behr et al., 2013; Winter et al., 2020), laat staan beschikbare empirische kennis die bevestigt dat een sociale mediafuik of filterbubbel op sociale mediaplatforms heeft bijgedragen aan de radicalisering van een persoon of diens rekrutering.

Er zijn dan ook verschillende redenen die het 'vaststellen' hiervan bemoeilijken. In de eerste plaats is het vrijwel onmogelijk om te reconstrueren dat een persoon in een sociale mediafuik of filterbubbel 'terechtgekomen' is. In dat geval zou men toegang moeten hebben tot de volledige internetgeschiedenis

van die persoon en op basis daarvan bepalen welke inhoud door diegene geconsumeerd is – en in hoeverre dit extremistischer van aard geworden is door het toedoen van aanbevelingsalgoritmen op sociale mediaplatforms. Wanneer er al dan niet aangetoond kan worden dat er sprake was van een fuik, dan geldt dat een fuik niet in een geïsoleerd universum bestaat. Of, anders gezegd, er kan verwacht worden dat naast deze fuik ook talloze andere online én offline mechanismen een rol spelen (denk bijvoorbeeld aan eerdergenoemde online echokamers; zie daarvoor paragraaf 4.3). Vraag is dan of een geradicaliseerd persoon daadwerkelijk geradicaliseerd is door het effect van de fuik of de bubbel, of dat dit toch veeleer te wijten is aan andere mechanismen en omstandigheden, waaronder de *eigen keuze* om extremistische content te consumeren, de *eigen keuze* voor specifieke platforms waarop extremistische content beschikbaar is, de *eigen keuze* om deel te zijn van bepaalde online gemeenschappen (die kunnen fungeren als echokamer), de *eigen keuze* om te abonneren op meer extreme accounts en als een gevolg daarvan extremere content ontvangen et cetera. Het is kortom nagenoeg onmogelijk om empirisch vast te stellen in hoeverre sociale mediafuiken als resultaat van aanbevelingsalgoritmen een bijdrage hebben gehad aan het radicaliseringsproces van een persoon. Wél kan er iets gezegd worden over de rol van de online context in bredere zin in verschillende beschreven radicaliseringscasussen die leidden tot geweldspleging of een aanslag.

## 6.2. Radicaliseringscasussen: online invloeden

Er zijn verschillende rechtsextremistische radicaliseringscasussen bekend die resulteerden in het plegen van geweld en/of een terroristische aanslag. Binnen deze casussen lijken sociale mediaplatforms, fora en andere relevante websites een bijdrage te hebben gehad in dit proces. In een studie van Rauf (2020) naar de rechtsterrorist Brenton Tarrant – die een terroristische aanval pleegde in Christchurch in 2019 op twee moskeeën – kwam naar voren dat sociale media een prominente rol speelden in de totstandkoming van deze aanslag. Zijn studie laat zien dat Tarrant veranderde van ideologie door frequent in

contact te komen met online gemeenschappen en groepen die wit nationalisme uitdragen (Rauf, 2020). Tarrant bleek dan ook een actief waar- en deelnemer van een web van online en offline extremisme (Barton, 2019). Verondersteld wordt dat zijn ideologische radicalisering met name in de hand werd gewekt door het lezen van online nieuws, online manifesten en haatdragende content op het internet. Opmerkelijk is de rol van YouTube in de terreuraanslag van Brenton Tarrant. In zijn laatste bericht voor de aanslag, schrijft Tarrant ‘Subscribe to PewDiePie’, waarmee de schutter gebruikers vraagt zich op YouTube te abonneren voor dit account. PewDiePie, ook wel Felix Kjellberg, is een bekende influencer op YouTube die video’s maakt waarbij hij videogames speelt of komische shows uitvoert. Kjellberg is eerder in de problemen gekomen door het maken van antisemitische en pro-nazistische opmerkingen en is een favoriet onder aanhangers van het witte nationalisme (Romano, 2019). Zo blijkt Kjellberg mannen te hebben ingehuurd om een bord met ‘dood aan alle Joden’ te dragen en gebruikte hij het n-woord in een van zijn livestreams. Desondanks profileert Kjellberg zich niet als rechtsextremist maar kan het kijken van zijn video’s indirect leiden tot aanbevelingen voor extremere content.

Een andere radicaliseringscasus is die van Buckley Wolfe: een man uit de Verenigde Staten die zijn broer doodde met een zwaard, omdat hij geloofde dat zijn broer een van vorm veranderend buitenaards reptiel was (Yesilada & Lewandowsky, 2022). Een journalist onderzocht dit incident door Buckley’s *like*-gedrag op YouTube te analyseren. De journalist vond een geleidelijke verschuiving in zijn interesses. Aanvankelijk hield Buckley vooral van video’s over vechtsporten en fitness. Echter, zijn *like*-gedrag verschoof uiteindelijk naar samenzweringstheorieën en rechtsextremistische content (View, 2019). In een andere casus lijkt eenzelfde proces te hebben plaatsgevonden. Voormalig rechtsextremist Caleb Cain vertelt in de New York Times over de sociale mediafuik op YouTube waarin hij beweerde te zijn terechtgekomen. Hij bekeek duizenden video’s over vrouwenhaat, racisme en complottheorieën (Roose, 2019). Cain gaf aan zich gehersenspoeld te voelen door de extreme inhoud en het gevoel te hebben deel uit te maken van een groep individuen



die exclusieve kennis hadden over sociale kwesties. Het geval van Caleb Cain is uniek aangezien hij inmiddels gederadicaliseerd is en actief vertelt over de impact die het aanbevelingssysteem had op zijn percepties. In het interview met *The New York Times* vertelt Cain zijn persoonlijke ervaring met YouTube's aanbevelingsalgoritme. In 2014 raadde het algoritme van YouTube een zelfhulpvideo aan van Stefan Molyneux, een Canadese talkshowpresentator en zelfbenoemd filosoof. Molyneux sprak over het overwinnen van ontberingen door zelfverbetering en hield zich bezig met vraagstukken zoals de vrije wil, samen met praktisch advies over onderwerpen als dating en sollicitatiegesprekken. De YouTuber had ook een politieke agenda waarin hij een voorstander was van mannenrechten. Cain beweert na verloop van tijd via deze video van Molyneux verder getrokken te worden door het aanbevelingsalgoritme van YouTube waardoor hij in de ban raakte van rechtsextremistische content.

Onderzoek uitgevoerd door Mølmen & Ravndal (2021) laat een vergelijking zien van drie rechtsextremistische terroristen en de rol van het internet in hun radicaliseringsproces. Binnen dit onderzoek hebben zij de casussen van Peter Mangs, Anders Breivik en Dylann Roof met elkaar vergeleken. Peter Mangs is bekend om de reeks aanslagen in Malmö in Zweden tussen 2009 en 2010 waarin de schutter het had gemunt op mensen met een donkere huidskleur of een niet-Zweeds uiterlijk. Anders Breivik is de dader van de aanslagen in Noorwegen in 2011, waarbij in totaal 77 mensen om het leven kwamen. Voorafgaand aan de aanslag schreef hij een manifest waaruit rechtsextremistische opvattingen blijken. Tenslotte is Dylan Roof bekend door zijn aanslag op zwarte kerkgoeters in Charleston in South Carolina in 2015. Ook de aanslag van Roof werd gekenmerkt door rechtsextremistisch gedachtegoed. Voor alle drie de gevallen blijkt dat het internet vooral een rol heeft gespeeld in het faciliteren van informatie. Anders Breivik had voor zijn online activiteiten waarschijnlijk enigszins radicale overtuigingen maar radicaliseerde verder tijdens zijn zelfopgelegde isolement. Tussen 2009 en 2010 schreef hij zijn politiek manifest dat volledig uit online bronnen werd samengesteld. Dit illustreert hoe het internet zijn ideologische ontwikkeling vergemakkelijkte. Mangs begon online onderzoek te

doen tijdens het schrijven van zijn ongepubliceerde politieke manifest 'De Germaanse Filosofie'. Ondanks uitgebreid gebruik van het internet tussen 2001 en 2009, bleef de kern van zijn ideologie sterk beïnvloed te zijn door zijn eerdere interacties met rechtsextremistische gedachtegoed offline. Het internet leek in zijn geval dus niet de initiële facilitator voor zijn radicalisering. Het vergemakkelijkte echter wel zijn verdere ideologische ontwikkeling. Voor zowel Breivik en Mangs leek het internet daarentegen wel cruciaal om te kunnen komen tot gewelddadige actie. In dat opzicht lijkt het internet dus bij te hebben gedragen aan hun radicalisering, omdat het toegang verschaft tot informatie waartoe zij anders niet gemakkelijk toegang tot hadden. Hierbij moet echter opgemerkt worden dat dit vooralsnog niets zegt over de eventuele rol van algoritmen op sociale mediaplatforms in die radicaliseringsprocessen.

Wat betreft Roof was het internet belangrijk voor de ontwikkeling van zijn ideologie. Hij kwam hier namelijk voor het eerst in aanraking met radicale ideologie. Dit werd een katalysator voor een snelle radicalisering waarin Roof snel opging in racistische blogs en fora. Het internet bevorderde Roof's radicalisering door propagandateksten en websites gemakkelijk beschikbaar te maken, waardoor de basis voor radicalisering werd gelegd. Tevens bleek dat een enkele zeer specifieke Google-zoekopdracht zijn visie op de wereld veranderde (Miller-Idriss 2020). De zoekopdracht 'Black-on-White crime' leidde hem naar de white-supremacist groep 'Council of Conservative Citizens' die onder meer rechtsextremistisch gedachtegoed verkondigde. In het onderzoek van Mølmen & Ravndal (2021) worden geen duidelijke aanwijzingen gegeven voor een rol van sociale mediafouken. Wel bespreken de auteurs de rol van echokamers in het radicaliseringsproces. Zowel Breivik als Roof frequenteerde extremistische echokamers en beide uitten zich offline in dezelfde termen als gangbaar was in deze echokamers. Concluderend op basis van deze casussen zien we dus dat er aanwijzingen zijn voor het radicaliserende effect van (online) echokamers, maar dat de het radicaliserende effect van sociale mediafouken als gevolg van aanbevelingsalgoritmen op platforms vrijwel niet is vast te stellen (zie daarvoor ook: paragraaf 6.1).

### 6.3. Online en offline rekrutering

Om de onderzoeksvraag te kunnen beantwoorden of er gevallen bekend zijn waarbij naast online rekrutering sprake was van actieve offline rekrutering, is het zinvol om te kijken naar de wisselwerking tussen de online en offline contexten bij radicaliseringsprocessen. Deze wisselwerking kan worden onderverdeeld in twee niveaus waarop deze plaatsvindt: het *individuele niveau* en het *groepsniveau*. We merken hier op dat het huidige onderzoek naar de wisselwerking tussen de online en offline contexten bij radicaliseringsprocessen beperkt is en verdere studie behoeft (Von Behr et al., 2013; Winter et al., 2020).

#### Wisselwerking op het individuele niveau

Allereerst is het van belang aandacht te hebben voor de fase voordat iemand radicaliseert. In deze fase kunnen de offline omstandigheden van een persoon namelijk bijdragen aan diens vatbaarheid voor radicale invloeden in de online context, wat de kans op offline en/of online rekrutering vergroot. Centraal hierin staan de risico- en triggerfactoren die personen kwetsbaar maken voor radicalisering (voor een uitgebreide beschrijving hiervan, zie: paragraaf 2.2). Deze factoren worden in meer of mindere mate bepaald door de directe offline context van personen. Wanneer personen zich in de offline context onvoldoende begrepen voelen, een gebrek aan contact met gelijkgestemden ervaren of simpelweg de eigen normen en waarden willen verkennen, is een stap naar de online context aantrekkelijk en gemakkelijk. De extreemrechtse ex-activist Michel Letteboer schetst bijvoorbeeld de volgende situatie: *“Op school liep ik een charismatische leeftijdsgenoot tegen het lijf. Een uitstekende verteller. Iemand die net als ik hield van geschiedenis en politiek. Hij spiegelde me voor dat Nederland in verval was dankzij een linkse regering. En dat moslims de vijand waren”* (Platform JEP, g.d.). Op verzoek van deze leeftijdsgenoot bezocht Michel verscheidene websites waarop rechtsextremistische ideeën werden verspreid, wat zijn aandacht trok. Hij stelt dan ook dat dit online ‘wereldje’ erg aantrekkelijk kan zijn voor jongeren. *“Als je zoekende bent en je identiteit nog aan het vormen bent. Zo’n wereld geeft een gevoel van thuiskomen en*

*ergens bij horen”* (Nieuwsuur, 2021). Niet onbelangrijk is bovendien dat rechtsextremistische platforms of fora zich specifiek richten op deze ontvankelijke personen, die vaak door hun offline omstandigheden in een kwetsbare positie terechtkomen (Liang & Cross, 2020).

#### Wisselwerking op het groepsniveau

Naast wisselwerking op individueel niveau kan er sprake zijn van wisselwerking op groepsniveau – of anders gezegd: hoe komen de online en offline context samen op het niveau van rechtsextremistische groepen? Dit blijkt erg veelzijdig. In deze paragraaf focussen we daarom op de interactie tussen de aanwezigheid van rechtsextremistische groepen op sociale mediaplatforms (de online context) en de offline context. Activiteiten in de online en offline context liggen vaak in het verlengde van elkaar of gaan hand in hand. Dit laat de voortdurende wisselwerking tussen het online en offline domein goed zien. De activiteiten die we in deze literatuurstudie bespreken zijn: (a) organisatorische activiteiten; (b) netwerken; en (c) online propaganda gebaseerd op sociaal-politieke gebeurtenissen in de offline context.

A. *Organisatorisch: adverteren, faciliteren en mobiliseren.* Rechtsextremistische groepen maken gebruik van zowel de online en offline context om hun succes en maatschappelijke impact te vergroten. In een studie van Gaudette et al. (2020) is aan de hand van interviews met ex-rechtsextremisten gekeken naar de rol van het internet in het verbinden van het online en offline domein. Verscheidene geïnterviewden gaven daarin aan dat online fora, chatrooms en sociale mediaplatforms bij uitstek geschikte plaatsen zijn voor het grootschalig adverteren en promoten van fysieke rechtsextremistische ‘events’. Denk hierbij aan concerten, vergaderingen en gezamenlijke sportactiviteiten, maar ook aan demonstraties en protestacties. Een bekend praktijkvoorbeeld van online mobilisering in de offline context was de Unite The Right rally in Charlottesville (Virginia) in 2017. *“The rally was among the first real-world events that brought together many loosely affiliated far right communities of different ideological leanings”* (Davey & Ebner, 2017). In veel meer gevallen gaat het om meer bescheiden acties. Zo

gebruikt het Nederlandse Identitair Verzet Facebook om kleinschalige activiteiten te adverteren, van het houden van een borrel tot het beklimmen van het dak van een moskee (Sterkenburg, 2021). Een ander praktijkvoorbeeld van het faciliteren van 'events' is het digitaal verspreiden van flyers. Leden kunnen deze flyers zelf afdrukken en in hun eigen buurt ophangen (Gaudette et al., 2020).

- B. *Netwerken*. Het internet en sociale mediaplatforms bieden rechtsextremistische groepen en individuen de mogelijkheid om grotere netwerken op te bouwen en deze te faciliteren (Davey & Ebner, 2017). Deze netwerkfunctie komt daarbij gedeeltelijk voort uit de eerdergenoemde organisatorische functie van het internet. Zowel de mobiliserende kracht van het internet als de organisatorische en faciliterende kracht ervan zorgen ervoor dat al dan niet ideologisch verschillende rechtsextremistische groepen gemakkelijk met elkaar in contact komen en via deze weg hun invloed kunnen vergroten (Gaudette et al., 2020).
- C. *Online propaganda als gevolg van sociaal-politieke gebeurtenissen*. Sociale mediaplatforms bieden rechtsextremistische groepen de mogelijkheid om propagandistische boodschappen te verspreiden onder een breed publiek. Om daarin relevant te zijn, helpt het om in te spelen op actuele sociaal-politieke gebeurtenissen. Zichtbaar is dan ook dat terroristische aanslagen als tegenreactie vaak tot een aanzienlijke toename van online haatspraak ter bevordering van extremistische narratieven (Lally & Birmingham, 2020). Onderzoek van Bliuc et al. (2020) wijst bijvoorbeeld uit dat sociaal-politieke gebeurtenissen een transformerende werking hebben op (a) rechtsextremistische online netwerken; (b) diens collectieve opvattingen over wat het betekent om lid te zijn van een dit netwerk - "wie wij zijn als groep en wie onze bondgenoten en vijanden zijn"; en (c) de zorgen over actuele onderwerpen die relevant zijn voor een specifieke groep.

### **Online en offline rekrutering?**

Eerder in deze literatuurstudie zijn verschillende voorbeelden genoemd van online rekrutering (zie daarvoor: Hoofdstuk 3). Zichtbaar is dat het internet online rekrutering faciliteert en vergemakkelijkt – onder meer door de verbindende en mobiliserende kracht ervan (Marwick et al., 2022; ESS, 2021). Sociale mediaplatforms (zowel mainstream als alt-tech) worden door rechtsextremisten gebruikt om nieuwe aanwas te rekruteren. Ook gameplatforms dienen als plek om soms kwetsbare jongeren te vinden. In de wetenschap dat online rekrutering steeds vaker plaatsvindt, is het relevant om ook te kijken naar vormen van offline – of eigenlijk 'traditionele' – rekrutering. In de woorden van Save the Children (2021): *"To understand the entirety of the radicalisation process, we must consider both online and face-to-face interaction. The online environment forms only a fraction of the radicalisation process. It does not replace an individual's need to encounter people in person during the process, nor can one form a similar confidential relationship online"*. Gaudette et al. (2020) laten bijvoorbeeld zien dat extreemrechtse activisten het internet gebruiken om lokaal te zoeken naar gelijkgestemden en hiermee offline contact te leggen. Save the Children (2021) noemt daarnaast enkele andere vormen van offline rekrutering: het organiseren van openbare evenementen door rechtsextremisten om nieuwe leden te werven en het verspreiden van flyers in de openbare ruimte of in brievenbussen. Echter, in de praktijk komt naar voren dat personen vaak naar extremistische activiteiten worden gelokt via hun vriendenkringen, familieleden of hobby's – en niet zozeer via 'goede' ronselaars (expertinterview in: Save the Children, 2021). Daadwerkelijke rekrutering lijkt daarmee vaak gebaseerd op persoonlijke relaties, zoals ook het eerdergenoemde voorbeeld van Michel Letteboer laat zien. Wanneer we kijken naar de onderzoeksvraag over specifieke casussen van offline rekrutering (naast online rekrutering), is het zinvol om terug te grijpen naar de wisselwerking tussen de online en offline context. Daar waar radicalisering een proces is dat zich kan voltrekken in zowel de online- als offline context, zonder deze contexten daarin als afzonderlijk te begrijpen, is het aannemelijk dat rekrutering eveneens plaatsvindt in zowel de online als de offline context.

## Samenvatting Hoofdstuk 6

In dit hoofdstuk zijn onderzoeksvragen 6, 7 en 8 beantwoord: Zijn er gevallen bekend waarbij de rechtsextremistische sociale mediafuik een bijdrage heeft geleverd aan radicalisering en uiteindelijke rekrutering van bezoekers van sociale mediaplatforms? Op welke wijze droeg de sociale mediafuik bij aan rechtsextremistische radicalisering en eventueel aan rekrutering? Zijn er gevallen bekend waarbij naast online rekrutering sprake was van actieve offline rekrutering door rechtsextremisten en/of -terroristen?

Vanuit de wetenschappelijke literatuur is betrekkelijk weinig bekend over casuïstieken waarbij is vastgesteld dat een sociale mediafuik heeft bijgedragen aan rechtsextremistische radicalisering of rekrutering. Het vaststellen hiervan is ook niet eenvoudig omdat de werking van aanbevelingsalgoritmen en filterbubbels zich niet een geïsoleerd universum voltrekt. Er zijn immers ook andere online- en offline mechanismen die bij een persoon kunnen bijdragen aan zijn of haar radicalisering. Daadwerkelijke rechtsextremistische (gewelddadige) radicalisering door een sociale mediafuik is nog minder beschreven. Daarbij is het moeilijk te meten of iemand specifiek door de sociale mediafuik is geradicaliseerd of dat andere factoren een grotere bijdrage hebben geleverd. Het is nauwelijks mogelijk om te concluderen dat een individu uitsluitend is geradicaliseerd door consumptie van eenzijdige online content op sociale mediaplatforms en dat diegene zonder deze media niet geradicaliseerd zou zijn.

Desalniettemin zijn er anekdotische aanwijzingen (casussen) waarin aanneemelijk is dat de sociale mediafuik een rol heeft gespeeld in het radicaliseringsproces. Vraag is dan echter wel in hoeverre deze fuik veroorzaakt wordt door bijvoorbeeld aanbevelingsalgoritmen op sociale mediaplatforms, of dat er sprake is van een ogenschijnlijke fuik die veeleer te wijten is aan het eigen online gedrag (keuzes) van gebruikers. De casussen van online radicalisering geven desalniettemin wel aanwijzingen voor het gegeven dat het internet kan fungeren als echokamer, hetgeen een radicaliserend effect kan hebben op personen.

Wat betreft de onderzoeksvraag of er gevallen bekend zijn waarbij naast online rekrutering sprake was van actieve offline rekrutering door rechtsextremisten of -terroristen, is het van belang dat de online en offline contexten bij radicaliseringsprocessen niet als twee zelfstandige domeinen worden begrepen. Deze contexten beïnvloeden elkaar en liggen in het verlengde van elkaar. Het huidige onderzoek naar de wisselwerking tussen de online en offline contexten bij radicaliseringsprocessen is beperkt en behoeft verdere studie.

# 7 Nieuwe regelgeving en handelingsperspectieven

## Onderzoeksvragen die worden beantwoord in dit hoofdstuk:

1. Welke haalbare mogelijkheden zijn er (voor de overheid) om rechtsextremistische radicalisering via online sociale mediaplatforms tegen te gaan?

## In dit hoofdstuk komt het volgende aan bod:

- Nuances omtrent de rol van aanbevelingsalgoritmen op online sociale mediaplatforms bij rechtsextremistische radicalisering (Paragraaf 7.1).
- Moderatiebeleid door sociale mediaplatforms ten aanzien van aanbevelingsalgoritmen en content (Paragraaf 7.2).
- Sociale mediabeleid door Nederland en Europese Unie (wat wordt er reeds ondernomen en welke nieuwe mogelijkheden biedt de implementatie van nieuwe Europese wetgeving) (Paragraaf 7.3).
- Handelingsperspectieven om (online) rechtsextremistische radicalisering op sociale mediaplatforms tegen te kunnen gaan (Paragraaf 7.4).
- Handelingsperspectieven om (offline) rechtsextremistische radicalisering op sociale mediaplatforms tegen te kunnen gaan (Paragraaf 7.5).

## 7.1. Nuances omtrent de rol van aanbevelingsalgoritmen bij rechtsextremistische radicalisering

Om de vraag "Welke haalbare mogelijkheden er zijn voor de overheid om rechtsextremistische radicalisering via online mediaplatforms tegen te gaan?" te kunnen beantwoorden is het allereerst van belang om de context van ontwikkelingspaden voor rechtsextremistische radicalisering in een breder perspectief te plaatsen. Zoals we eerder in deze literatuurstudie hebben opgemerkt, voltrekt het mechanisme van een "algoritme gestuurd aanbod van rechtsextremistische content op sociale mediaplatforms dat kan leiden tot een rechtsextremistische sociale media fuik" zich allesbehalve in een vacuüm. Ook uit de interviews van de onderzoekers met experts op het gebied van de impact van sociale mediaplatforms komt naar voren dat experts voorzichtig zijn om aanbevelingsalgoritmen een grote rol toe te schrijven bij rechtsextremistische radicalisering. De geïnterviewde experts wijzen erop dat personen zowel online als offline op allerlei manieren in aanraking kunnen komen met polariserende of extremistische content, en geven hiervan een aantal concrete voorbeelden:

1. Mensen delen binnen hun sociale netwerk links naar extremistische content met elkaar.
2. Mensen abonneren / volgen bepaalde kanalen op sociale mediaplatforms en worden op die manier selectief geconfronteerd met meer extreme content.
3. Gelijkgestemden ontmoeten elkaar in echokamers zoals chatgroepen op Telegram.

Ingezoomd op de werking van aanbevelingsalgoritmen zelf, hoeft er niet altijd sprake te zijn van sociale mediafuiken waar personen min of meer onbewust worden ingezogen door de aanbevelingsalgoritmen. Het kan ook zijn dat mensen bewust, uit zichzelf op zoek zijn naar extremere content en hiertoe doelbewust bepaalde zoektermen invoeren, bepaalde namen zoeken of op de meer extreme aanbevelingen doorklikken. De aldus verworven kennis kan

vervolgens ook weer van invloed zijn op nieuwe zoektermen die een persoon toepast. De verscheidene persoonlijke keuzes die gebruikers maken zijn daarmee relevant voor het online ontwikkelingspad van radicalisering dat zij mogelijkwijs bewandelen. Daarnaast kunnen er vraagtekens worden geplaatst bij het belang van de online context zélf bij ontwikkelingspaden van (rechtsextremistische) radicalisering. Op basis van een review van wetenschappelijke literatuur concludeert Whittaker (2022) dat studies naar de rol van aanbevelingsalgoritmen enkel focussen op het online domein, met als gevolg de rol van de offline context hierbij onderschat wordt (p. 27):

*“There has been an increased policy concern about the role of recommendation algorithms in the radicalization process. However, it is important to note that thirteen of the fifteen studies focused their research exclusively on the online domain. While it is intuitive to study an online effect in its own environment, it runs the risk of inflating the role of the Internet in cases of terrorism or extremism. Research has consistently shown that despite terrorists utilizing the Internet heavily, this does not come at the expense of offline interactions and that offline social networks may play a greater role than the web. Baugut and Neumann consider both domains in relation to each other. They do suggest an important role for algorithms recommending propaganda, but also find that the online and offline domains are inseparably intertwined: “Contact with online propaganda was usually followed by personal talks with peers or preachers, and preachers used their personal talks with recruits as opportunities to show them media propaganda.” An over-emphasis on the online environment – including algorithms – may foster a “streetlight effect” in which research focuses on what is easily available and therefore miss the true picture of contemporary radicalization”.*

Er zijn tevens vraagtekens te plaatsen bij de rol *an sich* van internet en sociale mediaplatforms bij *gewelddadige* radicalisering. Zoals eerder gesteld is er weinig empirisch bewijs dat radicalisering in de online context leidt tot gewelddadig extremisme noch dat gewelddadig extremisme rechtstreeks voortvloeit uit online radicalisering. Er is altijd sprake van een wisselwerking met de offline context die personen ertoe aanzet over te gaan tot het gebruik van (politiek) geweld. Kortom, er zijn veel online én offline factoren en mechanismen die kunnen bijdragen aan rechtsextremistische radicalisering. De rol van aanbevelingsalgoritmen lijkt hierbij beperkt. Al met al betekent dit dat het tegengaan van rechtsextremistische radicalisering een aanpak vereist waarbij aan meerdere knoppen tegelijkertijd gedraaid moet worden om tot gewenste resultaten te kunnen komen. Echter, dit is verre van eenvoudig te realiseren. Als het bijvoorbeeld gaat om het aanpakken van extremistische content op online echokamers dan concluderen Peeters et al. (2022) dat de mogelijkheden hiervoor beperkt zijn behalve een justitiële aanpak van uitspraken die strafbaar zijn. Daar waar het trekken van een grens tussen strafbare en niet-strafbare content in het geval van laatstgenoemde evenmin gemakkelijk is, zoals bijvoorbeeld lastig is bij het gebruik van ironie of ‘humoristische’ teksten verpakt in afbeeldingen. Peeters en collega’s (2022) bevelen aan om vooral *offline* mogelijkheden te benutten, waaronder het versterken van de digitale weerbaarheid van gebruikers tegen desinformatie en complottheorieën en het vergroten van het inzicht in de online leefwereld onder gezagsdragers en (jeugd)professionals. Als het gaat om het modereren van extremistische en borderline content op sociale mediakanalen zijn de sociale mediaplatforms aan zet. Ook dit is problematisch: bijvoorbeeld doordat dit een aanzienlijke capaciteit vraagt van de inzet van moderators. Het verwijderen van geplaatste content op sociale mediaplatforms brengt bovendien grote ethische en praktische problemen met zich mee, want: hoe bepaalt men welke content over de schreef gaat? Contentmoderatie brengt immers complexe issues met zich mee omtrent de spanning tussen de ‘freedom of speech’ enerzijds en het bestrijden van discriminatie, racisme en extremisme anderzijds. In een Trends Alert schrijft het Counter-Terrorism Committee van de Verenigde Naties hier het volgende over:

*“Although this initiative has been commended by many States, some Governments, researchers and human rights institutions have raised concerns about its commitment to remove “terrorist and violent extremist content”, which may be defined in such a way as to unduly restrict human rights -including the rights to freedom of opinion and expression, privacy, and freedom of religion—and limit the freedoms of human rights defenders, journalists and civil society organizations” (CTED, 2020).*

De rol van (aanbevelings)algoritmen op sociale mediaplatforms bij rechtsextremistische radicalisering lijkt in de recente jaren nog verder te zijn beperkt door de aangescherpte contentmoderatie van de mainstream sociale media platforms; en te zijn vervangen door niet strafbare borderline content welke wel meer extreme ideeën kan voeden. Extreme content is weliswaar minder aanwezig op de mainstream platforms maar nog wel aanwezig op kleinere fringe platforms. Op deze platforms is echter niet of nauwelijks sprake van (aanbevelings)algoritmen. Dit wordt in paragraaf 7.2 nader toegelicht.

## 7.2. Moderatie van algoritmen en content door sociale mediaplatforms

Voordat er in dit hoofdstuk wordt gekeken naar mogelijke handelingsperspectieven (voor overheden) om rechtsextremistische radicalisering via online sociale mediaplatforms tegen te gaan, is het inzichtelijk om op hoofdlijnen stil te staan bij de vraag wat sociale mediaplatforms zelf al doen aan contentmoderatie. Welke vormen van moderatie worden door de platforms gehanteerd?

Moderatie van content door sociale mediaplatforms vindt plaats onder druk, bijvoorbeeld wanneer ophef ontstaat onder gebruikers, in de media of in de politiek. Of, en misschien wel vooral, omdat adverteerders niet geassocieerd willen worden met bepaalde content. In de loop van de jaren hebben sociale mediaplatforms daarom contentmoderatiebeleid ontwikkeld. Hierover zijn

ook passages opgenomen in de gebruikersvoorwaarden, welke we hieronder zullen bespreken voor YouTube en TikTok. Aanpassingen in de aanbevelingsalgoritmen van deze sociale mediaplatforms vinden daarentegen slechts in zeer beperkte mate plaats. Deze algoritmen op sociale mediaplatforms dienen in zekere zin, zoals eerder uiteengezet, als verdienmodel voor het mediabedrijf (Kim, 2017). Daar waar er wel sprake is van contentmoderatie op deze platforms, zullen we dit in deze paragraaf benoemen. Wat relevant is om hier in het achterhoofd te houden, is dat met name grote, populaire sociale mediaplatforms in hun gebruikersrichtlijnen vaak erg strikt zijn in het moderatiebeleid dat zij uitdragen, terwijl de daadwerkelijke uitvoering ervan niet noodzakelijk hiermee overeen hoeft te komen.

### Moderatiebeleid van YouTube

#### Aanbevelingsalgoritmen

Op de aanbevelingsalgoritmen op YouTube wordt ingegaan in paragraaf 4.2. YouTube maakt zich zorgen over de zogeheten *borderline* content, dat zijn video's die niet direct extremistisch zijn of volledig uit desinformatie bestaan, maar daar wel al trekken van hebben of gebruikers richting nog extremere content kunnen leiden. Video's die als 'borderline' worden aangemerkt door het platform, worden al zo veel mogelijk geweerd uit de algoritmen die video's aanraden aan gebruikers. Deze strategie werd vanaf 2017 een meer prominent onderdeel van het algoritme zelf, omdat YouTube gebruikers die actief zoeken naar extremistische content omleidt naar andere content (Chi, 2021). Overigens maakt YouTube zich nog steeds zorgen over de verspreiding van deze content. Het bedrijf ziet het delen van dit soort video's via weblinks en de deelknop als problematisch, omdat er op die manier nog steeds veel mensen bij zo'n filmpje kunnen uitkomen. Om die reden denkt YouTube nu na over een manier om dit aan te pakken, waaronder het uitschakelen van de deelknop en het onbruikbaar maken van de link naar een 'borderline' video, zo schrijft het in een blogpost (Mohan, 2022).

### Contentmoderatie

YouTube heeft beleid geïmplementeerd tegen de verspreiding van bepaalde vormen van misleidende evenals extremistische informatie (YouTube Help, g.d. (1))<sup>31</sup>. Belangrijk is om vervolgens ook vast te stellen welke vormen van misleidende of extremistische informatie niet onder het moderatiebeleid vallen. YouTube heeft beleidsregels voor gebruikers opgesteld als het gaat om spam en misleidende praktijken, zoals scam, nabootsing van identiteit; een verbod op 'haatzaaiende en intimiderende content' en content 'die anderen aanmoedigt om gewelddadige handelingen uit te voeren'; een verbod op het verspreiden van misleidende of bedrieglijke content en daardoor ernstige schade veroorzaken, zoals het 'verstoren van democratische processen'. Ook externe links naar websites met deze content wordt niet toegestaan (YouTube Help, g.d. (2)). Dit beleid is van toepassing op video's, audio, videobeschrijvingen, reacties, vastgezette reacties, livestreams en andere YouTube-producten en -functies. Gebruikers kunnen ook andere gebruikers rapporteren als content het beleid schendt, bijvoorbeeld door misinformatie te verspreiden of illegale inhoud te bevatten. Als wordt geconstateerd dat de content van gebruikers het beleid van het platform schendt, verwijdert YouTube de content en stuurt aan de gebruiker een e-mail om dit te laten weten. Als dit de eerste overtreding is, krijgt de overtreder meestal een melding, maar worden er nog geen maatregelen genomen tegen het account. Als het niet de eerste keer is, geeft YouTube een waarschuwing. Als een gebruiker drie waarschuwingen binnen 90 dagen krijgt, wordt het kanaal beëindigd.

31 Wij hebben de beleidsbeschrijving geraadpleegd van YouTube zoals opgenomen in het YouTube Helpcentrum.

YouTube heeft het beleid voor contentmoderatie in bepalingen verwoord. Daarbij is het niet moeilijk om de invloed van actualiteiten te herkennen, waaronder desinformatie rond coronamaatregelen en de crisis en desinformatie rondom de afgelopen Amerikaanse presidentsverkiezingen in 2020. Niet toegestane content op het platform is bijvoorbeeld:

- Gemanipuleerde media, waarin content zo technisch gemanipuleerd wordt dat er een groot risico op ernstige schade kan ontstaan onder gebruikers.
- Onjuist toegeschreven content, waarbij bijvoorbeeld verkeerde ondertiteling wordt toegevoegd aan beeldmateriaal of ten onrechte wordt beweerd dat oud beeldmateriaal van een eerdere gebeurtenis van een huidige gebeurtenis is.
- Content die in strijd is met consensus onder experts over bepaalde veilige medische praktijken, zoals misinformatie en complottheorieën over vaccins.
- Het verspreiden van misleidende informatie over verkiezingen. Zo is niet toegestaan om content te plaatsen waarin twijfels worden gezaaid over methoden om een stem uit te kunnen brengen en mag content gebruikers niet aansporen om democratische processen te verstoren. Het is tenslotte niet toegestaan om materiaal te plaatsen waarin ten onrechte wordt beweerd dat er sprake is van grootschalige fraude, fouten of technische storingen tijdens verkiezingen.

Begin 2022 heeft YouTube het beleid omtrent desinformatie aangescherpt (Mohan, 2022). Er is specifiek aandacht gekomen voor het op tijd signaleren van desinformatie, zodat deze niet viral kan gaan; het terugdringen van het delen van desinformatie tussen sociale mediaplatforms; en meer aandacht voor de aanpak van desinformatie wereldwijd. Het zogeheten '4-R plan' is gelanceerd. Content dient sneller te worden verwijderd (Remove), gebruikers dienen te worden geïnformeerd over mogelijke desinformatie in video's (Raise), het delen van zulke video's wordt gelimiteerd (Reduce) en gebruikers die zich aan de richtlijnen houden worden beloond (Reward). Opvallend is dat ook hier desinformatie omtrent de Coronapandemie als één van de kernredenen voor



het aanscherpen van het beleid naar voren komt. Stappen om de regulatie van desinformatie concreet te verbeteren zijn onder andere het ontwikkelen van een systeem dat misinformatie sneller herkent, het zoeken naar betere classificatiemiddelen – denk bijvoorbeeld aan het gebruik van bepaalde labels – en het toevoegen van nieuwe talen om regionale analyses uit te voeren (Mohan, 2022). Vaak neemt het *factchecken* van informatie uit video's echter veel tijd in beslag. Daarom overweegt YouTube om specifieke labels te creëren die aangeven dat een video geen hoge kwaliteit heeft, zodat gebruikers weten dat de video niet gecontroleerd is op desinformatie (Mohan, 2022).

YouTube heeft daarnaast beleid geformuleerd omtrent het aanzetten tot haat. Content die geweld of haat tegen groepen promoot op basis van, bijvoorbeeld, leeftijd, beperking, etnische achtergrond, religie of genderidentiteit, wordt verwijderd. Zo mogen groepen niet worden ontmenselijkt door ze te vergelijken met dieren of ziekten. Verder mag niet worden beweerd dat bepaalde groepen lichamelijk of geestelijk minderwaardig zijn. Gebruikers wordt opgeroepen om schendingen van dit verbod te melden bij YouTube. Hiervoor geldt ook dat bij drie waarschuwingen een kanaal verwijderd wordt. Soms worden video's niet direct verwijderd, maar wordt de spreiding ervan tegengegaan door het delen ervan tegen te gaan, de video niet langer aan te raden en geen nieuwe reacties onder de video toe te staan (YouTube Help, g.d. (3)).

Er is, tot slot, ook beleid opgesteld voor het aanpakken van het zogeheten "cross-platform" probleem. Hiermee wordt de verspreiding van een video die op YouTube is geplaatst naar andere platforms bedoeld (Mohan, 2022). Zo kan een video met *borderline* content vaker bekeken worden als gebruikers de weblink van deze video op andere sociale mediaplatforms delen. Potentiële mogelijkheden om dit tegen te gaan zijn bijvoorbeeld het uitzetten van de deelfunctie of het onmogelijk maken van het delen van een link. Dit wordt echter nog niet breed toegepast op YouTube, omdat het moeilijk is om de grens te bepalen tussen enerzijds het beperken van de verspreiding van *borderline* content en anderzijds het schenden van de vrijheden van gebruikers (Mohan, 2022).

## Moderatiebeleid van TikTok

### Aanbevelingsalgoritmen

Op de aanbevelingsalgoritmen op TikTok wordt ingegaan in paragraaf 4.2. Inspanningen op het modereren van ongewenste effecten van aanbevelingsalgoritmen zijn gebaseerd op het inzicht van TikTok (2022) dat deze de gebruikerservaring kunnen beperken: *"they can inadvertently limit your experience – what is sometimes referred to as a filter bubble"*. Deze algoritmen zijn volgens TikTok (2022) daarom als volgt geconfigureerd om een homogene stroom aan video's te beperken:

- A. Repetitieve patronen worden onderbroken (geen video's achtereenvolgend met dezelfde 'sound', inhoud of maker).
- B. Diversiteit wordt opzettelijk ingebracht (hierdoor krijgt een gebruiker óók video's aanbevolen die niet direct relateren aan diens interesses).
- C. Video's met een illegale inhoud worden verwijderen uit het aanbevelingssysteem.

### Contentmoderatie

TikTok heeft een op hoofdlijnen vergelijkbaar moderatiebeleid als YouTube. Het platform heeft communityrichtlijnen opgesteld met een reeks normen en gemeenschappelijke gedragscodes met als doel om *"een veilige en inclusieve omgeving voor onze gemeenschap te behouden"*. TikTok verwijderd content en accounts die deze communityrichtlijnen schenden. Zo is gewelddadig extremisme en haatdragend gedrag niet toegestaan en wordt deze verwijderd. Bij meerdere waarschuwingen worden gebruikers geblokkeerd. Gebruikers die content plaatsen waarin de openbare veiligheid wordt bedreigd of die oproepen tot geweld, worden ook geblokkeerd. Bovendien komt content die het platform niet geschikt acht voor een publiek ouder dan 13 jaar niet in aanmerking voor aanbeveling in de For You-feed. Videomakers worden op de hoogte gebracht van de reden voor de verwijdering en krijgen de mogelijkheid om bezwaar aan te tekenen tegen de verwijdering (TikTok, g.d.).

In de communityrichtlijnen is onder meer bepaald dat de volgende content wordt verwijderd van het platform (TikTok, 2023):

1. *Misbruik, beschadiging, in gevaar brengen of uitbuiting van minderjarigen.*
2. *Aanstootgevende uitingen*, waaronder dreiging of kleinerende uitspraken die bedoeld zijn om een persoon te bespotten, te vernederen, in verlegenheid te brengen, te intimideren of te kwetsen. Om uitdrukking te kunnen geven aan zaken van algemeen belang, kunnen kritische opmerkingen over publieke figuren worden toegestaan; ernstig aanstootgevend gedrag ten opzichte van publieke figuren is echter verboden.
3. *Haatzaaiend gedrag.* TikTok definieert haatdragende taal of gedrag als content die aanvallend of bedreigend van aard is, aanzet tot geweld of anderszins onmenselijk is voor een individu of een groep op basis van de volgende beschermde eigenschappen: Ras, Etniciteit, Nationale afkomst, Religie, Seksuele geaardheid, Geslacht, Gender, Genderidentiteit, Handicap, Immigratiestatus.
4. *Haatdragende ideologieën.* TikTok definieert haatdragende ideologieën als ideologieën die blijk geven van duidelijke vijandigheid t.o.v. mensen vanwege hun beschermde kenmerken. Voorbeelden hiervan zijn:
  - Content die haatdragende ideologieën prijst, promoot, verheerlijkt of ondersteunt (bijv. blanke suprematie, vrouwenhaat, anti-LHBTQIA+, antisemitisme).
  - Content die namen, symbolen, logo's, vlaggen, slogans, uniformen, gebaren, saluten, beelden, portretten, liederen, muziek, teksten of andere voorwerpen bevat die verband houden met een haatdragende ideologie.
  - Content die conversietherapie of verwante programma's promoot, ondersteunt of er reclame voor maakt.
  - Content die ontkent dat goed gedocumenteerde en gewelddadige gebeurtenissen hebben plaatsgevonden die gevolgen hebben voor groepen met beschermde kenmerken (bijv. Holocaustontkenning).

- Claims van suprematie over een groep mensen met betrekking tot andere beschermde kenmerken.
  - Complottheorieën die gebruikt worden om haatdragende ideologieën te rechtvaardigen.
5. *Gewelddadig extremisme:*
    - TikTok beschouwt het aanzetten tot geweld als het ondersteunen, aansturen of aanmoedigen van andere mensen om geweld te plegen. Het platform staat niet toe dat er wordt bedreigd met of wordt aangezet tot geweld dat kan leiden tot ernstig lichamelijk letsel.
    - Gewelddadige extremistische organisaties en personen: TikTok staat geen personen of organisaties toe die geweld promoten of zich met geweld bezighouden, waaronder terroristische organisaties, georganiseerde haatgroeperingen, criminele organisaties en andere niet-staatelijke gewapende groepen.

TikTok brengt meerdere malen per jaar een rapport uit over de handhaving van communityrichtlijnen, waaronder het verbod op verspreiding van desinformatie en aanzetten tot geweld. In dit rapport wordt een overzicht weergegeven hoeveel video's er zijn verwijderd en om welke reden dit gebeurd is. Het doel van het rapport is om gebruikers transparantie te geven in het werk dat wordt gedaan en tot welke verbeteringen dit heeft geleid (Keenan, 2022). De communityrichtlijnen worden regelmatig herzien om mee te evolueren met nieuw gedrag en nieuwe risico's (TikTok, g.d.). Om de communityrichtlijnen effectief te handhaven, zet TikTok een combinatie van technologie en moderators in. Video's die op TikTok zijn geüpload, worden allereerst beoordeeld door geautomatiseerde moderatietechnologie, die tot doel heeft inhoud te identificeren die de communityrichtlijnen schendt. Deze systemen kijken naar verschillende signalen in de inhoud, waaronder trefwoorden, afbeeldingen, titels, beschrijvingen en audio. Als er geen overtreding wordt vastgesteld, kan deze op het platform worden bekeken. Als er een mogelijke schending wordt gevonden, geeft het geautomatiseerde moderatiesysteem dit door aan veiligheidsteams voor verdere beoordeling of wordt deze automatisch verwijderd

als er een hoge mate van zekerheid bestaat dat de inhoud in strijd is met de communityrichtlijnen. Deze geautomatiseerde verwijdering wordt toegepast wanneer de overtredingen het duidelijkst zijn, zoals bij naaktfoto's of jeugdveiligheid. Om de beoordeling van potentieel schendende inhoud te ondersteunen, werken moderators samen met de geautomatiseerde moderatiesystemen en houden ze rekening met aanvullende context en nuance die niet altijd door technologie worden opgepikt (TikTok, g.d.). Menselijke beoordeling helpt ook bij het verbeteren van de geautomatiseerde moderatiesystemen door feedback te geven voor de onderliggende machine learning-modellen om toekomstige detectiemogelijkheden te versterken. Ter ondersteuning van dit werk is technologie ontwikkeld die risicovolle of verdachte items, bijvoorbeeld wapens, in videoframes kan identificeren, zodat inhoudsmoderators de video en de context waarin deze verschijnt, zorgvuldig kunnen beoordelen (TikTok, g.d.).

### **Moderatiebeleid van alternatieve platforms als Gab, 4chan, 8kun en BitChute**

#### *Gab*

Interessant is om het beleid omtrent contentmoderatie van alternatieve 'alt-tech' platforms te vergelijken met het moderatiebeleid van grote online sociale mediaplatforms als YouTube en TikTok. De contentmoderatie door bijvoorbeeld Gab verschilt namelijk aanzienlijk van YouTube en TikTok. Aangezien op het sociale mediaplatform Gab nauwelijks sprake is van aanbevelingsalgoritmen, ontbreekt hierop ook het beleid. Waar YouTube en TikTok een moderatiebeleid hebben opgesteld welke in ieder geval een deel van misleidende of extreme content te weren, is Gab vrijwel uitsluitend gericht op het borgen van de vrijheid van meningsuiting (Gab, 2023). Enkele soorten content zijn desalniettemin niet toegestaan:

- Content die in strijd is met de federale wetten van de Verenigde Staten.
- Content waarin minderjarigen worden uitgebuit of mishandeld.
- Spam, ongewilde advertenties en gelijksoortige content.
- Content die fysieke schade of offline intimidatie aanmoedigt.

Gab claimt hun best te doen om "ervoor te zorgen dat alle besluiten omtrent contentmoderatie en uitvoeringen van servicevoorwaarden gebruikers niet zullen straffen voor het gebruiken van hun door God gegeven recht om vrijuit te spreken" (Gab, 2023). Mocht een gebruiker desondanks de voorwaarden schenden, dan kan een account geschorst of verwijderd worden of kan de gebruiker aangeklaagd worden.

#### *4chan, 8kun en BitChute*

Vergelijkbare statements omtrent contentmoderatie treffen we aan op online 'alt-tech' platforms als 4chan, 8kun en BitChute (voor meer informatie over deze platforms, zie daarvoor Paragraaf 3.5). Het online platform 4chan stelt op haar webpagina (g.d.): *"You will not post any of the following: Troll posts, Racism, Anthropomorphic ("furry") pornography, Grotesque ("guro") images, Loli/shota pornography, Dubs or GET posts, including 'Roll for X' images"*. Aanvullend geeft het platform aan dat het gebruikers verboden wordt om persoonlijke informatie van individuen te posten of aan te vragen (zogeheten 'doxxing') of op te roepen tot of deel te nemen aan invasies (zogeheten 'raids'). Er is een meldingssysteem op het platform waarbinnen gebruikers zelf melding kunnen doen van berichten die in strijd zijn met de algemene regels van het platform. Overtreders van deze regels kunnen (tijdelijk) worden verbannen en hun berichten kunnen worden verwijderd, wat uitgevoerd wordt door "het moderatieteam van 4chan" Bovendien geldt dat gebruikers die zich negatief uitlaten over het beleid of de contentmoderatie van 4chan of moderators intimideren eveneens verwijderd kunnen worden van het platform (4chan, g.d.). Vraag blijft echter wel in hoeverre 4chan er daadwerkelijk in slaagt om dit moderatiebeleid te handhaven en hiertoe bereid te zijn. Gekeken naar de contentmoderatie op het platform 8Kun, schrijft het platform hierover het volgende: *"We follow the first amendment on 8kun, which has five parts: Freedom of religion, Freedom of speech, freedom of the press, freedom of assembly, and the right to redress grievances. Suppose your posts are continually deleted, they may not be protected by freedom of speech. The usual types of posts which are not protected are advertising posts, posts that advocate violence, and posts that*

*don't pass the dost test.*” Laatstgenoemde verwijst naar een richtlijn (opgesteld in de Verenigde Staten in 1986) die helpt om te bepalen of een visuele afbeelding van een minderjarige pornografisch is.

Tot slot, het platform BitChute heeft evenals YouTube en TikTok een vrij uitgebreide webpagina met daarop communityrichtlijnen. Hierin is een lijst opgenomen met ‘verboden content’ (BitChute, g.d.): geweld; dierenmishandeling; kindermishandeling of bedreiging; pesterijen; schadelijke activiteiten; aanzetten tot haat – welke van kracht is in het Verenigd Koninkrijk, de EU en de EEA; seksueel expliciete content; bedreigingen en aanzetten tot geweld; en terrorisme en/of gewelddadig extremisme. Wat betreft laatstgenoemde: BitChute houdt een lijst bij met ‘entiteiten’ (groepen) die volgens het platform in strijd zijn met deze richtlijnen en niet aanwezig mogen zijn op het platform. Personen, organisaties en andere ‘entiteiten’ die betrokken zijn bij bovenstaande ‘verboden content’ worden dan verwijderd van het platform. Het platform verwacht daarbij van haar gebruikers dat zij met behulp van rapportagetools content melden wanneer deze in strijd is met de community richtlijnen (BitChute, g.d.).

#### *Contentmoderatie met behulp van tools*

Contentmoderatie wordt niet alleen uitgevoerd door de sociale mediaplatforms zelf. Verschillende externe bedrijven bieden ‘tools’ aan voor websites en applicaties om (meer) grip te krijgen op de content op online platforms. Organisaties met websites of applicaties kunnen dergelijke ‘tools’ kopen, hetgeen hen helpt in het modereren van hun content. Contentmoderatie kan dan worden uitgevoerd met behulp van software, getrainde medewerkers of een hybride vorm van deze twee. Er zijn meerdere websites die dit aanbieden. Een voorbeeld daarvan is WebPurify. Dit bedrijf heeft als doel om contentmoderatie beschikbaar en betaalbaar te maken voor elke website of applicatie-eigenaar.

### 7.3. Sociale mediabeleid in Nederland en de Europese Unie

Om verspreiding van extremistische en borderline content op sociale mediaplatforms te voorkomen, worden er verschillende maatregelen genomen door de Europese Unie en de Nederlandse overheid. Het beleid in Nederland omtrent online mediaplatforms wordt hoofdzakelijk bepaald op Europees niveau. Aanpassingen van het Nederlandse beleid volgen de implementatie van nieuwe richtlijnen vanuit de Europese Unie. De laatste aanpassing dateert uit 2018, waarin al aandacht werd gegeven aan het tegengaan van online haatzaaien en extremisme. Zo werden media-instellingen verplicht om maatregelen te nemen om te voorkomen dat content aanzet tot geweld of haat naar bepaalde bevolkingsgroepen, of terrorisme uitlokt (Europese Commissie, 2018, p.72-76). Over wat deze maatregelen concreet moeten inhouden en wanneer precies een medium deze al dan niet implementeert, ontbreken dwingende, concrete voorschriften. In een Kamerbrief over contentmoderatie en de vrijheid van meningsuiting op online platvormen (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 2021), wordt gesteld dat gebruikers een civielrechtelijke overeenkomst aangaan met het platform waar zij gebruik van maken. Online platforms zijn vrij om te bepalen wat er wel en niet mag binnen hun platform. Ze mogen hierbij actie ondernemen als de gebruiker de voorwaarden van het contract doorbreekt door bijvoorbeeld een account te verwijderen. In de Kamerbrief zegt de overheid zich er van bewust te zijn dat er echo-kamers zijn waar mensen enkel met gelijkgezinden in gesprek gaan of dat mensen terecht kunnen komen in sociale media fuiken. Het uitsluiten van bepaalde politieke overtuigingen binnen online communities wordt echter toegestaan, zolang daar een legitieme reden voor is en er andere fora zijn waar ook andere meningen gedeeld kunnen worden (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 2021). De vrijheid van online platforms om hun eigen content te reguleren geldt echter niet voor strafbare content. Hieronder vallen onder andere kinderpornografie, terroristische inhoud en haatzaaien. Vanaf 7 juni 2022 zijn EU-lidstaten en dus ook Nederland verplicht om een autoriteit op te richten om de verspreiding van kinderporno en terroristische

online inhoud tegen te gaan. Wat betreft het tegengaan van haatzaaien is in Nederland het Meldpunt Internet Discriminatie (MiND) verantwoordelijk voor het onafhankelijk en juridisch toetsen of er sprake is van discriminatie. MiND is het landelijke meldpunt voor strafbare, discriminerende uitingen op internet. MiND toetst uitingen aan de hand van het Wetboek van Strafrecht en de relevante jurisprudentie. Als er daadwerkelijk sprake blijkt van strafbare discriminatie gaat MiND met de melding aan de slag. Na het ontvangen van een melding over een discriminerende uiting onderzoekt MiND allereerst of de uiting (nog) online staat. Vervolgens beoordeelt MiND of de uiting strafbaar is op basis van artikel 137c t/m 137e van het Wetboek van Strafrecht en de bijbehorende jurisprudentie. Als dit het geval blijkt, dan verzoekt MiND de beheerder of moderator van de betreffende website om de uiting te verwijderen. Als aan een (herhaald) verzoek geen gehoor wordt gegeven zal MiND een aangiftedossier opstellen voor het Openbaar Ministerie (OM). De balans tussen strikte handhaving en ongerechtvaardigde beperking van de vrijheid van meningsuiting is echter precair, waardoor er het niet altijd helder is wat wel en niet strafbaar is.

Wat betreft het aanpakken van de verspreiding van desinformatie, bestond in Nederland in de afgelopen jaren nog geen duidelijk beleid. Zolang desinformatie geen duidelijke aanwijsbare schadelijke gevolgen heeft, zoals het schaden van iemands reputatie (smaad), konden er geen juridische stappen worden genomen om dit tegen te gaan. Met de nieuwe Europese DSA wetgeving (zie: hieronder) worden deze regels aangescherpt.

### **Nieuwe Europese wetgeving: DSA, DMA en EMFA**

Momenteel wordt er nieuwe Europese wetgeving geïmplementeerd welke relevant is voor het beleid in Nederland omtrent online sociale mediaplatforms, namelijk: de Digital Services Act (DSA), de Digital Markets Act (DMA) en de European Media Freedom Act (EMFA). Deze wetgeving gaat in 2024 van kracht.

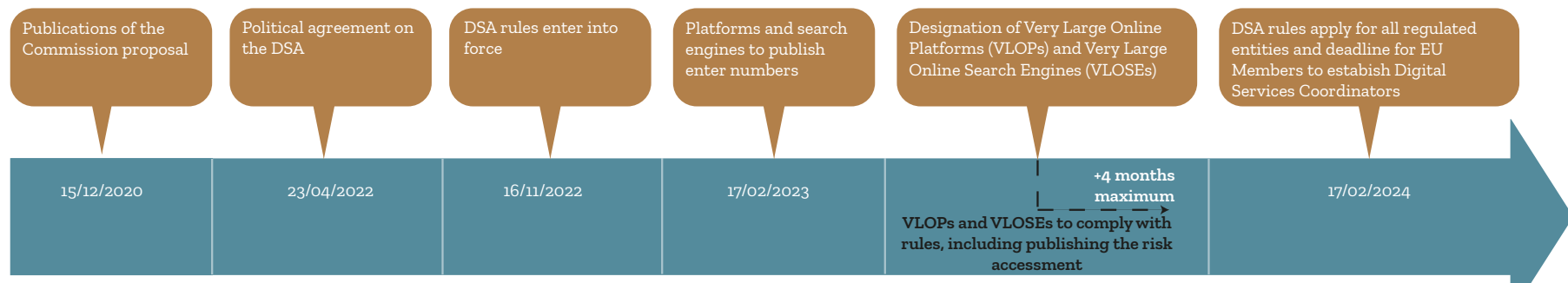
### *Digital Service Act (DSA)*

Met de verordening voor digitale diensten (Digital Services Act) komen er uitgebreide verplichtingen voor online platforms die consumenten in contact brengen met goederen, diensten of content (Europese Commissie, 2022). Met name op sociale mediaplatforms zal de impact hiervan groot zijn. Echter, ook intermediairs die netwerkinfrastructuur aanbieden, zoals internetproviders en hostingdiensten, krijgen ermee te maken. De Digital Services Act (DSA) beoogt schadelijke praktijken te beperken, online risico's tegen te gaan en gebruikersrechten online beter te beschermen. Alle online aanbieders van intermediaire diensten dienen zich te houden aan uitgebreide, nieuwe verplichtingen tot *transparantie*. Die moeten zorgen voor een sterkere verantwoordingsplicht en meer toezicht. Er komt meer beperking op de online verspreiding van illegale content en illegale producten. Daarnaast biedt de DSA ook de mogelijkheid om burgers beter te beschermen tegen desinformatie en nepnieuws en betere bescherming van grond- en mensenrechten. In welke mate online partijen met de nieuwe regels te maken krijgen, hangt af van hun rol, omvang en impact op het online ecosysteem. Voor de grote sociale mediaplatforms (als YouTube, Facebook, Instagram en TikTok) en zoekmachines met meer dan 45 miljoen gebruikers komen er meer verplichtingen. Deze online platforms en zoekmachines moeten namelijk (aanvullende) maatregelen nemen om risico's van hun dienstverlening voor onder andere de samenleving en de democratie te verminderen (Europese Commissie, 2022). Ook bevordert de DSA transparantie door door toegang tot data voor wetenschappelijke onderzoek beter te regelen, hierdoor krijgen we meer inzicht hoe deze platforms werken. De Europese Commissie bepaalt op basis van die gebruikersaantallen of een platform in de categorie 'zeer groot' thuishoort. Vervolgens hebben de platforms vier maanden de tijd om te voldoen aan de DSA-verplichtingen. Op 16 februari 2024 wordt de DSA volledig van toepassing voor alle diensten die daaronder vallen. In de onderstaande figuur is de tijdslijn zichtbaar van dit implementatieproces (Europese Commissie, 2023).

De Digital Service Act heeft als doel om de rechten van gebruikers te beschermen, transparantie en verantwoordelijkheid bij online platforms te verbeteren en groei binnen de markt te verbeteren (Europese Commissie, 2020). Op dit moment hebben overheden en gebruikers geen of weinig inzicht in de aanbevelingsalgoritmen die sociale mediaplatforms gebruiken, waardoor niet altijd nagegaan kan worden in hoeverre rechten van gebruikers beschermd zijn. Relevant is daarom de bepaling in de DSA die voorschrijft dat sociale mediaplatforms transparanter dienen te zijn richting haar gebruikers, overheden en onderzoekers. Dit vraagt bijvoorbeeld om transparantie over hun beleid omtrent contentmoderatie en het gebruik aan aanbevelingsalgoritmen. Samengevat geeft de DSA onder meer de volgende verplichtingen voor alle online platforms (Europese Commissie, 2022):

- Verplichting om ten minste een keer per jaar duidelijke, gemakkelijk te begrijpen en gedetailleerde rapporten te publiceren over eventuele inhoudsmoderatie die zij tijdens de betrokken periode hebben uitgevoerd (artikel 13).
- Verplichting om een intern klachtenafhandelingsstelsel in te voeren met betrekking tot beslissingen die worden genomen in verband met vermeende illegale inhoud of informatie die onverenigbaar is met hun algemene voorwaarden (artikel 17).
- Verplichting om ervoor te zorgen dat meldingen die worden gemaakt door zogeheten *trusted flaggers* met voorrang te behandelen (artikel 19).
- Onlineplatforms moeten de bevoegde rechtshandavingsinstanties informeren wanneer zij kennis krijgen van informatie die een vermoeden doet ontstaan van ernstige strafbare feiten (artikel 21).
- Onlineplatforms zijn verplicht om verslagen te publiceren over hun activiteiten met betrekking tot het verwijderen en ontoegankelijk maken van informatie die als illegale inhoud of in strijd met hun algemene voorwaarden wordt beschouwd (artikel 23).

## Timeline for Digital Services Act



Figuur 5. Tijdlijn van het implementatieproces van de Digital Services Act.

### Wat zijn 'trusted flaggers'?

Trusted flaggers zijn instanties of personen met bijzondere expertise en competentie voor het opsporen, identificeren en melden van illegale content of andere overtredingen van de wet op onder meer sociale mediaplatforms. De Digital Services Act (artikel 19) stelt dat online platforms meldingen van zogenaamde betrouwbare 'flaggers' met prioriteit moeten behandelen. Alleen 'entiteiten' die hebben aangetoond over deskundigheid en bekwaamheid te beschikken om illegale inhoud aan te pakken; collectieve belangen vertegenwoordigen; en onafhankelijk zijn van enig online platform kunnen de aanduiding van 'betrouwbare flagger' verkrijgen. Een entiteit kan de status van betrouwbare 'flagger' verkrijgen door dit aan te vragen bij de coördinator voor digitale diensten van de lidstaat waarin de aanvrager is gevestigd.

### Aanvullende verplichtingen voor de zeer grote platforms

Zeer grote online platforms met meer dan 45 miljoen gebruikers krijgen vanwege hun potentiële impact en de uitdagingen die gepaard gaan met een doeltreffend toezicht op deze platforms aanvullende verplichtingen opgelegd vanuit de DSA (Europese Commissie, 2022). Dit zijn onder meer verplichtingen op het gebied van risicobeheer, crisisrespons en transparantie. Concreet gaat dit voor de zeer grote sociale mediaplatforms om de volgende verplichtingen:

1. *Risicobeheer.* Zeer grote online platforms zijn verplicht om risicobeoordelingen uit te voeren met betrekking tot de systeemrisico's die voortvloeien uit of verband houden met de werking en het gebruik van hun diensten (artikel 26). Als platforms een risico vaststellen dan zijn zij verplicht om maatregelen te nemen ter vermindering van deze risico's (artikel 27). Het aanpassen van aanbevelingsalgoritmen kan hiervan een onderdeel zijn.
2. *Auditing.* Naast de zelfbeoordeling door de platforms geldt een verplichte audit van systeemrisico's. Zeer grote online platforms zijn verplicht om zich op eigen kosten aan externe en onafhankelijke controles te onderwerpen (artikel 28).

3. *Transparantie naar onderzoekers.* De zeer grote onlineplatforms moeten transparant zijn over een verscheidenheid aan zaken, waaronder het gebruik van aanbevelingsalgoritmen. Zij moeten door de Europese Commissie gemachtigde onderzoekers toegang geven tot belangrijke data, waaronder de aanbevelingsalgoritmen en deze uitleggen. Verder dienen onderzoekers toegang te krijgen om controles uit te voeren op de naleving van de verordening. Onderzoek naar de evolutie, werking en risico's van onlinesysteemrisico's zijn hierbij vooral belangrijk. Platforms mogen deze onderzoekers niet in hun werk beletten en moeten vrijwillig toegang verlenen voor deze onderzoekers. Dit maakt het vervolgens voor de EU beter mogelijk om bindende afspraken te maken over aanpassingen in de aanbevelingsalgoritmen opdat het aantal aanbevelingen voor schadelijke content terug kan worden gedrongen (artikel 31).
4. *Transparantie naar gebruikers.* Zeer grote onlineplatforms die gebruikmaken van aanbevelingssystemen, vermelden in hun algemene voorwaarden op een duidelijke, toegankelijke en gemakkelijk te begrijpen wijze de belangrijkste parameters die in hun aanbevelingssystemen worden gebruikt, alsook eventuele opties voor de afnemers van de dienst om deze beschikbaar gestelde belangrijkste parameters te wijzigen of te beïnvloeden. Ook verstrekken zij op hun online-interface een gemakkelijk toegankelijke functionaliteit waarmee de afnemer van de dienst zijn voorkeursoptie voor elk van de aanbevelingssystemen die de relatieve volgorde bepalen van informatie die hij gepresenteerd krijgt, te allen tijde kan selecteren en wijzigen (artikel 29).
5. *Openbare transparantie.* Zeer grote online platforms dienen een jaarlijks verslag uit te brengen over content moderatie, waarin onder meer de handhaving van hun algemene voorwaarden wordt besproken (artikel 33). Platforms dienen de Commissie in te lichten bij het doorvoeren van veranderingen, dienen een risicobeoordeling uit te voeren en hier verslag van te doen naar de Commissie.

Het toezicht op de zeer grote platforms zal worden uitgevoerd door de Europese Commissie. Daarnaast wijzen de EU-lidstaten bevoegde autoriteiten ('digitale-dienstencoördinatoren') aan die verantwoordelijk zullen zijn voor het toezicht op kleinere sociale mediaplatforms (Europese Commissie, 2022).

#### *Digital Markets Act (DMA)*

Per november 2022 is eveneens de Digital Markets Act (DMA) in werking getreden. De DMA is met name gericht op het reguleren van de digitale markt. In de DMA worden toezicht en maatregelen geïntroduceerd voor de wereldwijd opererende tien tot vijftien grootste online platforms met een zogenoemde poortwachterspositie. Dat zijn platforms waar gebruikers nauwelijks omheen kunnen. De DMA beoogt dat deze "poortwachters" zich op een transparante en eerlijke manier gedragen. De nieuwe Europese regels beogen bij te dragen aan een eerlijker speelveld op de digitale markt en stellen kaders aan spelers, ook die van buiten de EU. Dat gebeurt door strenger toezicht en vooraf ingrijpen. Deze groep grote online platforms moet zich ook aan diverse verboden en verplichtingen houden. Voorbeelden hiervan zijn een verbod op het bevoorraden van eigen diensten, de mogelijkheid om data van een platform naar een ander platform mee te nemen en de mogelijkheid vooraf geïnstalleerde apps te verwijderen. Verder komt er een meldingsplicht om fusies en overnames te kunnen beoordelen (Rijksoverheid, 2021). Mochten platforms zich niet aan de regelgeving houden, dan kunnen zij een boete krijgen die op kan lopen tot 10% van de jaarlijkse inkomsten van het bedrijf in kwestie en zelfs 20% procent als het platform herhaaldelijk de regels overtreedt. Hoewel de DMA al in 2020 goedgekeurd is door de Europese Raad, zal deze pas geïmplementeerd worden vanaf mei 2023. Vanaf dat moment zal de Europese Commissie met de grote platforms het beleid nader uitwerken en aanscherpen. Na zes maanden dienen de platforms aan het beleid te voldoen (Europese Commissie, 2020b; Europese Commissie, 2022).

Een derde regelgeving die per 2024 van kracht gaat, is de European Media Freedom Act (EMFA). Het doel van de EMFA is om mediavrijheid en pluralisme in de EU te waarborgen (Europese Commissie, 2022b). De belangrijkste doelen van dit beleid zijn de volgende:

- Geen interferentie met de besluitvorming van editors in de media.
- Geen spionage software tegen journalisten.
- Onafhankelijke en adequate financiële ondersteuning van publieke media.
- Transparantie omtrent de eigenaar.
- Transparantie en eerlijkheid omtrent dataverzamelingen van gebruikers en allocatie van advertenties van de staat.
- Het opzetten van een Europese Raad voor Media Services.
- Uitbreiding van bescherming tegen media die ten onrechte wordt verwijderd.
- Onderzoek naar de markt en maatregelen die nationale media beïnvloeden.

De EMFA is goedgekeurd op 16 september 2022, maar de implementatie zal tot begin 2024 duren (Europese Commissie, 2022b). Binnen de EMFA is er ook aandacht voor het aanpakken van desinformatie en het creëren van een gezond en divers medialandschap (Europese Commissie, 2022c). Gekeken zal worden naar mondiale mediaplatforms als Facebook, YouTube en TikTok waarop vaak sprake is van polariserende content of content die desinformatie bevat (Europese Commissie, 2022c). Verder wordt het belang van de nieuwe mediawet benadrukt om de integriteit van de interne mediamarkt te behouden. Zo zijn er momenteel aanbieders die zich systematisch bezighouden met het verspreiden van desinformatie, informatiemanipulatie en inmenging. Deze mediadiensten staan in sommige gevallen zelfs onder overheidscontrole of worden gefinancierd door andere staten (Europese Commissie, 2022c).



Met de implementatie van de EMFA beoogt de Europese Commissie dat Europese burgers toegang behouden tot media van hoge kwaliteit. Europese staten worden gevraagd in samenwerking met journalisten en redacteuren journalistieke normen op te stellen, waarin betrouwbare en actuele informatie op de voorgrond staat (Europese Commissie, 2022c). Om verspreiding van desinformatie te voorkomen, dienen de afgevaardigden van EU landen minstens een maal per jaar de dialoog aan te gaan met 34 zeer grote mediaplatforms als Facebook, YouTube en TikTok omtrent de verspreiding van desinformatie (Europese Commissie, 2022c). Er wordt dan nagegaan in hoeverre deze platforms zelfregulerende initiatieven toepassen die gebruikers beschermen tegen schadelijke inhoud. Enkele voorbeelden zijn het verbreden van de participatie, het wegnemen van financiële motieven om desinformatie te verspreiden, het ontdoen van het platform van nepaccounts, robots en *deep fakes* met negatieve intenties, het verlenen van toegang aan onderzoekers door hen toegang te geven tot data van dergelijke platforms en het opzetten van een *Transparency Centre and Task Force*. Deze vernieuwde richtlijn is gebaseerd op evaluatie van eerdere richtlijnen en heeft als doel om de risico's, die voortkomen uit de verspreiding van misinformatie, te verminderen. Verder veranderen er ook dingen voor gebruikers. Zo zullen *factcheckers* beschikbaar worden in meer talen en komen er meer mogelijkheden voor gebruikers om desinformatie aan te geven en geattendeerd te worden op mogelijke desinformatie.

#### 7.4. Handelingsperspectieven online

In deze paragraaf bespreken we mogelijkheden voor de Nederlandse overheid om bij te dragen aan het verminderen en/of tegengaan van rechtsextremistische radicalisering op online sociale mediaplatforms. Deze handelingsperspectieven zijn gebaseerd op zowel de literatuurstudie, de diepte-interviews en online sessies die zijn gevoerd met relevante experts als op de uitkomsten van de trials. Niet alle handelingsperspectieven die worden beschreven zijn eenvoudig of op korte termijn te realiseren. In veel gevallen is hierbij een lange adem vereist.

De handelingsperspectieven richten zich op het aanpakken van de volgende problemen:

1. Borderline content lijkt grotendeels de plaats van extremere content te hebben ingenomen op de mainstream sociale mediaplatforms. Er is sprake van tekortschietende moderatie van borderline content op de grote sociale mediaplatforms. Er bestaan onduidelijkheden over wat er onder borderline content moet worden verstaan.
2. Er is sprake van een gebrek aan transparantie over de werking van aanbevelingsalgoritmen op de grote sociale mediaplatforms.
3. Er is sprake van tekortschietende moderatie van extreme content op alt-tech mediaplatforms.
4. Het zoek/surfgedrag van gebruikers zelf speelt een belangrijke rol bij blootstelling aan radicale content.

Goed om hierbij te noemen, is dat de handelingsperspectieven van de Nederlandse overheid richting (zeer) grote sociale mediaplatforms en/of andere online platforms met de hoofdvestiging in het buitenland beperkt zijn. Immers, de Europese Commissie heeft slechts de rol van toezichthouder in respectievelijk het land van vestiging. Wat echter wel mogelijk is voor de Nederlandse overheid zijn de volgende handelingsperspectieven (per handelingsperspectief wordt aangegeven welk probleem wordt geadresseerd):

- Het ondersteunen van zogeheten betrouwbare 'flaggers'. Dit zijn erkende, deskundige melders van illegale content op online platforms (probleem 1).
- Het uitoefenen van invloed op de Europese Commissie bij de implementatie van de DSA-wetgeving. Bijvoorbeeld om toezicht of handhaving te prioriteren (problemen 1, 2,3).

Andere handelingsperspectieven (nb: deze worden toegelicht in de volgende alinea's) die mogelijk zijn ten aanzien van minder grote sociale mediaplatforms, zijn bijvoorbeeld:

- Dialoog met platforms over het sneller verwijderen van illegale content (probleem 3).
- Investeren in de samenwerking tussen overheden, online platforms en onderzoekers (problemen 1, 2, 3).
- Coördinatie tussen ministeries en de continue opbouw van kennis (zoals over algoritmen, moderatie, *borderline* content e.d.) (problemen 1, 2, 3).
- Toewerken naar meer co-regulering tussen mediaplatforms en de overheid (problemen 1, 2).
- Beter begrijpen van de problematiek rondom *borderline* content op mediaplatforms; en meer duidelijkheid creëren bij platforms over aanpak van deze *borderline* content (probleem 1).
- Onderzoek naar- en vergroten bewustzijn omtrent de rol van influencers bij toegang tot extremere narratieven (problemen 1, 3, 4).
- Bevorderen van contentmoderatie in de Nederlandse taal op sociale mediaplatforms (probleem 1).
- Het inzetten op discussies omtrent het belang van 'publieke waarden' (naast economische waarden) in het beleid richting sociale mediaplatforms en creëren van randvoorwaarden om hiermee te kunnen experimenteren (problemen 1, 2).
- Mogelijkheden onderzoeken of steun geven aan bestaande initiatieven voor een online ecosysteem van kleinere, lokale sociale mediaplatforms (problemen 1, 2).

Er kan tevens door de Nederlandse overheid ingezet worden op offline aanpakken om radicalisering op sociale mediaplatforms tegen te gaan. Concrete voorbeelden hiervan zijn het versterken van kritisch denken en digitale weerbaarheid bij jongeren, het creëren van meer inzicht in de online leefwereld van personen en het eigen gedrag hieromtrent (probleem 4), het betrekken van jeugdwerkers als betrouwbare 'flaggers' en het bevorderen van bewustwording bij werkgevers omtrent het belang van vergroten van online vaardigheden van jeugdwerkers. Deze handelingsperspectieven worden in paragraaf 7.5 nader toegelicht.

## Zeer grote sociale mediaplatforms

De Digital Services Act (DSA), de Digital Markets Act (DMA) en de European Media Freedom Act (EMFA) als onderdeel van nieuwe Europese wetgeving biedt aanzienlijk meer mogelijkheden dan in het verleden voor overheden om met sociale mediaplatforms bindende afspraken te maken over transparantie, contentmoderatie van desinformatie en het aanpassen van aanbevelingsalgoritmen indien die resulteren in onwenselijke uitkomsten. Voor zeer grote sociale mediaplatforms met meer dan 45 miljoen gebruikers (als YouTube, Facebook, TikTok en Instagram) is de Europese Commissie als toezichthouder aan zet. De rol en mogelijkheden van de Nederlandse overheid ten aanzien van de zeer grote sociale mediaplatforms is daardoor echter beperkt. Wat is er wel mogelijk?

- Ondersteuning van 'trusted flaggers' (bijvoorbeeld in financieel opzicht) met bijzondere expertise en competentie voor het opsporen, identificeren en melden van illegale content of andere overtredingen van de wet op onder meer sociale mediaplatforms.
- Het implementeren van de DSA impliceert dat er in de praktijk mee gewerkt gaat worden en er via deze weg ervaringen worden opgedaan. Dit betekent dat de Nederlandse overheid gedurende deze praktijktoets mogelijkwel invloed kan uitoefenen op de Europese Commissie. Denk bijvoorbeeld aan het stellen van bepaalde prioriteiten aan het toezicht of in de handhaving, maar ook dat er bepaalde normen of definities in de DSA gebruikt worden (online platforms, overheden, betrouwbare 'flaggers') die in het licht van de werkpraktijk vragen om verduidelijking om daar een goed toezicht op te kunnen houden.
- In de *European Board* zitten de nationale toezichthouders, waaronder de toezichthouder van Nederland. Het is belangrijk hoe er straks toezicht gehouden gaat worden op de correcte naleving van de DSA; iets waarover tot dusver onduidelijkheid lijkt te bestaan, maar vanzelfsprekend cruciaal is voor het resultaat. Gegeven dat de DSA-wetgeving zich procesmatig in de fase bevindt waarin alles in de steigers gezet gaat worden om het te implementeren in de praktijk, is het belangrijk om in deze fase hierop actief

te zijn – ook en misschien wel juist als nationale overheid. Ten aanzien van de due-diligence<sup>32</sup> verplichting voor onder meer sociale mediaplatforms is het bovendien relevant om als overheid hierop zicht te houden. Via die weg zullen wij als samenleving bijvoorbeeld meer inzicht krijgen in hoe deze platforms omgaan met de bescherming van publieke waarden. Ook zullen de dilemma's waarmee platforms te maken hebben inzichtelijker worden, wat een belangrijke eerste stap is in de richting van duurzame aanpak en oplossingen.

## Overige (kleinere) sociale mediaplatforms

Wat betreft de rol en mogelijkheden van de Nederlandse overheid ten aanzien van de overige (vaak kleinere) online sociale mediaplatforms geldt het volgende: iedere Europese lidstaat wijst één of meer bevoegde autoriteiten aan als verantwoordelijk voor de toepassing van de Digital Services Act. Eén van deze autoriteiten doet dienst als *coördinator voor digitale diensten*. Deze coördinator is verantwoordelijk voor de kwesties die verband houden met de handhaving en toepassing van de wet. In Nederland zal deze verantwoordelijkheid naar verwachting gegeven worden aan de Autoriteit Consument en Markt (ACM). Ook de Autoriteit Persoonsgegevens gaat hier een rol in spelen. In ieder EU-land heeft de coördinator voor digitale diensten als taak om platforms te onderzoeken op de naleving van de DSA en hebben hierbij zowel onderzoeks- als handhavingsbevoegdheden. Dit wordt nader gespecificeerd in de *Regeling bevoegdheden onderzoek en handhaving aanbevelingsalgoritmen*.

<sup>32</sup> Zie: <https://www.rvo.nl/blogs/due-diligence-6-stappen> voor een toelichting op de 'due-diligence' verplichting.

Onderzoek bevoegdheid	Handhavingsbevoegdheid
Verzoek aan een online platform om informatie voor controles binnen een redelijk termijn te verstrekken. Ook mag de coördinator anderen in de branche die waarschijnlijk meer weten over het te onderzoeken platform om informatie vragen.	De door platforms gedane toezegging met betrekking tot de naleving van de wet aanvaarden en bindend maken.
De coördinator mag inspecties ter plaatse uitvoeren of andere openbare autoriteiten dit laten doen. Hierbij mag de coördinator kopieën (laten) maken van de nodige informatie en de gegevens in beslag nemen. Dit mag zowel voor fysieke als digitale gegevens.	Bevelen dat het platform stopt met de overtreiding. Hiervoor mag coördinator maatregelen opleggen.
Werknemers en vertegenwoordigers van het platform vragen om uitleg bij de opgevraagde informatie.	Boetes en dwangsommen opleggen.
Algemene bevoegdheid.	De coördinator mag zelf tijdelijke maatregelen nemen om gevaar van 'ernstige schade' te voorkomen.

Bron: Seggelen (2022)

Nota bene: de beoogde coördinator voor digitale diensten (in Nederland dus de ACM) mag alleen optreden tegen online platforms waarvan de hoofdvestiging of wettelijke vertegenwoordiger ervan is gevestigd in de eigen lidstaat. Als het hoofdkantoor in een ander land staat, dan is het aan de bevoegde autoriteit in dat land om te zorgen dat het online platform zich aan de wet houdt. Stel nu dat het online platform alleen de DSA in Nederland overtreedt. Dan moet de coördinator de autoriteiten in het desbetreffende land van vestiging vragen om een onderzoek. Dit land zal dan binnen twee maanden na ontvangst van het verzoek de kwestie moeten hebben onderzocht. Als dit niet op tijd gebeurt, kan Nederland de kwestie bij de Europese Commissie voorleggen. Deze zal de kwestie bespreken met de autoriteiten van het betreffende land, of direct doorzetten naar de Europese Raad voor digitale diensten (Seggelen, 2022).

Hieronder worden verschillende handelingsperspectieven gegeven voor de Nederlandse overheid om (rechtsextremistische) radicalisering op sociale mediaplatforms tegen te gaan:

- *Naast grote platforms, ook aandacht besteden aan content moderatie door minder grote platforms.* De DSA lijkt vooral gericht op de hele grote platforms, terwijl er ook risico's bestaan bij de minder grote en kleine platforms en deze daarmee eveneens vragen om aandacht. Het verwijderen van illegale content is vaak op kleinere platforms een proces dat langzaam verloopt. Een mogelijke aanbeveling ter verbetering is dat vooral de kleinere platforms meer zouden kunnen doen om sneller illegale content te verwijderen, en de overheid zou hierin ook een rol in moeten spelen. De DSA lijkt mogelijkheden te bieden voor de beoogde coördinator (ACM) om hierover in gesprek te gaan met kleinere online platforms, in het bijzonder de in Nederland gevestigde online platforms. Gekeken naar de relatief kleinere platforms vereisen bijvoorbeeld gameplatforms (maar, niet uitsluitend) bovendien bijzondere aandacht als het gaat om de aanpak van online radicalisering. Op gameplatforms kunnen jongeren in aanraking komen met extreemrechtse content verpakt als spel of in interactie met medespelers waarbij ook rekrutering kan plaatsvinden (zie: paragraaf 3.6).
- *Dialogo aangaan met alt-tech platforms.* Op alt-tech platforms met veel extreme content (zoals genoemd in paragraaf 7.2) vindt weinig of geen contentmoderatie plaats. De Nederlandse overheid kan de mogelijkheden verkennen om met deze platforms in dialoog te gaan, of althans de Europese overheid tot deze dialoog te stimuleren, om meer in te zetten het verwijderen van illegale content.
- *Investeren in samenwerking tussen overheden, online platforms en onderzoekers.* Hoewel de DSA aandacht besteedt aan meer transparantie, is meer transparantie slechts een eerste stap in de goede richting: er is meer nodig dan transparantie alleen. Bijvoorbeeld, om als overheid een beter beeld te krijgen van de problematiek rond extremistische- en borderline content, is het aan te bevelen dat zowel overheden als platforms investeren in een goede samenwerking met onderzoekers.

- *Coördinatie tussen ministeries en continue opbouw van kennis.* Het up-to-date-houden van kennis omtrent online platforms, algoritmen en contentmoderatie is belangrijk en vergt afstemming tussen ministeries. Liever geen losstaande onderzoeken maar deze met elkaar in verband brengen (continue kennisopbouw).
- *Toewerken naar meer co-regulering.* Een handelingsperspectief voor de overheid is om meer toe te werken naar co-regulering met online platforms. In het huidige beleid wordt veel overgelaten aan de online platforms zelf om extreme – en borderline content en aanbevelingsalgoritmen aan te pakken, terwijl beleidsmakers tegelijkertijd opperen dat platforms meer zouden kunnen doen (Whittaker et al. 2021). De vooronderstelling is hierbij dat online platforms vanwege commerciële redenen niet erg geneigd zijn om uit zichzelf een voortvarende aanpak te implementeren voor moderatie van aanbevelingsalgoritmen en extreme – en borderline content. En de wetgeving van de overheid loopt in principe altijd achter op nieuwe technische en maatschappelijke ontwikkelingen, waardoor regulatie vanuit de overheid altijd achter de feiten aan zal lopen. Meer samenwerking is derhalve gewenst. Een mogelijk handelingsperspectief voor de overheid is dan ook om meer in te zetten op ‘co-regulering’ waarbij de overheid en online platforms samenwerken om problemen aan te pakken. Hoewel de betrokken partijen compromissen zullen moeten sluiten, kan zo de verantwoordelijkheid worden gedeeld en nieuwe oplossingen worden aangedragen vanuit de industrie (Whittaker et al. 2021).
- *Meer duidelijkheid verschaffen over borderline content: ‘legal yet harmful’ content.* Hoewel de Digital Services Act aandacht besteedt aan transparantie, zoals over het gebruik van algoritmen, zijn er momenteel nog weinig handvatten omtrent de omgang met borderline content (ook wel aangeduid als ‘legal yet harmful’ content). Het is ingewikkeld voor online platforms als de overheid zegt dat zij borderline content moeten aanpakken, terwijl duidelijkheid ontbreekt over wat precies aangepakt moet worden. De overheid zou kunnen investeren in het beter begrijpen van de problematiek rondom *borderline* content en faciliteren dat hierover

meer duidelijkheid komt. Aangezien de overheid zelf terughoudend dient te zijn in het beoordelen van ‘harmful content’ is hier primair een rol weggelegd voor ‘trusted flaggers’ met kennis van zake over mogelijke schadelijke effecten en illegaliteit van border line content. Een groot probleem met de aanpak van borderline content is immers de beoordeling wanneer deze content illegaal is. Eerder in het rapport wezen we op de complicerende rol van bijvoorbeeld beeldspraak en ironie waarin extremistische content wordt verpakt. ‘Freedom of speech’ is echter niet hetzelfde als ‘freedom of reach’. Online platforms kunnen zulke borderline content weliswaar niet verbieden maar ze kunnen wel de amplificatie ervan belemmeren, wat ze in de praktijk overigens ook al doen. Als betrouwbare flaggers meer duidelijkheid scheppen over de schadelijkheid van bepaalde borderline content, dan zal deze vorm van contentmoderatie door online platforms makkelijker ingezet kunnen worden.

- *Bevorderen van contentmoderatie in de Nederlandse taal.* Om te voorkomen contentmoderatie minder effectief is bij niet-Engelstalig content is een mogelijk handelsperspectief voor de overheid om hierover in gesprek te gaan met de online platforms, door om transparantie te vragen hoe hiermee wordt omgegaan en om hen zonodig te stimuleren hier meer aandacht aan te geven.
- *Influencers.* Online zijn talrijke influencers die via grote sociale mediaplatforms of gameplatforms een ingang kunnen zijn voor (jonge) personen om zich online verder te verdiepen in bepaald gedachtegoed of subculturen die een ‘gateway’ kunnen zijn naar rechtsextremistisch gedachtegoed (zie paragraaf 3.7). Optie kan zijn voor de overheid om de invloedrijke rol van influencers (en de risico’s van deze invloed) nader wetenschappelijk te onderzoeken, meer bewustzijn te creëren rondom dit thema en dit tevens te bespreken met de platforms.

- *Centraalstellen van publieke waarden.* Een punt dat in de gesprekken met experts meermalen is genoemd is dat het beleid omtrent content op sociale media niet alleen gebaseerd moet zijn op economische – en juridische uitgangspunten maar ook op basis van het borgen van publieke waarden. Hier zou meer discussie over moeten zijn en ruimte en randvoorwaarden worden gecreëerd om te experimenteren. De overheid zou dit kunnen faciliteren. Een voorbeeld waarnaar kan worden gekeken is de Stichting PublicSpaces, dat is een coalitie van verschillende publieke organisaties, zoals publieke omroepen, erfgoed-instellingen, festivals, bibliotheken, musea en onderwijs. Onder de paraplu van PublicSpaces experimenteren deze organisaties met oplossingen voor een gemeenschappelijk probleem: voor hun communicatie, informatie en mediacirculatie zijn ze grotendeels afhankelijk van Big Tech en hun platforms. Deze zijn primair gedreven door commerciële belangen en gaan dus niet uit van publieke waarden, zoals privacy, autonomie en transparantie. PublicSpaces onderneemt verschillende activiteiten om zo een digitaal ecosysteem te bouwen gebaseerd op publieke waarden.<sup>33</sup>
- *Bevorderen van bestaande initiatieven voor lokale online platforms.* Als alternatief -op de langere termijn- voor de grote internationaal werkende online platforms is het te overwegen om de mogelijkheden te onderzoeken of steun te geven aan initiatieven voor lokaal werkende online platforms of non-profit sociale media. De grote platforms hebben enorme aantallen gebruikers; de schade neemt toe zodra extreme- en borderline content grootschalig kan worden verspreid. Binnen een ecosysteem van kleinere lokale platforms speelt dit probleem minder. Bovendien kunnen gebruikers dan kiezen voor lokale platforms waar effectief en/of volgens bepaalde

33 De aanbeveling om publieke waarden meer centraal te stellen sluit aan bij de notie in *Rijksbrede Kamerbrief dat economische belangen een rol spelen bij de verspreiding van borderline content / desinformatie*. De *Praktijkcode tegen Desinformatie*, de DSA en de DMA bevatten maatregelen om dit verdienmodel aan te pakken. Het gebruik van publieke alternatieven voor online platformen en diensten bij publieke instellingen die voldoen aan publieke waarden dragen ook bij aan het inperken van het verdienmodel.

ethische normen wordt gemodereerd en de kans dus kleiner is dat ze met extreme en borderline content in aanraking komen. Het advies is hierbij wel omwille van de haalbaarheid aan te sluiten bij bestaande initiatieven, niet om zelf als overheid nieuwe initiatieven te initiëren. Ondersteuning van initiatieven voor lokaal werkende online platforms of non-profit sociale media sluit aan bij het voornemen van de Rijksoverheid (*Rijksbrede kamerbrief desinformatie dec 22*) om publieke alternatieven voor online platformen te stimuleren.<sup>34</sup>

- *Proactief betrekken van jeugdwerkers.* De rol van jeugdwerkers in de EU-wetgeving is onderbelicht. Zij weten het beste wat bij de jeugd actuele polariserende en radicaliserende content is. Zij zouden betrokken kunnen worden als betrouwbare 'flaggers'. De overheid kan hier proactief een rol in spelen door hierover in gesprek te gaan met beroepsorganisaties en 'koplopers' van jeugdwerkorganisaties. Tevens is bij jeugdwerkers behoefte aan (verdere) verdieping op online thema's, hoe zich daartoe te verhouden en te handelen. Dit betreft niet alleen communiceren online, maar ook signaleren en kennis van dynamieken in bepaalde contexten (zoals de online gamewereld). Daarbij kunnen de online en offline context niet los van elkaar gezien worden, deze zijn bij jongeren nauw aan elkaar verbonden. Meer bewustwording is gewenst bij werkgevers in het jeugd-domein omtrent het belang van online werken. Werkgevers moeten faciliteren dat jongerenwerkers hun online vaardigheid kunnen vergroten. De overheid kan dit stimuleren.

34 "Om online het publieke debat op een goede manier te kunnen voeren is het belangrijk dat er naast de bestaande grote online platformen ook alternatieven worden ontwikkeld waar mensen voor kunnen kiezen. De ontwikkeling van deze alternatieven stimuleren we" (*Kamerbrief Rijksbrede strategie effectieve aanpak van desinformatie; dec 22, pagina 9*).

## 7.5. Handelingsperspectieven offline

### Inzet op beschermende factoren

In dit rapport is op meerdere plaatsen beargumenteerd dat de online context en offline context niet los van elkaar gezien kunnen worden: deze liggen veeleer in elkaars verlengde (Gill et al., 2015; Mattheis, 2019; Whittaker & Herath, 2019; Whittaker, 2022). Behalve beleid dat erop gericht is om online (rechtsextremistische) radicalisering op sociale mediaplatforms tegen te gaan, moet er daarom ook worden geïnvesteerd in de digitale weerbaarheid van gebruikers van deze online platforms. Hiertoe kunnen er ten minste twee handelingsperspectieven voor de Nederlandse overheid gegeven worden:

#### *Bevorderen van kritisch denken en digitale weerbaarheid*

Een handelingsperspectief voor de overheid kan zijn om meer te investeren in kritisch denken, digitale weerbaarheid en mediawijsheid<sup>35</sup>. Het belang van individuele weerbaarheid van (in het bijzonder) jongeren, en dan in het bijzonder het vermogen om kritisch en flexibel te kunnen denken, neemt in de online context alleen maar toe. Jongeren in Nederland vinden het lastig om in te schatten of informatie op internet betrouwbaar is en, daarmee, nepnieuws te herkennen. Er kan gesteld worden dat jongeren minder digitaal vaardig en mediawijs zijn dan ze zelf denken. Ze zijn 'confident', maar niet 'competent' (Peeters et al., 2022). Belangrijk is dat beschermende programma's gericht op het bevorderen van kritisch denken en digitale weerbaarheid vertrekken vanuit het gegeven dat de online leefwereld van jongeren alomvattend is, in de zin dat zij nauw verweven is met hun fysieke leefwereld en vele facetten van hun leven beslaat. Rechtsextremistische groepen spelen handig in op de sociaal-emotionele behoeften van jongeren door de vorm aan te nemen van 'counterculture movements' die focussen op lifestyle, jeugdcultuur en kunst

(Ebner, 2019). Jongeren worden daarbij niet alleen uitgenodigd om rechtsextremistische content te consumeren, maar ook in staat gesteld om hier een actieve bijdrage aan te leveren en zelf content te creëren door middel van blogs, vlogs, memes of zelfs livestreams waarin maker en gebruiker direct op elkaar kunnen reageren en inspelen. Er wordt in dit verband gesproken over een 'participatory culture' (Peeters et al., 2022). Er is om deze reden veel te zeggen om digitale weerbaarheid en mediawijsheid – en het kritisch leren denken als onderdeel daarvan – over een lange periode te integreren in bredere programma's die recht doen aan de wijze waarop vormen van extremisme door kunnen dringen in het leven en het samenleven van jongeren. Denk bijvoorbeeld aan het koppelen van digitale weerbaarheid aan burgerschapstraining, waardoor het onder meer geïntegreerd kan worden in lesprogramma's op scholen.

#### *Investeren in inzicht in de online leefwereld*

Gegeven dat jongerenculturen altijd al dan niet gedeeltelijk onbegrepen zullen blijven voor (volwassen) gezagsdragers betekent niet men zich hier bij moet neerleggen. Overheden kunnen namelijk investeren in het bevorderen van meer inzicht bij ouders, docenten, jeugdprofessionals en gezagsdragers in de online leefwereld van jongeren. Het resultaat hiervan kan zijn dat zij beter leren te begrijpen hoe de aantrekkingskracht van (rechts)extremistische content en narratieven eruit kan zien voor jongeren en hoe jongeren in een dergelijk narratief of netwerk gelokt kunnen worden. Daarnaast kan gedegen kennis van deze online leefwereld bij ouders, docenten, jeugdprofessionals en gezagsdragers ervoor zorgen dat jongeren zich meer verbonden voelen met hen. Onder de juiste omstandigheden is het zelfs voorstelbaar dat dit kan helpen de voedingsbodem voor extremistisch gedachtegoed te verkleinen, omdat het jongeren de mogelijkheid geeft om hierover te praten met relevante 'volwassenen' in hun sociale netwerk. Preventie begint immers bij betekenisvol contact met jongeren en betekenisvol contact vereist inzicht in en kennis van de online leefwereld van jongeren (Peeters et al., 2022). Bovendien kan meer inzicht in de online leefwereld van jongeren gezagsdragers en andere relevante personen helpen om sneller signalen te herkennen wanneer zorgelijk

<sup>35</sup> Het bevorderen van kritisch denken en digitale weerbaarheidsluit aan bij de notie hieromtrent in de Rijksbrede kamerbrief desinformatie dec 22 .

gedachtengoed zich manifesteert bij jongeren. In dergelijke gevallen is het belangrijk dat er een dialoog tot stand komt binnen wat in de literatuur vaak aangeduid wordt als 'veilige ruimtes', zonder dat jongeren veroordeeld worden voor hun meningen en gedachten (R-Newt, 2019). Dit gesprek moet bij voorkeur niet, of in ieder geval niet alleen, door ouders worden gevoerd. Jongeren zetten zich immers in de adolescentie vaak juist af tegen de ouders. Andere familieleden en/of mensen uit de eigen sociale omgeving zijn hier geschikter voor. Ook jongerenwerkers kunnen hier een belangrijke rol spelen, omdat ze de (online) leefwereld van jongeren kennen en laten zien dat je kritisch kunt zijn op de maatschappij zonder dat je daarbij in extremen hoeft te vervallen. In dit verband is het belangrijk om te wijzen op de actieve rol die (lokale) overheden daarin kunnen hebben: het bevorderen van inzicht in deze online leefwereld van jongeren bij relevante volwassenen met wie zij in aanraking komen – bijvoorbeeld docenten of jongerenwerkers.

## **Samenvatting Hoofdstuk 7**

### *Onderzoeksvraag*

Dit hoofdstuk (in het bijzonder paragraaf 7.4) beantwoordt onderzoeksvraag 9: Welke haalbare mogelijkheden zijn er (voor de overheid) om rechtsextremistische radicalisering via online sociale mediaplatforms tegen te gaan. Paragrafen 7.1 tot en met 7.3 gaan in op de context voor mogelijke handelingsperspectieven.

### *Nuances omtrent de rol van aanbevelingsalgoritmen bij rechtsextremistische radicalisering (par. 7.1)*

Deze paragraaf nuanceert de invloed van algoritme gestuurde content op online media platforms bij rechtsextremistische radicalisering. Er zijn veel online én offline factoren en mechanismen die kunnen bijdragen aan rechts-extremistische radicalisering, waaronder het uitwisselen van informatie en links, echokamers met gelijkgestemden en abonnementen op bepaalde online kanalen. Algoritme gestuurd aanbod van rechtsextremistische content op online media platforms is derhalve slechts één van de mechanismen die een rol speelt. Ook als wordt ingezoomd op de werking van aanbevelingsalgoritmen zelf, dan hoeft er niet altijd sprake te zijn van sociale mediafuiken waar personen min of meer onbewust worden ingezogen door de aanbevelingsalgoritmen. Het gedrag van mensen zelf is zeer bepalend. Het kan zijn dat mensen bewust, uit zichzelf op zoek zijn naar extremere content en hiertoe doelbewust bepaalde zoektermen invoeren, bepaalde namen zoeken of op de meer extreme aanbevelingen doorklikken. Het tegengaan van rechtsextremistische radicalisering vereist kortom een aanpak waarbij aan meerdere knoppen tegelijkertijd gedraaid moet kunnen worden om tot gewenste resultaten te kunnen komen



### *Moderatiebeleid door sociale mediaplatforms (Paragraaf 7.2)*

Deze paragraaf beschrijft wat online mediaplatforms momenteel zelf al doen aan het tegengaan van de verspreiding van extreme – en borderline content via aanpassingen in aanbevelingsalgoritmen en contentmoderatie.

### *Sociale mediabeleid Nederland en Europese Unie (Paragraaf 7.3)*

De nieuwe Europese wetgeving Digital Services Act (DSA), de Digital Markets Act (DMA) en de European Media Freedom Act geeft overheden meer mogelijkheden dan in het verleden om de verspreiding van extreme – en borderline content tegen te gaan. De verplichtingen aan online platforms voor transparantie over contentmoderatie en aanbevelingsalgoritmen zijn aangescherpt. Aanvullende verplichtingen worden opgelegd aan de zeer grote online platforms (meer dan 45 miljoen gebruikers). Deze paragraaf bespreekt deze nieuwe mogelijkheden en verplichtingen.

### *Mogelijke handelingsperspectieven voor de overheid (Paragraaf 7.4)*

Handelingsperspectieven kunnen zich richten op het aanpakken van de volgende problemen:

1. Borderline content lijkt grotendeels de plaats te hebben ingenomen van extremere content op mainstream sociale mediaplatforms. Er is sprake van tekortschietende moderatie van borderline content op de grote sociale mediaplatforms. Er bestaan onduidelijkheden over wat er onder borderline content moet worden verstaan.
2. Er is sprake van een gebrek aan transparantie over de werking van aanbevelingsalgoritmen op de grote sociale mediaplatforms.
3. Er is sprake van tekortschietende moderatie van extreme content op alt-tech mediaplatforms.
4. Het zoek/surfgedrag van gebruikers zelf speelt een belangrijke rol bij blootstelling aan radicale content.

De beschreven handelingsperspectieven zijn niet altijd eenvoudig of snel te realiseren en vereisen een lange adem. De handelingsperspectieven van de Nederlandse overheid richting (zeer) grote sociale mediaplatforms en/of andere online platforms met de hoofdvestiging in het buitenland zijn beperkt. Immers, de Europese Commissie heeft slechts de rol van toezichthouder in respectievelijk het land van vestiging. Wat echter wel mogelijk is voor de Nederlandse overheid zijn de volgende handelingsperspectieven:

- Het ondersteunen van zogeheten betrouwbare ‘flaggers’: erkende, deskundige melders van illegale content op online platforms (probleem 1).
- Het uitoefenen van invloed op de Europese Commissie bij de implementatie van de DSA-wetgeving. Bijvoorbeeld om toezicht of handhaving te prioriteren (problemen 1, 2,3).

Handelingsperspectieven ten aanzien van minder grote sociale mediaplatforms zijn bijvoorbeeld:

- Dialoog met platforms over het sneller verwijderen van illegale content (probleem 3).
- Investeren in de samenwerking tussen overheden, online platforms en onderzoekers (problemen 1, 2,3).
- Coördinatie tussen ministeries en de continue opbouw van kennis (zoals over algoritmen, moderatie, borderline content e.d.) (problemen 1, 2, 3);
- Toewerken naar meer co-regulering tussen mediaplatforms en de overheid (problemen 1, 2).
- Beter begrijpen van de problematiek rondom *borderline* content op mediaplatforms; en meer duidelijkheid creëren bij platforms over aanpak van deze borderline content (probleem 1).
- Bevorderen van contentmoderatie in de Nederlandse taal op sociale mediaplatforms (probleem 1).
- Onderzoek naar- en vergroten bewustzijn omtrent de rol van influencers bij toegang tot extremere narratieven (problemen 1, 3, 4).

- Het inzetten op discussies omtrent het belang van 'publieke waarden' (naast economische waarden) in het beleid richting sociale mediaplatforms en creëren van randvoorwaarden om hiermee te kunnen experimenteren (problemen 1, 2).
- Mogelijkheden onderzoeken of steun geven aan bestaande initiatieven voor een online ecosysteem van kleinere, lokale sociale mediaplatforms (problemen 1, 2).

Tevens kan ingezet worden op het bevorderen van offline aanpakken om radicalisering op sociale mediaplatforms tegen te gaan. Concrete voorbeelden hiervan zijn het versterken van kritisch denken en digitale weerbaarheid bij jongeren, het creëren van meer inzicht in de online leefwereld van personen, het betrekken van jeugdwerkers als betrouwbare 'flaggers' en het bevorderen van bewustwording bij werkgevers omtrent het belang van vergroten van online vaardigheden van jeugdwerkers.

## 8 Conclusies

In dit onderzoek hebben we in kaart gebracht wat verschillende ontwikkelingspaden van rechtsextremistische radicalisering zijn en wat hierbij de rol is van sociale mediaplatforms. Er was daarbij specifiek aandacht voor de werking van aanbevelingsalgoritmen en de rol van de offline context in ontwikkelingspaden van rechtsextremistische radicalisering. Op basis van de bevindingen hebben we handelingsperspectieven beschreven die haalbaar geacht worden voor de overheid om bij te dragen aan het voorkomen van rechtsextremistische radicalisering via online platforms. Het onderzoek hebben we uitgevoerd door middel van een uitgebreide literatuurstudie; het opzetten en uitvoeren van aanvullende trials om de werking van aanbevelingsalgoritmen te bestuderen op de sociale media platforms YouTube en TikTok; en door diepte-interviews en online groepsgesprekken met experts op verschillende gebieden. Hieronder worden de belangrijkste conclusies uit deze rapportage beschreven.

### **Grote variëteit aan risicofactoren en ontwikkelingspaden**

Uit de wetenschappelijke literatuur blijkt dat een groot aantal risicofactoren op individueel-, groeps- en samenlevingsniveau kunnen bijdragen aan het ontstaan van radicaliseringsprocessen. Daarbij zijn er meer overeenkomsten dan verschillen in de kenmerken, risicofactoren en mechanismen van radicaliseringsprocessen met uiteenlopende ideologische achtergronden (links- en rechtsextremisme, religieus gemotiveerd en single-issue extremisme). Ontwikkelingspaden van radicalisering zijn uiterst complex, waarbij zowel online als offline factoren een rol spelen en op elkaar in werken. Om de invloed van de online context op extremistische overtuigingen en geweld te kunnen begrijpen, dient te worden gekeken naar deze wisselwerking tussen de online en offline context. Op basis van empirische studies blijkt immers dat de meeste

personen die grotendeels online radicaliseren niet overgaan tot gewelddadig gedrag, wat maakt dat aandacht voor offline factoren eveneens van belang is. Er zijn zeer uiteenlopende manieren waarop rechtsextremistisch gedachtengoed de online leefwereld van personen kan binnendringen. Zo kunnen verschillende online strategieën van rechtsextremistische individuen en netwerken bijdragen aan de blootstelling, normalisering en acceptatie van rechtsextremistische content. Een belangrijke strategie is bijvoorbeeld het 'vermommen' van rechtsextremistische content door het gebruik van humor, ironie en *dog whistles*. Een langdurige blootstelling aan rechtsextremistische content en -strategieën kan personen hiervoor steeds ontvankelijker maken voor deze content. In de literatuur worden hiertoe meerdere zogeheten 'cognitieve fasen' geïdentificeerd. Het gaat om de volgende fasen: normalisering, acclimatisering en dehumanisering. Deze fasen overlappen elkaar, zijn niet exclusief of strikt lineair, maar veeleer cyclisch van aard.

### **Gateways voor rechtsextremistische radicalisering**

Radicaal- en rechtsextremistische groepen zijn actief op verschillende sociale mediaplatforms, maar zij hebben nog altijd het grootste bereik via mainstream-platforms zoals Facebook, Twitter, YouTube en in iets mindere mate Instagram. Deze grote platforms hebben in de afgelopen jaren streng opgetreden tegen kanalen en pagina's die gelinkt zijn aan rechtsextremistische groepen of alternatieve media. Alternatieve ofwel 'fringe' platforms als Gab, Telegram en 4chan zijn mede daardoor populairder geworden onder rechtsextremistische groepen, omdat hier meer extremistische content kan worden gepost. Aangezien het bereik op alternatieve platforms kleiner is, wordt door diezelfde gebruikers op de grotere platforms soms meer 'gematigd'/minder extreem materiaal gepost dat aan de voorwaarden van het platform voldoet en dus niet verwijderd wordt. Wel blijkt dat gebruikers via deze 'gematigde' content soms via links doorgeleid worden, of zelf op zoek gaan, naar meer extreme content elders en zo uitkomen bij alternatieve platforms. Ook verschillende

online subculturen (zoals conspiratorialiteit, de manosphere en de Vikingcultuur) kunnen fungeren als zogeheten 'gateways' naar rechtsextremistisch gedachtegoed. Het is belangrijk te noemen dat deze subculturen op zichzelf niet georiënteerd zijn op (rechts)extremistisch gedachtegoed, maar door de interpretatie van bepaalde uitgangspunten of ideeën een aanleiding kunnen geven voor personen om zich te interesseren in rechtsextremistische narratieven of geconfronteerd te worden met inhoud die daaraan relateert. Een andere mogelijke gateway naar extremistische content wordt gevormd door sommige populaire online influencers als Andrew Tate. Online is sprake van talrijke invloedrijke figuren die via grote sociale mediaplatforms of gameplatforms een ingang kunnen zijn voor (jonge) personen om zich online verder te verdiepen in bepaald gedachtegoed of bepaalde gemeenschappen.

### **Online en offline rekrutering**

Vanuit de wetenschappelijke literatuur is betrekkelijk weinig bekend over casuïstieken waarbij is vastgesteld dat een sociale mediafuik heeft bijgedragen aan rechtsextremistische radicalisering of rekrutering. Het vaststellen hiervan is ook niet eenvoudig omdat de werking van aanbevelingsalgoritmen en filterbubbels zich niet een geïsoleerd universum voltrekt. Er zijn immers ook andere online- en offline mechanismen die bij een persoon kunnen bijdragen aan zijn of haar radicalisering. Daarbij is het moeilijk te meten of iemand specifiek door de sociale mediafuik is geradicaliseerd of dat andere factoren een grotere bijdrage hebben geleverd. Het is kortom nauwelijks mogelijk om te concluderen dat een individu uitsluitend is geradicaliseerd door consumptie van eenzijdige online content op sociale mediaplatforms en dat diegene zonder deze media niet geradicaliseerd zou zijn. Desalniettemin zijn er anekdotische aanwijzingen (casussen) waarin aannemelijk is dat de sociale mediafuik een rol heeft gespeeld in het radicaliseringsproces. Vraag is dan echter wel in hoeverre deze fuik veroorzaakt wordt door bijvoorbeeld aanbevelingsalgoritmen op sociale mediaplatforms, of dat er sprake is van een ogenschijnlijke fuik die veeleer te wijten is aan het eigen online gedrag (keuzes) van gebruikers.

### **Rol van aanbevelingsalgoritmen bij rechtsextremistische radicalisering is beperkt**

Zoals hierboven is opgemerkt, voltrekt het mechanisme van een "algoritme gestuurd aanbod van rechtsextremistische content op sociale mediaplatforms dat kan leiden tot een rechtsextremistische sociale media fuik" zich allesbehalve in een vacuüm. De rol van aanbevelingsalgoritmen en sociale mediafuiken bij rechtsextremistische radicalisering lijkt kleiner dan in het publieke discours soms wordt voorondersteld. Personen kunnen immers zowel online als offline op allerlei manieren in aanraking komen met polariserende of extremistische content. Verschillende *persoonlijke keuzes* die platformgebruikers maken zijn daarbij in het bijzonder relevant. Enkele voorbeelden zijn: gebruikers delen binnen hun sociale netwerk weblinks naar extremistische content met elkaar; mensen abonneren of volgen bepaalde kanalen op sociale media en worden om die reden selectief geconfronteerd met extreme(re) content; gelijkgestemden kunnen elkaar ontmoeten in (soms gesloten) chatgroepen of pagina's, welke fungeren als echokamers; en het eigen online gedrag van gebruikers kan bepalend zijn voor de content die aan hen getoond wordt, doordat zij uit zichzelf op zoek gaan naar extremistische content en daarom bewust bepaalde zoektermen invoeren of op bepaalde content klikken.

Er kunnen daarnaast vraagtekens geplaatst worden bij de rol *an sich* van het internet en sociale mediaplatforms bij gewelddadige radicalisering. Er is namelijk weinig empirisch bewijs dat radicalisering in de online context leidt tot *gewelddadig* extremisme, noch dat gewelddadig extremisme rechtstreeks voortvloeit uit online radicalisering / radicalisering die met name online plaatsvindt. Kortom, er zijn veel online én offline factoren en mechanismen die kunnen bijdragen aan rechtsextremistische radicalisering. De rol van aanbevelingsalgoritmen lijkt hierbij relatief beperkt gegeven dat recente empirische studies naar de werking van deze algoritmen op platforms steeds minder vaak aanwijzingen geven voor veronderstelde 'filterbubbels' of 'sociale mediafuiken'.

Al met al betekent dit dat het tegengaan van rechtsextremistische radicalisering een aanpak vereist waarbij aan meerdere knoppen tegelijkertijd gedraaid moet worden om tot gewenste resultaten te kunnen komen. Nadrukkelijk oog hebben voor de wisselwerking tussen het online en offline domein is daarbij van belang.

### **Moderatie extremistische content is verbeterd, ook gaan focussen op zoekgedrag gebruikers**

Op basis van de literatuurstudie en de uitgevoerde trials op YouTube en TikTok kan geconcludeerd worden dat de contentmoderatie door (mainstream) sociale mediaplatforms steeds beter op orde lijkt en zijn beoogde doelen lijkt te bereiken. Een belangrijke bevinding van de trials is namelijk dat de algoritmen op basis van *neutrale* zoekwoorden geen extremistische (illegale) content en relatief weinig borderline content aanbevelen. Gebruikers kunnen desondanks wel in aanraking komen met radicalere/borderline content. Echter, om meer suggesties door het platform voor borderline content te krijgen, moet hierop actief, expliciet en herhaaldelijk gezocht worden door platformgebruikers. Een aanbeveling naar de platforms is daarom dat zij zich dus óók moeten focussen op ook het gedrag van gebruikers zélf.

### **'Kuil' van borderline content**

Effecten van contentmoderatie van illegale extremistische content op de platforms zijn zichtbaar. Dit ondersteunt de notie dat contentmoderatie belangrijk en nodig is. Desalniettemin blijft er een grijs gebied van content over – ofwel content die ervaren kan worden als onwenselijk, maar die niet op grond van wettelijke kaders verwijderd kan worden. In dit verband wordt er gesproken van 'borderline content'. Borderline content lijkt op de platforms grotendeels de plaats te hebben ingenomen van extremistische content. De bevindingen uit de literatuurstudie en trials suggereren daarentegen ook dat er zeker nog ruimte is voor verdere verbetering om borderline content accurater te detecteren en te weren uit de tijdlijn van de gebruikers. Borderline content tast soms de grenzen van de contentmoderatie af, welke door het gebruik

van gecodeerde taal, symbolen en ambigu taalgebruik steeds meer expertise vereist om deze correct te identificeren. Het gebrek aan deze expertise is op dit ogenblik een zwakke schakel in de strijd tegen borderline content op de platforms. Een effectieve aanpak van borderline content staat of valt met de correcte identificatie ervan, wat vraagt om praktisch handzame definities. Een rol hierbij kan worden gespeeld door zogeheten '*trusted flaggers*' (conform de nieuwe DSA wetgeving: instanties of personen met bijzondere expertise en competentie voor het opsporen, identificeren en melden van illegale content of andere overtredingen van de wet op onder meer sociale mediaplatforms).

### **Mogelijke handelingsperspectieven voor de overheid**

De nieuwe Europese wetgeving, in het bijzonder de Digital Services Act (DSA), alsook de Digital Markets Act (DMA) en de European Media Freedom Act, geeft overheden meer mogelijkheden dan in het verleden om de verspreiding van extremistische- en borderline content tegen te gaan op het internet. De in deze rapportage gegeven handelingsperspectieven richten zich op de aanpak van de volgende problemen:

1. Borderline content lijkt grotendeels de plaats te hebben ingenomen van extremere content op mainstream sociale mediaplatforms. Er is sprake van tekortschietende moderatie van borderline content op de grote sociale mediaplatforms. Er bestaan onduidelijkheden over wat er onder borderline content moet worden verstaan.
2. Er is sprake van een gebrek aan transparantie over de werking van aanbevelingsalgoritmen op de grote sociale mediaplatforms.
3. Er is sprake van tekortschietende moderatie van extreme content op alt-tech mediaplatforms.
4. Het zoek/surfgedrag van gebruikers zelf speelt een belangrijke rol bij blootstelling aan radicale content.

Handelingsperspectieven voor de Nederlandse overheid richting zeer grote (mainstream) sociale mediaplatforms of andere platforms met een hoofdvestiging in het buitenland zijn beperkt, omdat de Europese Commissie respectievelijk het land van vestiging hier de rol van toezichthouder heeft. Wel kan de Nederlandse overheid *trusted flaggers* identificeren, betrekken en ondersteunen; en hiernaast proberen invloed uit te oefenen op Europese Commissie bij implementatie DSA wetgeving, zoals op prioriteiten toezicht of handhaving.

Wat betreft minder grote sociale mediaplatforms, lijken er meer handelingsperspectieven mogelijk voor de Nederlandse overheid. Deze handelingsperspectieven zijn te categoriseren langs vier lijnen:

1. Het sneller verwijderen van extreme, illegale content op de kleinere platforms. De DSA lijkt vooral gericht op de hele grote mediaplatforms, terwijl er ook risico's bestaan bij de minder grote en kleine platforms. Een mogelijke aanbeveling ter verbetering is dat vooral de kleinere platforms meer zouden kunnen doen om sneller illegale content te verwijderen. De overheid kan daarin een rol spelen. De DSA lijkt mogelijkheden immers te bieden voor de beoogde coördinator (ACM) om hierover in gesprek te gaan met kleinere online platforms, in het bijzonder de in Nederland gevestigde online platforms.
2. Samenwerking met platforms (verschillende handelingsperspectieven);
  - Investeren in samenwerking overheden, online platforms en onderzoekers. Op dit moment hebben overheden en gebruikers geen of weinig inzicht in de aanbevelingsalgoritmen die sociale mediaplatforms gebruiken. Relevant is daarom de bepaling in de DSA die voorschrijft dat sociale mediaplatforms transparanter dienen te zijn richting haar gebruikers, overheden en onderzoekers. Dit vraagt bijvoorbeeld om transparantie over hun beleid omtrent contentmoderatie en het gebruik van aanbevelingsalgoritmen. Om een beter beeld te krijgen van de problematiek, ook omtrent de aanpak van borderline content, is het aan te bevelen dat zowel overheden als platforms investeren in een goede samenwerking met onderzoekers.

- Toewerken naar meer co-regulering tussen platforms en overheid waarbij de overheid en online platforms samenwerken om problemen aan te pakken.
- Er zijn momenteel nog weinig handvatten omtrent de omgang met borderline content. Het is ingewikkeld voor online platforms als de overheid zegt dat zij borderline content moeten aanpakken, terwijl duidelijkheid ontbreekt over wat precies aangepakt moet worden. De overheid zou kunnen investeren in het beter begrijpen van de problematiek rondom borderline content en faciliteren dat hierover meer duidelijkheid komt. Aangezien de overheid zelf terughoudend dient te zijn in het beoordelen van 'harmful content' is hier primair een rol weggelegd voor *trusted flaggers* met kennis van zaken over mogelijke schadelijke effecten en illegaliteit van border line content. Een groot probleem met de aanpak van borderline content is immers de beoordeling wanneer deze content illegaal is, zoals bij bijvoorbeeld beeldspraak en ironie waarin extremistische content wordt verpakt. 'Freedom of speech' is echter niet hetzelfde als 'freedom of reach'. Online platforms kunnen zulke borderline content weliswaar niet verbieden maar ze kunnen wel de amplificatie ervan belemmeren, wat ze in de praktijk overigens ook al doen. Als betrouwbare *flaggers* meer duidelijkheid scheppen over de illegaliteit en schadelijkheid van bepaalde borderline content, dan zal deze vorm van contentmoderatie door online platforms makkelijker ingezet kunnen worden.
- Om te voorkomen contentmoderatie minder effectief is bij niet-Engelstalige content is een mogelijk handelsperspectief voor de overheid om hierover in gesprek te gaan met de online platforms, transparantie te vragen hoe hiermee wordt omgegaan en hen zonodig te stimuleren hier meer aandacht aan te geven.
- Centraal stellen van publieke waarden. Het beleid omtrent sociale media platforms zou niet alleen gebaseerd moet zijn op economische – en juridische uitgangspunten maar ook op basis van het borgen van publieke waarden. Hier zou meer discussie over moeten zijn en ruimte

en randvoorwaarden worden gecreëerd om te experimenteren. De overheid zou dit kunnen faciliteren.

3. Samenwerking tussen overheden / organisaties / ruimte voor experimenteren (verschillende handelingsperspectieven);
  - Het up-to-date houden van kennis omtrent online sociale mediaplatforms, hun algoritmen en contentmoderatie is belangrijk en vergt daarom afstemming tussen Nederlandse ministeries. Het is zodoende belangrijk dat onderzoeken hiernaar niet op zichzelf blijven staan, maar juist met elkaar in verband worden gebracht of voortbouwen op elkaar, zodat een continue kennisopbouw (beter) gegarandeerd kan worden.
  - Als alternatief voor de grote internationaal werkende online platforms is het te overwegen om de mogelijkheden te onderzoeken of steun te geven aan initiatieven voor lokaal werkende online platforms of non-profit sociale media. De grote platforms hebben enorme aantallen gebruikers; de schade neemt toe zodra extreme- en borderline content grootschalig kan worden verspreid. Binnen een ecosysteem van kleinere lokale platforms speelt dit probleem minder. Bovendien kunnen gebruikers dan kiezen voor lokale platforms waar effectief en/of volgens bepaalde ethische normen wordt gemodereerd en de kans dus kleiner is dat ze met extreme en borderline content in aanraking komen. Het advies is hierbij wel omwille van de haalbaarheid aan te sluiten bij bestaande initiatieven, niet om zelf als overheid nieuwe initiatieven te initiëren.
  - Proactief betrekken van jeugdwerkers. De rol van jeugdwerkers in de EU-wetgeving is onderbelicht. Zij weten het beste wat bij de jeugd actuele polariserende en radicaliserende content is. Zij zouden betrokken kunnen worden als trusted flaggers. De overheid kan hier proactief een rol in spelen door hierover in gesprek te gaan met beroepsorganisaties en 'koplopers' van jeugdwerkorganisaties.
  - Online zijn talrijke influencers die via grote sociale mediaplatforms of gameplatforms een ingang kunnen zijn voor (jonge) personen om zich online verder te verdiepen in bepaald gedachtegoed of subculturen

die een 'gateway' kunnen zijn naar rechtsextremistisch gedachtegoed. Deze invloedrijke rol van influencers en subculturen (en de risico's hiervan) als 'gateways naar rechtsextremistisch gedachtegoed' kan nader wetenschappelijk worden onderzocht om meer bewustzijn creëren rondom dit thema.

4. Inzet op offline aanpakken (verschillende handelingsperspectieven);
  - Een handelingsperspectief voor de Nederlandse overheid kan zijn om meer te investeren in kritisch denken, digitale weerbaarheid en mediawijsheid. Het belang van individuele weerbaarheid van (in het bijzonder) jongeren, en dan in het bijzonder het vermogen om kritisch en flexibel te kunnen denken, neemt in de online context alleen maar toe.
  - Dat jongerencultuur altijd deels onbegrepen zal blijven voor (volwassen) gezagsdragers betekent niet dat we ons hier bij moeten neerleggen. Overheden kunnen investeren in het bevorderen van meer inzicht bij ouders, docenten, jeugdprofessionals en gezagsdragers in de online leefwereld van jongeren en zo beter te begrijpen wat de aantrekkingskracht van rechts-extremisme kan zijn en hoe jongeren zo'n narratief en/of netwerk ingelokt kunnen worden.
  - Het bevorderen van meer bewustwording bij werkgevers omtrent het belang van vergroten van online vaardigheden van jeugdwerkers.
  - Opvallende bevinding in de trials op YouTube was dat de meest radicale content te vinden was onder (in de commentsecties) van artikelen in de nieuwsmedia. Dergelijke berichten kunnen aantrekkingskracht hebben op personen of bepaalde reacties van hen uitlokken. Er worden veel mediaberichten becommentarieerd; opnieuw geplaatst; boze of opruiende comments onder video's waarin enkel nieuws wordt weergegeven. Suggestie zou kunnen zijn dat de nationaal coördinator hierover het gesprek aangaat met nationale of internationale nieuwsmedia om hen hiervan bewuster te maken. Platforms en media kunnen weblinks naar extremere content verwijderen.

## Slotbeschouwing

Dit onderzoek geeft antwoorden op de vraag hoe ontwikkelingspaden van rechtsextremistische radicalisering eruitzien en wat daarbij in het bijzonder de rol is van online sociale mediaplatforms. Hierin wordt een onderscheid gemaakt tussen mainstream platforms als YouTube en TikTok en zogeheten 'alt-tech' platforms. Na het bestuderen van de relevante wetenschappelijke literatuur over deze thema's is het duidelijk geworden dat ontwikkelingspaden van radicalisering uiterst complex zijn, en een samenspel van zowel online als offline factoren behelzen. Er is echter weinig empirisch bewijs dat radicalisering in de online context leidt tot *gewelddadig* extremisme, noch dat gewelddadig extremisme rechtstreeks voortvloeit uit online radicalisering. Het is van belang om de online en offline contexten bij radicaliseringsprocessen niet als twee zelfstandige domeinen te zien: de beide contexten beïnvloeden elkaar en liggen veeleer in het verlengde van elkaar. Het huidige onderzoek naar de wisselwerking tussen de online en offline contexten bij radicaliseringsprocessen is beperkt en behoeft daarom verdere studie.

Tevens hebben we vastgesteld dat de rol van aanbevelingsalgoritmen beperkt lijkt bij rechtsextremistische radicalisering in de online context. Persoonlijke keuzes die gebruikers maken blijken minstens zo belangrijk – zo laten verschillende empirische studies zien. Een belangrijk inzicht op basis van recente empirische studies, waaronder onze uitgevoerde trials, is dat de mainstream sociale mediaplatforms de moderatie van extremistische, illegale content beter op orde lijken te hebben dan voorheen. In de afgelopen jaren is hier namelijk veel in geïnvesteerd door deze platforms na kritiek uit verschillende hoeken. Recentere empirische studies laten dan ook zien dat gebruikers via aanbevelingsalgoritmen veel minder vaak/gemakkelijk extremistische content gesuggereerd krijgen door het platform – tenzij gebruikers zelf doelbewust zoeken naar deze content of abonneren op accounts die deze content plaatsen. Echter, er kunnen een aantal disclaimers geplaatst worden bij deze conclusies: (1) extremistische content lijkt op mainstream sociale media platform grotendeels vervangen te zijn door borderline content; (2) extremistische content is verplaatst naar

alternatieve 'alt-tech' platforms; en (3) platformgebruikers kunnen nog steeds in aanraking komen met extremistische content door het eigen online gedrag, zoals het delen van weblinks, het abonneren of volgen van bepaalde kanalen en het zoeken/vinden van gelijkgestemden in bepaalde online (gesloten) chatgroepen of op fora. Een belangrijke oplossingsrichting in het tegengaan en voorkomen van rechtsextremistische radicalisering in de online context is dan ook het centraalstellen (en problematiseren) van aanwezige borderline content op sociale mediaplatforms. Dit omvat niet-illegale, maar vaak wel radicale content die op grond van de community- of gebruikersrichtlijnen van een platform niet verwijderd/gedetecteerd wordt. Belangrijk is daarom dat er handvatten komen die helpen om deze content beter te begrijpen en daarmee beter te kunnen modereren. De risico's van borderline content zijn immers aanwezig. Andere oplossingsrichtingen zijn (a) het besteden van (meer) aandacht voor illegale content op alternatieve, 'alt-tech' mediaplatforms die vaak relatief makkelijk te vinden is voor gebruikers van deze platforms; en (b) het al dan niet gedeeltelijk 'terugleggen' van de verantwoordelijkheid bij gebruikers zelf. Laatstgenoemde vloeit voort uit het feit dat het in aanraking komen met extremistische content op mediaplatforms voor een deel te wijten is aan het eigen online internetgedrag, zoals blijkt uit verschillende empirische studies. Belangrijk is daarom om – naast de gegeven handelingsperspectieven die aanspraak doen op online platforms – méér in te zetten op de online mediawijsheid én weerbaarheid van (jonge) internetgebruikers. Extremistische content zal immers altijd beschikbaar zijn op het onuitputtelijke internet, wat maakt dat het in de eerste plaats uiterst belangrijk is om mensen weerbaar te maken voor deze en andere radicale/borderline content. Inmiddels zijn er meerdere geëvalueerde lesprogramma's beschikbaar die breder ingezet kunnen worden online weerbaarheid en mediawijsheid te vergroten.



# Bijlage 1 Glossary

## *Algoritmen / aanbevelingsalgoritmen*

Een algoritme op sociale mediaplatforms is een technische formule waarmee platforms 'berekenen' wat zij een gebruiker willen laten zien. Algoritmen zijn daarmee een tool die platforms in staat stelt invloed uit te oefenen op de verblijfsduur op het platform van haar gebruikers. Door relevante content aan te bieden aan gebruikers, wordt beoogd dat zij geïnteresseerd/betrokken blijven en meer tijd doorbrengen op het platform. Algoritmen op sociale mediaplatforms kunnen op verschillende manieren bijdragen aan het genereren van relevante content voor gebruikers. *Zoekalgoritmen*: deze algoritmen betrekken zich op de zoekopdrachten van een gebruiker en het genereren van relevante zoeksuggesties. Een voorbeeld hiervan is het automatisch aanvullen van (het begin van) een zoekopdracht die ingevoerd wordt door een gebruiker. *Aanbevelingsalgoritmen*: deze algoritmen genereren content die naar verwachting relevant is of interessant gevonden wordt door een gebruiker. Een voorbeeld hiervan zijn de 'aanbevolen video's' op YouTube of video's op de 'For You-page' van TikTok. *Ad-tech algoritmen*: deze algoritmen hebben als doel het aanbod van advertenties op sociale mediaplatforms te optimaliseren naar de relevantie ervan voor een gebruiker. Vaak gebeurt dit op basis van data over het online gedrag en demografische kenmerken van een gebruiker. *In het rapport focussen we op de rol van aanbevelingsalgoritmen.*

## *Alt-tech of 'fringe' platforms*

Sociale mediaplatforms met een relatief beperkt bereik (in vergelijking met mainstream sociale media platforms). Een aantal alt-tech platforms zijn zo ingericht dat deze rechtsextremistische netwerken faciliteren. Daarnaast zijn er alt-tech platforms die 'ideologisch neutraal' zijn, maar door beperkte moderatie vooralsnog ruimte creëren voor rechtsextremistische content.

## *Borderline content (ook wel aangeduid als 'legal yet harmful' content)*

Content op sociale media platforms die ervaren kan worden als onwenselijk, maar die niet op grond van wettelijke kaders verwijderd kan worden. De EU Counter-Terrorism Coordinator (2020) geeft de volgende definitie: *"Borderline content' is toegestane content, omdat het voldoet aan de servicevoorwaarden van de sociale media platforms, maar komt in de buurt van verboden content. Borderline-inhoud zou legale schadelijke inhoud omvatten. Vanuit juridisch oogpunt kan 'content op het randje' worden opgevat als legale content waarvan de wettigheid onduidelijk of dubbelzinnig blijft (bijv. geweld versus aanzetten tot geweld)".*

## *Conspiriteit*

Op het eerste gezicht lijken de werelden van het door links gedomineerde New Age (met daarbinnen een positieve focus op het zelf) en de door rechts gedomineerde complottheorieën (met daarbinnen een negatieve focus op de wereldpolitiek) heel tegenstrijdig. Echter, er is een punt waarop deze twee werelden hand in hand lijken te gaan. Dit 'punt' wordt ook wel 'conspiriteit' genoemd. Conspiriteit is een snelgroeiend online fenomeen waarbinnen een ideologie uitgedrukt wordt die zich voedt met politieke desillusie en de populariteit van alternatieve wereldbeelden. Het biedt een brede politiek-spirituele filosofie die zich baseert op twee kernovertuigingen: (a) een geheime groep controleert heimelijk de politieke en sociale orde en probeert deze te beheersen – deze overtuiging valt samen met traditionele complottheorieën; en (b) de mensheid ondergaat een 'verschuiving' in het bewustzijn – deze overtuiging is geworteld in de New Age-beweging en refereert naar een wereldbeeld waarin sprake is van een 'ontwaakt nieuw paradigma'. Zie paragraaf 3.6 voor nadere toelichting op conspiriteit.

### *Digital Services Act*

De Digital Services Act (DSA) is nieuwe Europese wetgeving welke relevant is voor het beleid omtrent online sociale mediaplatforms en thans wordt geïmplementeerd. Met de DSA komen er uitgebreide verplichtingen voor online platforms die consumenten in contact brengen met goederen, diensten of content. De DSA beoogt schadelijke praktijken te beperken, online risico's tegen te gaan en gebruikersrechten online beter te beschermen. Er komt meer beperking op de online verspreiding van illegale content en illegale producten. Daarnaast biedt de DSA ook de mogelijkheid om burgers beter te beschermen tegen desinformatie en nepnieuws en betere bescherming van grond- en mensenrechten. In welke mate online partijen met de nieuwe regels te maken krijgen, hangt af van hun rol, omvang en impact op het online ecosysteem. Voor de grote sociale mediaplatforms (als YouTube, Facebook, Instagram en TikTok) en zoekmachines met meer dan 45 miljoen gebruikers komen de meeste verplichtingen. Begin 2024 wordt de DSA volledig van toepassing voor alle diensten die daaronder vallen. Zie paragraaf 7.3 voor nadere toelichting op de Digital Services Act.

### *Dog whistles*

Dog whistles is een retorisch middel dat gebruikt wordt door onder andere rechtsextremisten. Dog whistles zijn 'verborgen' politieke boodschappen in tekst of afbeeldingen die bedoeld zijn voor een specifieke groep en vaak enkel door hen begrepen worden. De impliciete betekenis van dog whistles kan altijd ontkend worden, met als gevolg dat gebruikers het norm-overschrijdende karakter ervan kunnen ontkennen en moderatie door sociale mediaplatforms kunnen ontwijken. Zie paragraaf 3.3 voor nadere toelichting op dog whistles.

### *Echokamers*

Een echokamer op sociale media is een omgeving of ecosysteem waarin gebruikers voornamelijk of uitsluitend overtuigingen tegenkomen die hun reeds bestaande overtuigingen bevestigen. In een online echokamer komen gebruikers niet of nauwelijks tegengestelde opvattingen tegen. Echokamers kunnen sociale en politieke polarisatie en extremisme vergroten. Zie paragraaf 4.3 voor nadere toelichting op echokamers.

### *Filterbubbels / sociale media fuik*

Gebruikers van sociale media kunnen als gevolg van aanbevelingsalgoritmen in een 'fuik' van content terecht komen, hetgeen ook wel een 'filterbubbel' of 'rabbit hole' wordt genoemd. Algoritmen bevelen aan een gebruiker content aan waarvan 'berekend' is dat deze gebruiker dit interessant zal vinden. Vaak is deze content afgestemd op eerder online gedrag van de gebruiker. Content die niet relevant wordt geacht voor een gebruiker, wordt door het aanbevelingsalgoritme niet of minder aan hem getoond. Het wegfilteren van andere meningen, attitudes en houdingen kan een filterbubbel veroorzaken met eenzijdige content. Hierdoor wordt gebruikers tot op zekere hoogte de mogelijkheid ontnomen om een kritische attitude te ontwikkelen op basis van objectieve en veelzijdige informatie. Zie paragraaf 4.3 voor nadere toelichting op filterbubbels.

### *Incels subcultuur*

De term incels is een afkorting van involuntary celibates. Dit omvat een subcultuur waarbinnen mannelijke leden geloven dat de maatschappij hun verlangen om seks te hebben, te daten of relaties te hebben met vrouwen, verhindert. De gemeenschap bevindt zich vrijwel volledig online. Incels gaan uit van een sociale hiërarchie met "chads" (aantrekkelijke mannen) aan de top, gevolgd door "normies", de normale meerderheid van de bevolking en op de bodem de incels. De ervaren oneerlijkheid onder incels leidt onder sommige van hen tot het goedpraten van geweld tegen de samenleving, met name verbaal of fysiek geweld tegen vrouwen. Zie paragraaf 3.6 voor nadere toelichting op de incels subcultuur.

### *Mainstream platforms*

Sociale mediaplatforms met een groot bereik zoals Facebook, Twitter, YouTube en Instagram.

### *Manosfeer/masculiene subculturen*

De zogeheten 'manosfeer' behelst een gevarieerde verzameling van online 'mannen-ondersteuningsgemeenschappen' die zijn ontstaan als reactie op onder meer het feminisme, vrouwelijk empowerment, maatschappelijke veranderingen en de vervreemdende effecten van het neoliberalisme. Binnen deze subcultuur wordt aan vrouwen en feministen de schuld gegeven van talloze problemen in de samenleving. De manosfeer of de daaraan verwante nadruk op masculiniteit is op zichzelf niet rechtsextremistisch, maar kent wel overlap met diverse extreemrechtse thema's. Zie paragraaf 3.6 voor een nadere toelichting op manosfeer/masculiene subculturen.

### *Preppers community*

De Preppers community is een politiek neutrale gemeenschap van mensen die zich willen voorbereiden op een extreme noodsituatie als gevolg van een ramp of sociale, politiek en/of economische opschudding en ongeremdheden. Preppers hebben zich vaak medische kennis en zelfverdedigingstechnieken eigen gemaakt en zorgen voor allerhande voorzieningen die hen helpen te overleven en zo min mogelijk afhankelijk te zijn van de staat. Ook binnen rechtsextremistische kringen geniet dit 'preppen' de nodige aandacht. De aantrekkingskracht ligt hier vooral in het niet langer afhankelijk zijn van de staat – waarin al het vertrouwen is verloren. Zie paragraaf 3.6 voor nadere toelichting op Preppers community.

### *Qanon*

Eén van de bekendste hedendaagse complottheorieën die soms door zowel aanhangers van conspiratorialiteit als door rechtsextremisten onderschreven wordt, is Qanon. Oorspronkelijk is de QAnon-beweging in 2017 ontstaan in Amerikaanse far right politieke kringen. Aanhangers hiervan geloven in de aanwezigheid van een onzichtbare en invloedrijke elite met satanistische

ideeën "die achter de schermen stiekem de touwtjes in handen heeft" – de zogeheten deep state die achter de rug van officiële overheden het wereldtoneel beheerst. Zie paragraaf 3.6 voor een nadere toelichting op Qanon.

### *Rechtsextremisme*

Tot extreemrechts gedacht goedgevoerd worden denkbeelden gerekend zoals vreemdelingenhaat (inclusief: antisemitisme en antimoslims) en haat jegens vreemde (cultuur)elementen en ultranationalisme. Hierbij houden *extreemrechtse* activisten zich veelal aan de wet, overtreden *rechtsextremisten* de wet en extreemrechtse *terroristen* plegen geweld gebaseerd op extreemrechtse opvattingen. In dit rapport baseren we ons op de volgende driedeling van de AIVD (2023b) tussen activisme, extremisme en terrorisme. (1) Activisme kenmerkt zich door de afwezigheid van ondemocratische doelen of middelen (en is daarmee niet in strijd met de wet); (2) extremisme kenmerkt zich door het nastreven of gebruiken van ondemocratische doelen en middelen, waardoor het in strijd is met de wet, maar niet per definitie gewelddadig is; en (3) verondersteld wordt dat extremisme in zijn uiterste vorm kan leiden tot terrorisme. Onder terrorisme vallen acties die maatschappij-ontwrichtende schade aanrichten en/of grote groepen mensen angst aanjagen. Zie paragraaf 1.1 voor een nadere toelichting de definitie van rechtsextremisme.

### *Red-pilling*

De term 'Red-pilling' is ontleent uit de film *The Matrix* waarin de hoofdpersoon Neo moet kiezen tussen de rode en de blauwe pil. Het innemen van de rode pil zorgt ervoor dat hij de 'waarheid' kan zien en zich zal realiseren dat hij weer uit 'vrije wil' kan handelen. De blauwe pil heeft tot gevolg dat Neo doorgaat met het leven zoals het al was – een wereld vol leugens. Rechtsextremisten passen het onderscheid tussen deze pillen toe op de huidige samenleving waarin 'bluepillers' – ook wel 'normies' – de media en het systeem blindelings geloven, alsook de veronderstelde leugens die zij uitdragen en vertegenwoordigen. De 'redpillers' daarentegen zien volgens hen de wereld voor hoe deze 'echt' is en hebben het systeem dat bestaat uit allerlei leugens door. Zie paragraaf 3.3 voor een nadere toelichting op red-pilling.

### *Trials*

We hebben op YouTube en TikTok zogeheten 'online trials' uitgevoerd. Hierin is er onderzocht in hoeverre algoritmen op basis van gepersonaliseerd zoekgedrag door gebruikers kunnen leiden tot suggesties voor extreme(re) content. Wij voerden deze trials uit als een aanvulling op de literatuurstudie. Zie hoofdstuk 5 voor een andere beschrijving van de uitgevoerde trials.

### *Trusted Flaggers*

Trusted flaggers zijn instanties of personen met bijzondere expertise en competentie voor het opsporen, identificeren en melden van illegale content of andere overtredingen van de wet op onder meer sociale mediaplatforms. De Digital Services Act (artikel 19) stelt dat online platforms meldingen van zogenaamde betrouwbare 'flaggers' met prioriteit moeten behandelen. Alleen 'entiteiten' die hebben aangetoond over deskundigheid en bekwaamheid te beschikken om illegale inhoud aan te pakken; collectieve belangen vertegenwoordigen; en onafhankelijk zijn van enig online platform kunnen de aanduiding van 'betrouwbare flagger' verkrijgen. Een entiteit kan de status van betrouwbare 'flagger' verkrijgen door dit aan te vragen bij de coördinator voor digitale diensten van de lidstaat waarin de aanvrager is gevestigd.

### *Vikingcultuur*

De zogeheten 'Vikingcultuur' is nauw verwant aan het (neo-)paganisme (nieuwe heidendom) en kent veel gelijkwaardige symbolen. Hoewel de meeste aanhangers van de Vikingcultuur de progressieve waarden van de bredere paganistische gemeenschap delen, is er aan de flanken sprake van een aantrekkingskracht van sommige ideeën die verwant zijn aan het rechtsextremistische narratief. Deze aantrekking ligt onder meer in de 'eigen' voorchristelijke cultuur, vrij van diversiteit en migratie evenals de pseudowetenschappelijke theorie dat Europa oorspronkelijk enkel door 'pure blanken' werd bevolkt en de bijhorende wens om dit te herstellen en behouden. Het merendeel van de paganisten verwerpen deze ideeën met klem en pleiten sterk voor inclusiviteit. Zie paragraaf 3.6 voor nadere toelichting op Vikingcultuur.

## Bijlage 2 Referenties

Abbas, T. (2017). "Ethnicity and Politics in Contextualising Far Right and Islamist Extremism." *Perspectives on Terrorism* 11, no. 3: 54–61.

Aiston, J. (2021). What is the manosphere and why is it a concern? *Internet Matters.Org*. Geraadpleegd via: <https://www.internetmatters.org/hub/news-blogs/what-is-the-manosphere-and-why-is-it-a-concern/>.

Alava, S., Chaouni, N., & Charles, Y. (2020). How to characterise the discourse of the far-right in digital media? *Interdisciplinary approach to preventing terrorism. International Conference on Knowledge-Based Intelligent Information & Engineering Systems*.

Alizadeh, M., Weber, I., Cioffi-Revilla, C., Fortunato, S., & Macy, M. (2019). Psychology and morality of political extremists: Evidence from Twitter language analysis of alt-right and antifa. *EPJ Data Science*, 8(17). <https://doi.org/10.1140/epjds/s13688-019-0193-9>.

Algemene Inlichtingen- en Veiligheidsdienst (2018). Rechts-extremisme in Nederland. Een fenomeen in beweging. Geraadpleegd via: [https://www.aivd.nl/binaries/aivd\\_nl/documenten/publicaties/2018/10/02/rechts-extremisme-in-nederland-een-fenomeen-in-beweging/Rechts-extremisme-in+Nederland+-+een-fenomeen-in+beweging.pdf](https://www.aivd.nl/binaries/aivd_nl/documenten/publicaties/2018/10/02/rechts-extremisme-in-nederland-een-fenomeen-in-beweging/Rechts-extremisme-in+Nederland+-+een-fenomeen-in+beweging.pdf).

Algemene Inlichtingen- en Veiligheidsdienst (2023a). Rechts-extremisme. Wat verstaan wij onder deze vorm van extremisme? Geraadpleegd via: <https://www.aivd.nl/onderwerpen/extremisme/rechts-extremisme>.

Algemene Inlichtingen- en Veiligheidsdienst (2023b). Extremisme. Wat verstaat de AIVD onder extremisme? Geraadpleegd via: <https://www.aivd.nl/onderwerpen/extremisme>.

Algotransparency (weblink). Our manifesto. Why the world needs algorithmic transparency. Geraadpleegd via: <https://www.algotransparency.org/our-manifesto.html>.

Alonso, R., Bjørgo, T., Della Porta, D., Coolseat, R., Khosrokhavar, F., Lohlker, R., Ranstorp, M., Reinares, F., Schmid, A., Silke, A., Taarnby, M., & De Vries, G. (2008). *Radicalisation Processes Leading to Acts of Terrorism*. Geraadpleegd via <https://biblio.ugent.be/publication/446365/file/6814706>.

Archetti, C. (2012). *Understanding terrorism in the age of global media: A communication approach*. Springer.

Baier, D., Manzoni, P. & Bergmann, M. (2016). Einflussfaktoren des politischen Extremismus im Jugendalter — Rechtsextremismus, Linksextremismus und islamischer Extremismus im Vergleich / Influencing Factors of Political Extremism in Adolescence — Right-Wing Extremism, Left-Wing Extremism and Islamic Extremism Compared. *Monatsschrift für Kriminologie und Strafrechtsreform / Journal of Criminology and Penal Reform*, 99(3), 171-198 <https://doi.org/10.1515/mkr-2016-0302>.

Beauchamp, Z. (2018). "Incel, the misogynist ideology that inspired the deadly Toronto attack, explained." *Vox*. Geraadpleegd via: <https://www.vox.com/world/2018/4/25/17277496/Inceltoronto-attack-alekminassian>.

Bartlett, J., Birdwell, J., & King, M. (2010). *The edge of violence: a radical approach to extremism*. London: Demos. Geraadpleegd via [https://www.demos.co.uk/files/Edge\\_of\\_Violence\\_-\\_web.pdf](https://www.demos.co.uk/files/Edge_of_Violence_-_web.pdf).

Barton, G. (2019). Christchurch attacks are a stark warning of toxic political environment that allows hate to flourish. ABC News. Geraadpleegd via: <https://www.abc.net.au/news/2019-03-17/christchurch-attacks-stark-warning-of-toxic-hate-flourishing/10909394>.

Bedingfield, W. (2021). The far right are running riot on Steam and Discord. WIRED UK. Geraadpleegd via: <https://www.wired.co.uk/article/steam-discord-far-right>.

Beelmann, A. (2020). A social-developmental model of radicalization: A systematic integration of existing theories and empirical research. *International Journal of Conflict and Violence*, 14, 1–14. <https://doi.org/10.4119/ijcv-3778>.

Bell, K. (2022). YouTube could 'break' sharing on borderline content to fight misinformation. Engadget. Geraadpleegd via: <https://www.engadget.com/YouTube-could-break-sharing-on-borderline-content-to-fight-misinformation-201819354.html>.

Benslama, F. (2017). *Der Übermuslim: Was junge Menschen zur Radikalisierung treibt [What Drives Young People to Radicalization]*. Matthes & Seitz.

Berardi, F. (2008). "(T)error and poetry," *Radical Philosophy*, nr. 149, pp. 39–45. Geraadpleegd via: <https://www.radicalphilosophyarchive.com/article/terror-and-poetry>.

Bergmann, M. C., Kliem, S., Krieg, Y., & Beckmann, L. (2019). *Jugendliche in Niedersachsen: Ergebnisse des Niedersachsensurveys 2017 [Young people in Lower Saxony: Results of the Lower Saxony survey 2017]* (KFN-Research Reports, Vol. 141). Kriminologisches Forschungsinstitut Niedersachsen.

Binder, J., & Kenyon, J. (2022). "Terrorism and the internet: How dangerous is online radicalization?" *Frontiers in Psychology*, vol. 13. Geraadpleegd via: <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.997390/full>.

BitChute (g.d.). *Community Guidelines*. Geraadpleegd via (2023, januari): <https://support.bitchute.com/policy/guidelines/>.

Bjørge, T., & Carlsson, Y. (2005). *Early intervention with violent and racist youth groups*. Working paper No. 677. Oslo: Norwegian Institute of International Affairs. Geraadpleegd via: <http://nupi.brage.unit.no>.

Bjørge, T. & Horgan J., eds. (2009). *Leaving Terrorism Behind: Individual and Collective Disengagement*. New York: Routledge.

Bliuc, A., Betts, J., Vergani, M., Iqbal, M. & Dunn, K. (2020). The growing power of online communities of the extreme-right: deriving strength, meaning, and direction from significant socio-political events 'in real life'. *International Centre for Counter-Terrorism*. Geraadpleegd via: <https://www.jstor.org/stable/resrep25261?seq=1>.

Böckler, N. & Zick, A. (2015). *Wie gestalten sich Radikalisierungsprozesse im Vorfeld jihadistisch-terroristischer Gewalt? Perspektiven aus der Forschung*. In D. Molthagen (Ed.), *Handlungsempfehlungen zur Auseinandersetzung mit islamistischem Extremismus und Islamfeindlichkeit: Arbeitsergebnisse eines Expertengremiums der Friedrich-Ebert-Stiftung* (pp. 99–121). Friedrich-Ebert-Stiftung, Forum Berlin.

Boehnke, K., Hagan, J., & Merkens, H. (1998). Right-wing extremism among German adolescents: Risk factors and protective factors. *Applied Psychology*, 47, 109-126.

Bolle, J. (2022, 30 december). Vrouwonvriendelijke influencer Andrew Tate in Roemenië gearresteerd op verdenking van mensenhandel. De Volkskrant. Geraadpleegd via: <https://www.volkskrant.nl/nieuws-achtergrond/vrouwonvriendelijke-influencer-andrew-tate-in-roemenie-gearresteerd-op-verdenking-van-mensenhandel~b47a26c9/>.

Borger, G. (2021). Josh Hawley and the 'left-wing attack on manhood'. CNN Politics. Geraadpleegd via: <https://edition.cnn.com/2021/11/08/politics/josh-hawley-masculinity-video-games-porn/index.html>.

Borum, R. (2011). Radicalization into violent extremism I: A review of social science theories. *Journal of Strategic Security*, 4, 7-35.

Bouhana, N., Corner, E., Gill, P., & Schuurman, B. (2018). Background and Preparatory Behaviours of Right-Wing Extremist Lone Actors: A Comparative Study. *Perspectives on Terrorism*, 12(6), 150–163. <https://www.jstor.org/stable/26544649>.

Bouma, R. (2020). Amerikaanse complottheorie QAnon ook in Nederland in opkomst. Nederlandse Omroep Stichting. Geraadpleegd via: <https://nos.nl/nieuwsuur/artikel/2349814-amerikaanse-complottheorie-qanon-ook-in-nederland-in-opkomst>.

Bowman, E. & Stewart, I. (2017). The Woman Behind The 'Alt-Right'. National Public Radio. Geraadpleegd via: <https://www.npr.org/2017/08/20/544134546/the-women-behind-the-alt-right>.

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, 15(3), 209–227. <https://doi.org/10.1007/s10676-013-9321-6>.

Braddock, K. (2019). Vaccinating Against Hate: Using Attitudinal Inoculation to Confer Resistance to Persuasion by Extremist Propaganda. *Terrorism and Political Violence*, 1–23. <https://doi.org/10.1080/09546553.2019.1693370>.

Broderick, R. (2022, 16 mei). You can't always blame algorithms. *Garbage Day*. Geraadpleegd via: <https://www.garbageaday.email/p/you-cant-always-blame-algorithms>.

Brown, K. (2019). Gender-specific approaches in exit work. *Radicalisation Awareness Network EXIT*. Geraadpleegd via [https://home-affairs.ec.europa.eu/system/files\\_en?file=2019-12/ran\\_exit\\_gender\\_specific\\_approaches\\_rome\\_22-23\\_102019\\_en.pdf](https://home-affairs.ec.europa.eu/system/files_en?file=2019-12/ran_exit_gender_specific_approaches_rome_22-23_102019_en.pdf).

Brown, M., Bisbee, J., Lai, A., Bonneau, R., Nagler, J. & and Tucker, J. (2022). Echo Chambers, Rabbit Holes, and Algorithmic Bias: How YouTube Recommends Content to Real Users. Geraadpleegd via: SSRN: <https://ssrn.com/abstract=4114905> or <http://dx.doi.org/10.2139/ssrn.4114905>.

Brunt, B., Murphy, A., & Zedginidze, A. (2017). An Exploration of the Risk, Protective, and Mobilization Factors Related to Violent Extremism in College Populations. *Violence and Gender*. 4. 81-101. 10.1089/vio.2017.0039.

Brzuszkiewicz, S. (2020). Incel radical milieu and external locus of control. *International Centre for Counter-Terrorism (ICCT)*. Geraadpleegd via: <https://journals.sagepub.com/doi/pdf/10.1177/1781685820915972>.

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, 20(1), 30–44. <https://doi.org/10.1080/1369118X.2016.1154086>.

Burley, S. (2020). Total Life Reform. The Real Consequences of the Far Right's Self-Help Grift. Political Research Associates. Geraadpleegd via: <https://politicalresearch.org/2020/11/09/total-life-reform>.

Caramellino, D., Melacarne, C., Ducol, B. (2022). Transformative Learning and Microradicalization. In: Nicolaidis, A., Eschenbacher, S., Buergelt, P.T., Gilpin-Jackson, Y., Welch, M., (eds) The Palgrave Handbook of Learning for Transformation. Palgrave Macmillan, Cham: Londen. Geraadpleegd via: [https://link.springer.com/chapter/10.1007/978-3-030-84694-7\\_43#citeas](https://link.springer.com/chapter/10.1007/978-3-030-84694-7_43#citeas).

Chen, A., Nyhan, B., Reifler, J., Robertson, R., & Wilson, C. (2022). Subscriptions and external links help drive resentful users to alternative and extremist YouTube channels. Geraadpleegd via: <https://cpb-us-e1.wpmucdn.com/sites.dartmouth.edu/dist/5/2293/files/2022/04/YouTube.pdf>.

Chermak, S., & Gruenewald, J. (2015) Laying a Foundation for the Criminological Examination of Right-Wing, Left-Wing, and Al Qaeda-Inspired Extremism in the United States, *Terrorism and Political Violence*, 27:1, 133-159.

Chi, C. (2021, 15 april). YouTube Algorithm: The Constantly Updated Guide to YouTube's Updates & Changes. Geraadpleegd via: <https://blog.hubspot.com/marketing/YouTube-algorithm>.

Conway, M., Scrivens, R. & Macnair, L. (2019). Right-Wing Extremists' Persistent Online Presence: History and Contemporary Trends. ICCT. <https://doi.org/10.19165/2019.3.12>.

Copeland, S. & Marsden, S. (2020). Right-Wing Terrorism: Pathways and Protective Factors. CREST. Geraadpleegd via <https://crestresearch.ac.uk/download/3418/20-004-01.pdf>.

Corner, E., Gill, P., & Mason, O. (2016). Mental Health Disorders and the Terrorist: A Research Note Probing Selection Effects and Disorder Prevalence. *Studies in Conflict & Terrorism*, 39(6), 560-568. doi:10.1080/1057610X.2015.1120099.

Counter-Terrorism Committee Executive Directorate (2020, april). Member states concerned by the growing and increasingly transnational threat of extreme right-wing terrorism. CTED Trends Alert. Geraadpleegd via: [https://www.un.org/securitycouncil/ctc/sites/www.un.org.securitycouncil.ctc/files/files/documents/2021/Jan/cted\\_trends\\_alert\\_extreme\\_right-wing\\_terrorism.pdf](https://www.un.org/securitycouncil/ctc/sites/www.un.org.securitycouncil.ctc/files/files/documents/2021/Jan/cted_trends_alert_extreme_right-wing_terrorism.pdf).

Covington, P., Adams, J., & Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. Geraadpleegd via: <https://storage.googleapis.com/pub-tools-public-publication-data/pdf/45530.pdf>.

Crenshaw, M. (2011). *Explaining Terrorism: Causes, Processes, and Consequences*. London; New York: Routledge.

Crosset, V., Tanner, S. & Campana, A. (2018). Researching far right groups on Twitter: Methodological challenges 2.0. *New Media & Society*, 21(4), 939–961. <https://doi.org/10.1177/1461444818817306>.

Dalgaard-Nielsen, A. (2008). "Studying violent radicalization in Europe I: The potential contribution of socio-psychological and psychological approaches," Danish Institute for International Studies, Working Paper, number 2008/2.

Damen, F. & Kraak, H. (2021). 'Een gewelddadige cultuur die doordrenkt is van een geromantiseerd ondergangsdnken'. In: *De Volkskrant*. Geraadpleegd via: <https://www.volkskrant.nl/nieuws-achtergrond/een-gewelddadige-cultuur-die-doordrenkt-is-van-een-geromantiseerd-ondergangsdnken~b-56cd597/>.



Davey, J. (2021). Gamers Who Hate: An Introduction to ISD's Gaming and Extremism Series. Institute for Dialogue. Geraadpleegd via: <https://www.isdglobal.org/wp-content/uploads/2021/08/01-gaming-report-intro.pdf>.

Davey, J., & Ebner, J. (2017). The Fringe Insurgency: Connectivity, Convergence and Mainstreaming of the Extreme Right. Institute for Strategic Dialogue (ISD). Geraadpleegd via: [https://www.isdglobal.org/wp-content/uploads/2017/10/The-Fringe-Insurgency-221017\\_2.pdf](https://www.isdglobal.org/wp-content/uploads/2017/10/The-Fringe-Insurgency-221017_2.pdf)

Daymon, C., & Margolin, D. (2022). Women in American violent extremism: An examination of far-right and Salafi jihadist movements. Washington, D.C.: Program on Extremism at George Washington University.

Decker, O., Kiess, J., Schuler, J., Handke, B., Pickel, G. & Brähler, E. (2020). Die Leipziger Autoritarismus-Studie 2020: Methode, Ergebnisse und Langzeitverlauf [The Leipzig Authoritarianism Study 2020: Method, Results, and Long-Term Trajectory]. In O. Decker & E. Brähler (Eds.), *Autoritäre Dynamiken. Alte Ressentiments – neue Radikalität. Leipziger Autoritarismus Studie 2020* (pp. 27–88). Psychosozial-Verlag.

Demant, F., Slootman, M., Buijs, F., & Tillie, J. N. (2008). Teruggang en uittreding: processen van deradicalisering ontleed. IMES.

DeVito, M. A. (2016). From Editors to Algorithms. *Digital Journalism*, 1–21. <https://doi.org/10.1080/21670811.2016.1178592>.

De Wit, G. Nieuw-rechts en het heidendom. De Fabel van de illegaal, 38/39. Geraadpleegd via: <https://www.doorbraak.eu/gebladerte/10651f38.htm>.

De Wolf, A. & Doosje, B. (2010/2015). Aanpak van radicalisme. Een psychologische analyse. Amsterdam: SWP.

De Zutter, J. (2010) Heidenen voor het blok: radicaal-rechts en het nieuwe heidendom. Houtekiet.

Dier, A. & Baldwin, G. (2022). Masculinities and Violent Extremism. International Peace Institute & UN Security Council Counter-Terrorism Committee Executive Directorate. Geraadpleegd via <https://www.ipinst.org/wp-content/uploads/2022/06/Masculinities-and-VE-Web.pdf>.

Dijkman, A. (2021). 'De Radicaliseringsindustrie' – Van pionieren naar professionaliseren. Amsterdam University Press: Amsterdam.

Doosje, B., Moghaddam, F., Kruglanski, A., Wolf, A., Mann, L., & Feddes, A. (2016). Terrorism, radicalization and de-radicalization. *Current Opinion in Psychology*. 11. 79-84. 10.1016/j.copsyc.2016.06.008.

Doosje, B., Van den Bos, K., Loseman, A., Feddes, A.R. and Mann, L. (2012), "My In-group is Superior!": Susceptibility for Radical Right-wing Attitudes and Behaviors in Dutch Youth. *Negotiation and Conflict Management Research*, 5: 253-268. <https://doi.org/10.1111/j.1750-4716.2012.00099.x>.

Dreigingsbeeld Terrorisme Nederland nr. 56 (2022). Geraadpleegd via: <https://open.overheid.nl/documenten/ronl-764a8385e3f406fd4be3129e1c949f11c592cbcc/pdf>.

Ducol, B. (2015). "A Radical sociability: in defense of an online/offline multi-dimensional approach to radicalization," in *Social Networks, Terrorism and Counter-Terrorism: Radical and Connected*. New York: Routledge, 82–104.

Duff, M. (2019). Gaming culture and the alt-right: The weaponisation of hate. Stuff. co.nz.

Eberl, M. (2019). Privacy Analysis of TikTok's App and Website. Rufposten. Geraadpleegd via: <https://rufposten.de/blog/2019/12/05/privacy-analysis-of-tiktoks-app-and-website/>.

Ebner, J. (2017). *The Rage: The Vicious Circle of Islamist and Far-Right Extremism*. Bloomsbury Publishing.

Ebner, J. (2019). Counter-Creativity: Innovative Ways to Counter Far-Right Communication Tactics, in M. Fielitz & N. (eds.), *Post-Digital Cultures of the Far Right Online Actions and Offline Consequences in Europe and the US*. Bielefeld: Transcript.

Ebner, J. (2020). *Going Dark. The Secret Social Lives of Extremists*. Bloomsbury Publishing Plc: Londen.

Eslami, M. (2015). "I always assumed that I wasn't really that close to [her]". CHI'15 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 153–162. <https://doi.org/10.1145/2702123.2702556>.

Emmelkamp, J., Asscher, J. J., Wissink, I. B., & Stams, G. J. J. M. (2020). Risk factors for (violent) radicalization in juveniles: A multilevel meta-analysis. *Aggression and Violent Behavior*. <https://doi.org/10.1016/j.avb.2020.101489>.

Emore & Leader (2020) *Paganism and Its Discontents: Enduring Problems of Racialized Identity*, Cambridge Scholars Publishing, retrieved <https://www.cambridgescholars.com/resources/pdfs/978-1-5275-5770-3-sample.pdf>.

Europese Commissie (2018). "Richtlijnen". Publicatieblad van de Europese Unie. Geraadpleegd via: <https://support.bitchute.com/policy/guidelines/>.

Europese Commissie. (2020). *The Digital Services Act: ensuring a safe and accountable online environment*. European Commission Website. Geraadpleegd via: [https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment\\_en](https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en).

European Commission. (2020b). *The Digital Markets Act: ensuring fair and open digital markets*. European Commission Website. [https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets\\_en](https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en).

Europese Commissie (2022). "Verordeningen". Publicatieblad van de Europese Unie. Geraadpleegd via: <https://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=CELEX:32022R1925&from=EN>.

Europese Commissie. (2022b). *European Media Freedom Act*. European Commission Website. Geraadpleegd via: [https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/european-democracy-action-plan/european-media-freedom-act\\_en](https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/european-democracy-action-plan/european-media-freedom-act_en).

Europese Commissie (2022c). *Voorstel voor een VERORDENING VAN HET EUROPEES PARLEMENT EN DE RAAD tot vaststelling van een gemeenschappelijk kader voor mediadiensten op de interne markt ("verordening media-vrijheid") en tot wijziging van Richtlijn 2010/13/EU*. Geraadpleegd via: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0457>.

Europese Commissie (2023, 5 mei). *The Digital Services Act package*. Geraadpleegd via: <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package> Gab (2023, 18 januari). *Terms of Service*. Geraadpleegd via: <https://gab.com/about/tos>.

Expertise-unit Sociale Stabiliteit (2021). Factsheet Online radicaal- en extremistisch gedrag. Geraadpleegd via: <https://www.platformjep.nl/binaries/platformjep/documenten/publicaties/2021/02/12/factsheet-online-radicaal-en-extremistisch-gedrag/> Factsheet+online+radicaal+en+extremistisch+gedrag.pdf.

Feddes, A. R., Mann, L., & Doosje, B. (2013). Empirical study as part of a scientific approach to finding indicators of and responses to radicalisation (SAFIRE). Report presented to the European Commission.

Feddes, A. R., Nickolson, L., & Doosje, B. (2015). Triggerfactoren in het radicaliseringsproces. Expertise-unit Sociale Stabiliteit/Universiteit van Amsterdam.

Feddes, A. R., Huijzer, A., Van Ooijen, I., & Doosje, B. (2019). Fortress of democracy: Engaging youngsters in democracy results in more support for the political system. *Peace and Conflict*, 25(2), 158–164. <https://doi.org/10.1037/pac0000380>.

Fielitz, M., & Ahmed, R. (2021). It's not funny anymore. Far-right extremists' use of humour. Radicalisation Awareness Network. Geraadpleegd via: [https://home-affairs.ec.europa.eu/system/files/2021-03/ran\\_ad-hoc\\_pap\\_fre\\_humor\\_20210215\\_en.pdf](https://home-affairs.ec.europa.eu/system/files/2021-03/ran_ad-hoc_pap_fre_humor_20210215_en.pdf).

Finnerty, C., & Grelich, K. (2020). The use of social media and alt-tech platforms by threat actors. The Counterterrorism Group. Geraadpleegd via: <https://www.counterterrorismgroup.com/post/the-use-of-social-media-and-alt-tech-platforms-by-threat-actors>.

Floridi, L. (2016). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford: Oxford University Press.

Freelon, D., Marwick, A., & Kreiss, D. (2020). False equivalencies: Online activism from left to right. *Science* (New York, N.Y.), 369(6508), 1197–1201. <https://doi.org/10.1126/science.abb2428>.

Fuchs, M. (2003). Rechtsextremismus von Jugendlichen: Zur Erklärungskraft verschiedener theoretischer Konzepte. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 55, 654-678.

Gambetta, D., & Hertog, S. (2017). *Engineers of jihad: The curious connection between violent extremism and education*. Princeton, USA: Princeton University Press.

Gaudette, T. (2020). Upvoting Extremism: Collective identity formation and the extreme right on Reddit. *New Media and Society*. <https://doi.org/10.1177/1461444820958123>.

Gaudette, T., Scrivens, R. & Venkatesh, V. (2020). The Role of the Internet in Facilitating Violent Extremism: Insights from Former Right-Wing Extremists. *Terrorism and Political Violence*. Geraadpleegd via: <https://doi.org/10.1080/09546553.2020.1784147>.

Geyser, W. (2023, 6 maart). How Does the YouTube Algorithm Work: A Peek into its Changes in 2023. *Influencer Marketing Hub*. Geraadpleegd via: <https://influencermarketinghub.com/how-does-the-YouTube-algorithm-work/#toc-3>.

Gielen, A.J. (2008) *Radicalisering en identiteit: Radicale rechtse en moslimjongeren vergeleken*. Amsterdam: Aksant.

GIFCT (2021, juli). Content-Sharing Algorithms, Processes, and Positive Interventions Working Group. Part 1: Content-Sharing Algorithms and Processes. *Global Internet Forum to Counter Terrorism*. Geraadpleegd via: <https://gifct.org/wp-content/uploads/2021/07/GIFCT-CAP1-2021.pdf>.

Gill, P., Corner, E., Thornton, A. & Conway, M. (2015). What are the roles of the internet in terrorism? Measuring online behaviours of convicted UK terrorists. *Voxpol*. Geraadpleegd via: [https://www.voxpol.eu/download/vox-pol\\_publication/What-are-the-Roles-of-the-Internet-in-Terrorism.pdf](https://www.voxpol.eu/download/vox-pol_publication/What-are-the-Roles-of-the-Internet-in-Terrorism.pdf).

Gill, P., Corner, E., Conway, M., Thornton, A., Bloom, M. & Horgan, J. (2017). Terrorist Use of the Internet by the Numbers. Quantifying Behaviours, Patterns, and Processes. *Criminology & Public Policy*, vol. 16(1), 99-117. Geraadpleegd via: <https://onlinelibrary.wiley.com/doi/full/10.1111/1745-9133.12249>.

Ging, D. (2019). Alphas, betas, and incels: Theorizing the masculinities of the manosphere. *Men and Masculinities*, 22(4), 638-657.

Google LLC (webpagina, g.d.(a)). Google Privacy Policy. Geraadpleegd via (18 januari 2023): [https://www.gstatic.com/policies/privacy/pdf/20221215/qi5mvafl/google\\_privacy\\_policy\\_en-GB\\_eu.pdf](https://www.gstatic.com/policies/privacy/pdf/20221215/qi5mvafl/google_privacy_policy_en-GB_eu.pdf).

Google LLC (webpagina, g.d. (b)). YouTube Help. Je aanbevelingen en zoekresultaten beheren. Geraadpleegd via (18 januari 2023): <https://support.google.com/YouTube/answer/6342839>.

Google LLC (webpagina, g.d. (c)). YouTube Help. About the YouTube Flagger program. Geraadpleegd via: <https://support.google.com/YouTube/answer/7554338?hl=en>.

Grant, J., & MacDonald, F. (2018). "The New/"Alt" Right, Toxic Masculinity, and Violence". Geraadpleegd via: <https://www.cpsa-acsp.ca/documents/conference/2018/783.Grant.MacDonald.pdf>.

Grant, J., & MacDonald, F. (2018). *Turbulent Times, Transformational Possibilities?: Gender and Politics Today*, University of Toronto. Geraadpleegd via: [https://books.google.be/books?hl=en&lr=&id=P8bhDwAAQBAJ&oi=fnd&pg=PA368&dq=Grant+%26+MacDonald,+evolution+of+passive+to+active+sexuality&ots=O5luca56cs&sig=5WvIEPFfows\\_8-hzW2Ukc9TjzzI#v=onepage&q=LGBTI&f=false](https://books.google.be/books?hl=en&lr=&id=P8bhDwAAQBAJ&oi=fnd&pg=PA368&dq=Grant+%26+MacDonald,+evolution+of+passive+to+active+sexuality&ots=O5luca56cs&sig=5WvIEPFfows_8-hzW2Ukc9TjzzI#v=onepage&q=LGBTI&f=false) p376.

Greve, W. (2007). Selbst und Identität im Lebenslauf [Self and identity in the life course]. In J. Brandstädter & U. Lindenberger (Eds.), *Entwicklungspsychologie der Lebensspanne* (pp. 305–336). Kohlhammer.

Grover, T., & Mark, G. (2012) "Detecting Potential Warning Behaviors of Ideological Radicalization in an Alt-Right Subreddit." *Proceedings of the International AAAI Conference on Web and Social Media* 13: 193–204.

Gualeni, S. (2015). *Virtual worlds as philosophical tools: How to philosophize with a digital hammer*. New York: Palgrave Macmillan.

Hart, M., Davey, J., Maharasingam-Shah, E., Gallagher, A., & O'Connor, C. (2021). *An Online Environmental Scan of Right-wing Extremism in Canada*. ISD Global.

Hassan, G., Brouillette-Alarie, S., Alava, S., Frau-Meigs, D., Lavoie, L., Fetiu, A., Varela, W., Borokhovski, E., Venkatesh, V., Rousseau, C., & Sieckelinc, S. (2018). Exposure to extremist online content could lead to violent radicalization: A systematic review of empirical evidence. *International Journal of Developmental Science*, 12(1-2), 71–88. <https://doi.org/10.3233/DEV-170233>.

Hassan, G., Brouillette-Alarie, S., Ousman, S., Kilinc, D., Savard, É. L., Varela, W., Lavoie, L., Fetiu, A., Harris-Hogan, S., Borokhovski, E., Pickup, D., Madriaza, P., Rousseau, C., Thompson, S. K., McCoy, J., Venkatesh, V., Boivin, M., Srimathi Narayana, M., Morin, D., ... the CPN-PREV team. (2021). A systematic review on the outcomes of primary and secondary prevention programs in the field of violent radicalization. *Canadian Practitioners Network for the Prevention of Radicalization and Extremist Violence*. Geraadpleegd via: <https://cpnprev.ca/systematic-review-2/>.

Heldt, A. (2020). Borderline speech: caught in a free speech limbo? Internet Policy Review. Journal on Internet Regulation. Geraadpleegd via: <https://policy-review.info/articles/news/borderline-speech-caught-free-speech-limbo/1510>.

Hermansson, P., Lawrence, D., Mulhall, J., & Murdoch, S. (2020). The International Alt-Right. Routledge <https://doi.org/10.4324/9780429032486>.

Hern, A. (2022). The TikTok takeover. How TikTok's algorithm made it a success: 'It pushes the boundaries'. The Guardian. Geraadpleegd via: <https://www.theguardian.com/technology/2022/oct/23/tiktok-rise-algorithm-popularity>.

Hernández-Ramírez, R. (2017). "Technology and self-modification: Understanding technologies of the self after Foucault," Journal of Science and Technology of the Arts, vol. 9, nr. 3, at <http://artes.ucp.pt/citarj/article/view/423>.

H.M. Government. (2019). Online Harms White Paper. The Stationary Office. Geraadpleegd via: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/793360/Online\\_Harms\\_White\\_Paper.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf).

Hosseinmardi, H., Ghasemian, A., Clauset, A., Rothschild, D., Mobius, M., & Duncan J. Watts, D. (2020) "Evaluating the scale, growth, and origins of right-wing echo chambers on YouTube." Physics. Geraadpleegd via: <https://arxiv.org/abs/2011.12843>.

Institute of Strategic Dialogue. (2021). Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok. [https://www.isdglobal.org/wp-content/uploads/2021/08/HateScape\\_v5.pdf](https://www.isdglobal.org/wp-content/uploads/2021/08/HateScape_v5.pdf).

Jost, J.T., Glaser, J., Kruglanski, A.W., & Sulloway, F.J. (2003). Political conservatism as motivated social cognition. Psychological bulletin, 129(3), 339-375.

Kaldor, S. (2021). Extremist Manifestos through the Lens of Ressentiment. International Centre for Counter-Terrorism (ICCT) Research Paper. Geraadpleegd via <https://icct.nl/app/uploads/2021/05/Far-Right-Violent-Extremism-as-a-Failure-of-Status.pdf>.

Keenan, C. (2022). Een update over ons werk om misinformatie tegen te gaan. TikTok Newsroom. Geraadpleegd via: <https://newsroom.tiktok.com/nl-nl/een-update-over-ons-werk-om-desinformatie-tegen-te-gaan>.

Keller, D. (2021). Amplification and Its Discontents. Why regulating the reach of online content is hard. Knight First Amendment Institute at Columbia University. Geraadpleegd via: <https://knightcolumbia.org/content/amplification-and-its-discontents>.

Kennisnet, 2017. Monitor Jeugd en Media 2017. Zoetermeer: Kennisnet.

Kenyon, J., Binder, J., and Baker-Beall, C. (2021b). Exploring the role of the internet in radicalization and offending of convicted extremists. Geraadpleegd via: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1017413/exploring-role-internet-radicalisation.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1017413/exploring-role-internet-radicalisation.pdf).

Kessler, T., Mummendey, A., & Klink, A. (1999). Soziale Identität und relative Deprivation: Determinanten individuellen und kollektiven Verhaltens in Ostdeutschland nach der Vereinigung. In M. Schmitt & L. Montada (Eds.), Gerechtigkeitserleben im wiedervereinigten Deutschland (pp. 213–262). Leske + Budrich.

Kim, S.A. (2017). "Social Media Algorithms: Why You See What You See". 2 GEO. L. TECH. REV. 147. Geraadpleegd via: <https://georgetownlawtechreview.org/social-media-algorithms-why-you-see-what-you-see/GLTR-12-2017/>.

Kimmel, M. (2018). Almost all violent extremists share one thing: their gender. *The Guardian*. Geraadpleegd via: <https://www.theguardian.com/world/2018/apr/08/violent-extremists-share-one-thing-gender-michael-kimmel>.

King, M., & Taylor, D. (2011). The radicalization of home grown Jihadists. A review of theoretical models and social psychological evidence. *Terrorism and Political Violence*, 23, 602–622.

Klandermands, B., & Mayer, N. (Eds.). (2006). *Extreme right activists in Europe: Through the magnifying glass*. London and New York: Routledge.

Kleuver, J., Dijk, B., & Ramirez, A. (2017). *Zorgelijke Signalen*. Geraadpleegd via [https://www.dsp-groep.nl/wp-content/uploads/11bd\\_aanpak\\_radicalisering\\_DSP\\_2017.pdf](https://www.dsp-groep.nl/wp-content/uploads/11bd_aanpak_radicalisering_DSP_2017.pdf).

Koehler, D. (2014). The radical online: Individual radicalization processes and the role of the internet. *Journal for Deradicalization*, 1, 116–134.

Köhler, D., & Ebner, J. (2019). "Strategies and Tactics: Communication Strategies of Jihadists and Right-Wing Extremists." In *Hate Speech and Radicalisation Online*. London, UK: Institute for Strategic Dialogue.

Korteweg, R., Gohel, S., Heisbourg, F., Ranstorp, M., & De Wijk, R. (2010). Background contributing factors to terrorism: radicalization and recruitment. In M. Ranstorp (Eds.), *Understanding violent radicalization: Terrorist and jihadist movements in Europe*. Abingdon: Routledge.

Krasenberg, J. & Wouterse, L. (2019). *Understanding the Mental Health Disorders Pathway to Violent Extremism*. Ex-post-paper. Radicalisation Awareness Network.

Kreissel, P., Ebner, J., Urban, A., & Guhl, J. (2018): "Hass Auf Knopfdruck. Rechtsextreme Trollfabriken und das Ökosystem Koordinierter Hasskampagnen im Netz." Institute for Strategic Dialogue Geraadpleegd via: [https://www.isdglobal.org/wp-content/uploads/2018/07/ISD\\_Ich\\_Bin\\_Hier\\_2.pdf](https://www.isdglobal.org/wp-content/uploads/2018/07/ISD_Ich_Bin_Hier_2.pdf).

Kruglanski, A. W., Chen, X., Dechesne, M., Fishman, S., & Orehek, E. (2009). Volledig toegewijd: De motivatie van zelfmoordterroristen en de zoektocht naar persoonlijke betekenis. *Politieke psychologie*, 30(3), 331–357. <https://doi.org/10.1111/j.1467-9221.2009.00698.x>.

Kruglanski, A., Webber, D., & Koehler, D., (2019). "Right-Wing Extremism in Germany", *The Radical's Journey: How German Neo-Nazis Voyaged to the Edge and Back*. New York: Oxford Academic.

Küpper, B., & Zick, A. (2011). *Geschlecht und Vorurteile: Eine empirische Analyse [Gender and prejudice: An empirical analysis]*. In U. Birsl (Ed.), *Rechtsextremismus und Gender* (pp. 187–210). Verlag Barbara Budrich.

Küpper, B., Zick, A., & Rump, M. (2021). Rechtsextreme Einstellungen in der Mitte 2020/21 [Right-wing attitudes in the middle 2020/21]. In A. Zick & B. Küpper (Eds.), *Die geforderte Mitte: Rechtsextreme und demokratiegefährdende Einstellungen in Deutschland 2020/21* (pp. 75–111). Dietz.

Lacoviello, V., Valsecchi, G., Berent, J., Borinca, I., & Falomir-Pichastor, J. M. (2021). The Impact of Masculinity Beliefs and Political Ideologies on Men's Backlash Against Non-Traditional Men: The Moderating Role of Perceived Men's Feminization. *International Review of Social Psychology*, 34(1), 28. DOI: <http://doi.org/10.5334/irsp.588>.

Lally, C. & Bermingham, R. (2020). Online extremism. UK Parliament Post, no. 622. Geraadpleegd via: <https://post.parliament.uk/research-briefings/post-pn-0622/>.

Ledwich, M., & Zaitsev, A. (2019). Algorithmic Extremism: Examining YouTube's Rabbit Hole of Radicalization. <http://arxiv.org/abs/1912.11211>.

Lee, B. (2021). Only Playing: Extreme-Right Gamification (crestresearch.ac.uk).

Lefelon, P. & De Roeck, M. (2022). Beruchte complotdenker Yannick V. doodgeschoten bij huiszoeking in Antwerpen, ook politie beschoten. Brabants Dagblad. Geraadpleegd via: <https://www.bd.nl/buitenland/beruchte-complotdenker-yannick-v-doodgeschoten-bij-huiszoeking-in-antwerpen-ook-politie-beschoten~ad9456de/?referrer=https%3A%2F%2Fwww.google.nl%2F>.

Leidig, E. (26 oktober 2021). "We Are Worth Fighting for": Women in Far-Right Extremism. ICCT. Geraadpleegd via: <https://icct.nl/publication/women-far-right-extremism/>.

Lewis, B. & Marwick, A. E. (2017). Media Manipulation and Disinformation Online. Data & Society Research Institute. [https://datasociety.net/pubs/oh/DataAndSociety\\_MediaManipulationAndDisinformationOnline.pdf](https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf).

Liang, C.S., & Cross, M.J. (2020). White Crusade: How to Prevent Right-Wing Extremists from Exploiting the Internet. Strategic Security Analysis. Geraadpleegd via: <https://dam.gcsp.ch/files/doc/white-crusade-how-to-prevent-right-wing-extremists-from-exploiting-the-internet>.

Liht, J. & Savage, S. (2013). Preventing Violent Extremism through Value Complexity: Being Muslim Being British. *Journal of Strategic Security*, 6(4): 44-66.

Lindsay, A. 2020. Swallowing the Black Pill: A Qualitative Exploration of Incel Antifeminism within Digital Society. Geraadpleegd via: <http://researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/8915/thesisJaccess.pdf?sequence=1>.

Little, O. (2021). TikTok is prompting users to follow far-right extremist accounts. Media Matters. Geraadpleegd via: <https://www.mediamatters.org/tiktok/tiktok-prompting-users-follow-far-right-extremist-accounts>.

Llansó, E., Van Hoboken, J., Leerssen, P., & Harambam, J. (2020). Artificial Intelligence, Content Moderation, and Freedom of Expression. Transatlantic Working Group. Geraadpleegd via: <https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf>.

Lösel, F., & Farrington, D. P. (2012). Direct protective and buffering protective factors in the development of youth violence. *American Journal of Preventive Medicine*, 43(2S1), 8-23.

Lösel, F., King, S., Bender, D., & Jugl, I. (2018). Protective factors against extremism and violent radicalization: A systematic review of research. *International Journal of Developmental Science*, 12(1-2), 89-102.

Luke, T. (2021) Beyond Prepper Culture as Right-wing Extremism: Selling Preparedness to Everyday Consumers as How to Survive the End of the World on a Budget, Volume 18 -1 Fast Capitalism.

Macfarlane, S. (2022). TikTok faces growing national security concerns: "Its's not just the collection or theft of the data". CBS. Geraadpleegd via: <https://www.cbsnews.com/news/tiktok-security-concerns-china-data/>.

Malter, J. (2017). Alt-Tech platforms: A haven for fringe views online. CNN Business. Geraadpleegd via: <https://money.cnn.com/2017/11/10/technology/culture/divided-we-code-alt-tech/index.html>.

Maly, I. (2018). Nieuw rechts. Antwerpen: Epo Uitgeverij.

Marwick, A., Clancy, B. & Furl, K. (2022). Far-Right Online Radicalization: A Review of the Literature. *The Bulletin of Technology & Public Life*. <https://doi.org/10.21428/bfcb0bff.e9492a11>.

Marwick, A., & Lewis, R. (2017). "Media manipulation and disinformation online," Data & Society Research Institute. Geraadpleegd via: [https://datasociety.net/pubs/oh/DataAndSociety\\_MediaManipulationAndDisinformationOnline.pdf](https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf).

Mattheis, A. (2018). Shieldmaidens of whiteness:(Alt) maternalism and women recruiting for the far/alt-right. *Journal for Deradicalization*, (17), 128-162.

Mattheis, A. (2019). Disrupting the Digital Divide: Extremism's Integration of Offline/ Online Practice. In: *Interventionen: Zeitschrift für Verantwortungspädagogik*, issue 14: 4-18. Geraadpleegd via: [https://violence-prevention-network.de/wp-content/uploads/2020/01/Interventionen\\_14-2019.pdf](https://violence-prevention-network.de/wp-content/uploads/2020/01/Interventionen_14-2019.pdf).

McCauley, C. & Moskalenko, S. (2008). Mechanisms of political radicalization: pathways toward terrorism. *Terrorism and Political Violence*, 20(3), 415-433.

McClintock, A. (2022). The Woman Paradox: Misogyny and Women In the Far-Right. Australian Institute of International Affairs. Geraadpleegd via: <https://www.internationalaffairs.org.au/australianoutlook/the-woman-paradox-misogyny-and-women-in-the-far-right/>.

Mediawijsheid (webpagina). Filterbubbel. *Mediawijsheid.nl*. Geraadpleegd via (18 januari 2023): <https://www.mediawijsheid.nl/filterbubbel/>.

Meleagrou, A., & Kaderbhai, N. (2017). Research Perspectives on Online Radicalisation. A literature review, 2006-2017. *VOX-Pol*. Geraadpleegd via: <http://icsr.info/2017/05/icsr-vox-pol-paper-research-perspectives-online-radicalisation-literature-review-2006-2016/>.

Meloy, J., Hoffmann, J., Guldemann, A. & James, D. (2012) "The Role of Warning Behaviors in Threat Assessment: An Exploration and Suggested Typology." *Behavioral Sciences & the Law* 30, nr. 3: 256–79. <https://doi.org/10.1002/bsl.999>.

Menczer, F. (2021, 7 oktober). Here's exactly how social media algorithms can manipulate you. *Big Think*. Geraadpleegd via: <https://bigthink.com/the-present/social-media-algorithms-manipulate-you/>.

Meta Transparency Center (webpagina, g.d. (a)). Richtlijnen voor de Facebook-community. Geraadpleegd via: <https://transparency.fb.com/nl-nl/policies/community-standards>.

Meta Transparency Center (webpagina, g.d. (b)). Haatdragend taalgebruik. Geraadpleegd via: <https://transparency.fb.com/nl-nl/policies/community-standards/hate-speech/>.

Miller-Idriss, C. (2020). *Hate in the Homeland: The New Global Far Right*. Princeton University Press. <https://doi.org/10.2307/j.ctv10tq6km>.

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. (2022, 22 februari). Kamerbrief over contentmoderatie en vrijheid van meningsuiting op online platforms. Kamerstuk | Rijksoverheid.nl. <https://www.rijksoverheid.nl/documenten/kamerstukken/2021/09/21/kamerbrief-over-contentmoderatie-en-vrijheid-van-meningsuiting-op-online-platforms>.

Moghaddam, F. M. (2005). The staircase to terrorism: a psychological exploration. *American Psychologist*, 60, 161-170.

Mohan, N. (2022, 17 februari). Inside Responsibility: What's next on our misinfo efforts. *Blog.YouTube*. Geraadpleegd via: <https://blog.YouTube/inside-YouTube/inside-responsibility-whats-next-on-our-misinfo-efforts/>.



Mølmen, G. (2021). The Role of the Internet in the Radicalisation of Extreme-Right Lone Actors. Global Network on Extremism & Technology. Geraadpleegd via: [https://gnet--research-org.translate.goog/2021/12/14/the-role-of-the-internet-in-the-radicalisationof-extreme-right-lone-actors/?\\_x\\_tr\\_sl=en&\\_x\\_tr\\_tl=nl&\\_x\\_tr\\_hl=nl&\\_x\\_tr\\_pto=sc](https://gnet--research-org.translate.goog/2021/12/14/the-role-of-the-internet-in-the-radicalisationof-extreme-right-lone-actors/?_x_tr_sl=en&_x_tr_tl=nl&_x_tr_hl=nl&_x_tr_pto=sc).

Mølmen, G. & Ravndal, J. (2021). Mechanisms of online radicalisation: how the internet affects the radicalisation of extreme-right lone actor terrorists. Behavioral Sciences of Terrorism and Political Aggression. Geraadpleegd via <https://www.tandfonline.com/doi/citedby/10.1080/19434472.2021.1993302?scroll=top&needAccess=true>.

Momas, B. (2021). The Radical Right Is Targeting Neurodivergent Groups. Rant Media. Geraadpleegd via <https://rantt.com/are-autistic-people-now-at-a-greater-risk-of-radicalization-than-ever>.

Morosoli, S., Van Aelst, P. & van Erkel, P. (2022). To Convince, to Provoke or to Entertain? A Study on Individual Motivations behind Engaging with Conspiracy Theories Online. Convergence: The International Journal of Research into New Media Technologies, 28(4), 1030–1059. <https://doi.org/10.1177/13548565221105792>.

Moskalenko, S., González, J. F. G., Kates, N., & Morton, J. (2022a). Incel ideology, radicalization and mental health: A survey study. The Journal of Intelligence, Conflict, and Warfare, 4(3), 1-29.

Moskalenko, S., Kates, N., González, J. F. G., & Bloom, M. (2022b). Predictors of Radical Intentions among Incels: A Survey of 54 Self-identified Incels. Journal of Online Trust and Safety, 1(3).

Mudde, C. (2000). The Ideology of the Extreme Right. Manchester University Press: Manchester.

Mudde, C. (2019). The far right today. Polity Press.

Mulhall, J. & Khan-Ruf, S. (eds.) (2021). State of hate: Far-right extremism in Europe 2021. London: HOPE not hate Charitable Trust.

Munger, K. & Phillips, J. (2020). "Right-Wing YouTube: A Supply and Demand Perspective". The International Journal of Press/Politics, 27, 186-219.

Munn, L. (2019). "Alt-right pipeline: Individual journeys to extremism online. First Monday. Geraadpleegd via: <https://firstmonday.org/ojs/index.php/fm/article/view/10108/7920>.

Nakamura, L. (2013). "Glitch racism: Networks as actors within vernacular Internet theory," Culture Digitally. Geraadpleegd via: <http://culturedigitally.org/2013/12/glitch-racismnetworks-as-actors-within-vernacular-internet-theory/>.

Napoli, P. M. (2014). Automated media: An institutional theory perspective on algorithmic media production and consumption. Communication Theory, 24(3), 340–360. <https://doi.org/10.1111/comt.12039>.

Napoli, P. M. (2015). Social media and the public interest: Governance of news platforms in the realm of individual and algorithmic gatekeepers. Telecommunications Policy, 39(9), 751–760. <https://doi.org/10.1016/j.telpol.2014.12.003>.

Nationaal Coördinator Terrorismebestrijding en Veiligheid (2021a). Definities gebruikt in het Dreigingsbeeld Terrorisme Nederland. Geraadpleegd via: <https://www.nctv.nl/onderwerpen/dtn/definities-gebruikt-in-het-dtn>.

Neo, L. S. (2016). An Internet-Mediated Pathway for Online Radicalisation: RECRO. In M. Kader, L. S. Neo, G. Ong, E. T. Mingyi, & J. Chin. pp. 197–224.

Neumann, P. (2013). Options and Strategies for Countering Online Radicalization in the United States. *Studies in Conflict & Terrorism*, 36(6), 431-459.

Nickolson, L., Van Bergen, N., Feddes, A., Mann, L., & Doosje, B. (2021). Extremistisch Denken en Doen. Een systematische studie van empirische bevindingen over het radicaliseringsproces. WODC. Geraadpleegd via <https://repository.wodc.nl/bitstream/handle/20.500.12832/3063/3050-extremistisch-denken-en-doen-volledige%20tekst.pdf?sequence=1&isAllowed=y>.

Niemi, R. G., & Hepburn, M. A. (2010). The rebirth of political socialization. *Perspectives on Political Science*, 24(1), 7–16. <https://doi.org/10.1080/10457097.1995.9941860>.

Nienierza, A., Reinemann, C., Fawzi, N., Riesmeyer, C., & Neumann, K. (2019). Too dark to see? Explaining adolescents' contact with online extremism and their ability to recognize it. *Information, Communication & Society*, 1–18. <https://doi.org/10.1080/1369118x.2019.1697339>.

Nieuwsuur (2021). 'Radicale rechts-extremistische jongeren zijn te redden, net als ik'. NOS Nieuws. Geraadpleegd via: <https://nos.nl/nieuwsuur/artikel/2362831-radicale-recht-extremistische-jongeren-zijn-te-redden-net-als-ik>.

NOS (2022a). Elon Musk nu eigenaar van Twitter, onduidelijk hoe het nu verder gaat. NOS Nieuws. Geraadpleegd via: <https://nos.nl/artikel/2450041-elon-musk-nu-eigenaar-van-twitter-onduidelijk-hoe-het-nu-verdergaat>.

NOS (2022b). Elon Musk kondigt vertrek als topman Twitter aan. NOS Nieuws. Geraadpleegd via: <https://nos.nl/artikel/2457198-elon-musk-kondigt-vertrek-als-topman-twitter-aan>.

NOS (2022c). Omstreden Andrew Tate 'vaderfiguur' voor sommige tieners. NOS Nieuws. Geraadpleegd via: <https://nos.nl/artikel/2442815-omstreden-andrew-tate-vaderfiguur-voor-sommige-tieners>.

NOS (2023, 21 maart). Kabinet verbiedt TikTok op werktelefoons van rijksambtenaren. NOS Nieuws. Geraadpleegd via: <https://nos.nl/artikel/2468306-kabinet-verbiedt-tiktok-op-werktelefoons-van-rijksambtenaren>.

NTA (2018). De nieuwe generatie extreemrechts en haar online schaduw. Geraadpleegd via: <https://ntadvies.nl/media/1009/de-nieuwe-generatie-extreemrechts-en-haar-online-schaduw.pdf>.

Obaidi, M. (2020). What are the psychological characteristics of people holding far-right beliefs? University of Oslo Center for Research on Extremism (C-REX). Geraadpleegd via [https://www.sv.uio.no/c-rex/english/groups/compendium/what-are-the-psychological-characteristics-of-people-holding-far-right-beliefs.html#\\_ftn28](https://www.sv.uio.no/c-rex/english/groups/compendium/what-are-the-psychological-characteristics-of-people-holding-far-right-beliefs.html#_ftn28).

Oboler, A. (2008). "Online antisemitism 2.0: 'Social antisemitism' on the 'social Web'." Jerusalem Center for Public Affairs, number 67. Geraadpleegd via: <http://jcpa.org/article/onlineantisemitism-2-0-social-antisemitism-on-the-social-web/>.

Oboler, A. (2014). "The antisemitic meme of the Jew," Online Hate Prevention Institute. Geraadpleegd via: <http://ohpi.org.au/the-antisemitic-meme-of-the-jew/>.

O'Callaghan, D. (2015). Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems. *Social Science Computer Review*, 33(4), 459–478. <https://doi.org/10.1177/0894439314555329>.

O'Callaghan, D. J., Greene, D., Macdonald, S., Carthy, J., & Cunningham, P. (2013). An Analysis of Interactions within and between Extreme Right Communities in Social Media. *Lecture Notes in Computer Science*, 88–107. [https://doi.org/10.1007/978-3-642-45392-2\\_5](https://doi.org/10.1007/978-3-642-45392-2_5).

Odag, Ö., Leiser, A., & Boehnke, K. (2019). Reviewing the role of the Internet in radicalization processes. *Journal for deradicalization*, (21), 261-300.

Ofcom (2019). Use of AI in online content moderation. Cambridge Consultants. Geraadpleegd via: [https://www.ofcom.org.uk/\\_\\_data/assets/pdf\\_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf](https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf).

Ondrak, J. (2020). Do You Know Where Your Children's Data Is? Reverse-engineering a TikTok Moral Panic. Logically. Geraadpleegd via: <https://www.logically.ai/articles/reverse-engineering-tiktok-moral-panic>.

Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. Penguin Press.

Pauwels, A. (2019). Contemporary manifestations of violent right-wing extremism in the EU: An overview of P/CVE practices. Radicalisation Awareness Network. Geraadpleegd via [https://home-affairs.ec.europa.eu/system/files/2021-04/ran\\_adhoc\\_cont\\_manif\\_vrwe\\_eu\\_overv\\_pcve\\_pract\\_2021\\_en.pdf](https://home-affairs.ec.europa.eu/system/files/2021-04/ran_adhoc_cont_manif_vrwe_eu_overv_pcve_pract_2021_en.pdf).

Pauwels, L., & De Waele, M. (2014). Youth involvement in politically motivated violence: Why do social integration, perceived legitimacy, and perceived discrimination matter? *International Journal of Conflict and Violence*, 8, 134-153.

Pauwels, L. J. R., & Svensson, R. (2017). How robust is the moderating effect of extremist beliefs on the relationship between self-control and violent extremism? *Crime and Delinquency*, 63, 1000-1016.

Peeters, T., Van Wonderen, R., Burggraaff, D., & De Wit, N. (2022). Online extreemrechtse radicalisering. Handvatten voor een preventieve aanpak. Verwey-Jonker Instituut: Utrecht.

Phelan, A., White, J., Paterson, J., & Wallner, C. (2023). Misogyny And Masculinity: Toward A Typology Of Gendered Narratives Amongst The Far-Right. CREST Research. Geraadpleegd via: <https://crestresearch.ac.uk/comment/misogyny-and-masculinity-toward-a-typology-of-gendered-narratives-amongst-the-far-right/>.

Phillips, W. (2018). *The Oxygen of Amplification: Better Practices for Reporting on Extremists, Antagonists, and Manipulators Online*. New York: Data and Society Research Institute.

Pisoiu, D. (2015) "Subcultural Theory Applied to Jihadi and Right-Wing Radicalization in Germany." *Terrorism and Political Violence* 27, no. 1: 9–28. <https://doi.org/10.1080/09546553.2014.959406>.

Platform JEP (g.d.). Hoe een extreemrechtse jongen deradicaliseerde. Geraadpleegd via: <https://www.platformjep.nl/ervaringsverhalen/michel-letteboer>.

Platform JEP (2019). Handreiking Licht Verstandelijke beperking (LVB) en radicalisering. Geraadpleegd via: <https://www.platformjep.nl/binaries/platformjep/documenten/publicaties/2019/12/13/licht-verstandelijke-beperking-en-radicalisering/LVB+en+radicalisering.pdf>.

Politie's Sikkerhetstjeneste. (2019). Theme report: What is the background of right-wing extremists in Norway? Politie's Sikkerhetstjeneste. Geraadpleegd via [https://www.pst.no/globalassets/artikler/utgivelser/theme-report\\_-what-is-the-background-of-rightwing-extremists-in-norway.pdf](https://www.pst.no/globalassets/artikler/utgivelser/theme-report_-what-is-the-background-of-rightwing-extremists-in-norway.pdf).

Pousaz, L. & Roten, C. (2022). A Short linguistic meta-analysis of Qanon authorship: confirming Ron Watkins as the most likely author. Orphan Analytics. Geraadpleegd via: [https://www.orphananalytics.com/en/news/white-paper202201/OA\\_QAnon\\_meta-analysis\\_2022-02.pdf](https://www.orphananalytics.com/en/news/white-paper202201/OA_QAnon_meta-analysis_2022-02.pdf).

Precht, T. (2007). Home grown terrorism and Islamist radicalisation in Europe. From conversion to terrorism. Danish Ministry of Justice.

Raad van de Europese Unie (2020). The Role of Algorithmic Amplification in Promoting Violent and Extremist Content and its Dissemination on Platforms and Social Media. Geraadpleegd via: <https://data.consilium.europa.eu/doc/document/ST-12735-2020-INIT/en/pdf>.

Rauf, A. (2020). "New Moralities for New Media? Assessing the Role of Social Media in Acts of Terror and Providing Points of Deliberation for Business Ethics". *Journal of Business Ethics*, 170, 229-251. <https://doi.org/10.1007/s10551-020-04635-w>.

Reed, A. et al. (2019) "Radical Filter Bubbles: Social Media Personalization Algorithms and Extremist Content." *Global Research Network on Terrorism and Technology*, vol. 8.

Ribeiro, M. H. (2019). Auditing Radicalization Pathways on YouTube. *ACM Symposium on Neural Gaze Detection*. <http://arxiv.org/abs/1908.08313>.

Rich, B. & Buljaka, E. (2023). The Draw of the 'manosphere': understanding Andrew Tate's appeal to lost men. *The Conversation*. Geraadpleegd via: <https://theconversation.com/the-draw-of-the-manosphere-understanding-andrew-tates-appeal-to-lost-men-199179>.

Rijksoverheid (2021). EU-ministers akkoord met regelgeving voor digitale diensten en markten. Geraadpleegd via: <https://www.rijksoverheid.nl/actueel/nieuws/2021/11/24/eu-ministers-akkoord-met-regelgeving-digitale-diensten-en--markten>.

R-Newt (2019). Safe Spaces om Radicalisering te voorkomen. R-Newt Jongeren. Geraadpleegd via: <https://r-newt.nl/algemeen/safe-spaces-om-radicalisering-te-voorkomen/>.

Rogers, C. (2020). "Islamic and Domestic Radicalization: Does One Size Fit Most?" *Naval Postgraduate School*.

Roks, R., & Schoot, J. van der. (2019). Het aanpassingsdilemma online: een verkennend onderzoek naar extreemrechts op social media. *Tijdschrift voor Criminologie*, (61)3.

Romano, A. (2019). How the Christchurch shooter used memes to spread hate. *Vox*. Geraadpleegd via: from <https://www.vox.com/culture/2019/3/16/18266930/christchurch-shooter-manifesto-memes-subscribe-to-pewdiepie>.

Roose, K. (2019) The Making of a YouTube Radical. *The New York Times*. Geraadpleegd via: <https://www.nytimes.com/interactive/2019/06/08/technology/YouTube-radical.html>.

Rottweiler, B., & Gill, P. (2022). Individual Differences in Personality Moderate the Effects of Perceived Group Deprivation on Violent Extremism: Evidence From a United Kingdom Nationally Representative Survey. *Frontier Psychology*. <https://doi.org/10.3389/fpsyg.2022.790770>.

Rowa, J.Y. (2020). Part 1: Algorithmic Deconstruction in the Context of Online Extremism. *Global Network on Extremism & Technology (GNET)*. Geraadpleegd via: <https://gnet-research.org/2020/09/15/part-1-algorithmic-deconstruction-in-the-context-of-online-extremism/>.

Rowa, J.Y. (2022). The Contextuality of Lone Wolf Algorithms: An Examination of (Non)Violent Extremism in the Cyber-Physical Space. *Global Internet Forum to Counter Terrorism*. Geraadpleegd via: <https://gifct.org/wp-content/uploads/2022/09/GIFCT-22WG-ContextualityIntros-1.1.pdf>.

Saarijärvi, J., & Siafaka, A. (2021). The role of women in far-right extremism. The Counterterrorism Group. Geraadpleegd via <https://www.counterterrorismgroup.com/post/the-role-of-women-in-far-right-extremism>.

Sageman, M. (2004). Understanding terror networks. Philadelphia: University of Pennsylvania Press.

Save the Children (2021). Youth as a target for extremist recruitment. Geraadpleegd via: [https://pelastakaalapset.s3.eu-west-1.amazonaws.com/main/2022/02/17133213/youth-as-a-target-for-extremist-recruitment\\_stc\\_en.pdf](https://pelastakaalapset.s3.eu-west-1.amazonaws.com/main/2022/02/17133213/youth-as-a-target-for-extremist-recruitment_stc_en.pdf).

Schleicher, F. (2018). Waarom Jordan Peterson gevaarlijker is dan hij lijkt. BNNVARA. Geraadpleegd via: <https://www.bnnvara.nl/joop/artikelen/waarom-jordan-peterson-gevaarlijker-is-dan-hij-lijkt>.

Schmid, A. P. (2013). Radicalisation, de-radicalisation, counter-radicalisation: A conceptual discussion and literature review. ICCT Research Paper, 97.

Schneider, J. (2023). TikTok Can Now Detect Borderline Content To Help Keep Young Eyes Safe. PetaPixel. Geraadpleegd via: <https://petapixel.com/2023/01/03/tiktok-can-now-detect-borderline-content-to-help-keep-young-eyes-safe/#:~:text=Newsletters-,TikTok%20Can%20Now%20Detect%20'Borderline'%20Content%20to,Help%20Keep%20Young%20Eyes%20Safe&text=TikTok%20says%20that%20it%20is,been%20more%20challenging%20to%20identify>.

Schröder, C., Bruns, J., Lehmann, L., Goede, L., Bliesener, T., & Tomczyk, S., (2022). Radicalisation in Adolescence: the Identification of Vulnerable Groups. European Journal on Criminal Policy and Research 28: 177-201. <https://doi.org/10.1007/s10610-022-09505-x>.

Schumpe, B. M., Bélanger, J. J., Moyano, M., & Nisa, C. F. (2020). The role of sensation seeking in political violence: An extension of the Significance Quest Theory. Journal of personality and social psychology, 118(4), 743–761. <https://doi.org/10.1037/pspp0000223>.

Seaver, N. (2018). Captivating algorithms: Recommender systems as traps. Journal of Material Culture. <https://doi.org/10.1177/1359183518820366>.

Sieckelinck, S. & Gielen, A. (2020). Veerkracht bevorderen bij opgroeiende jongeren: Inzetten op beschermende factoren. Platform JEP, Expertise-unit Sociale Stabiliteit.

Sieckelinck, S., Sikkens, E., San, M. V., Kotnis, S., & De Winter, M. (2019). Transitional journeys into and out of extremism. A biographical approach. Studies in Conflict and Terrorism, 42(7), 662-682. <https://doi.org/10.1080/1057610X.2017.1407075>.

Sikkens, E., Van San, M., De Winter, M., & Sieckelinck, S. (2017). Verschillende extreme idealen, dezelfde radicale zoektocht? Push- en pullfactoren volgens jongeren met extreemlinks, extreemrechtse en extreem Islamitische idealen. PED 37(2): 157-171. Amsterdam: Amsterdam University Press.

Silber, M. D., Bhatt, A., & Analysts, S. I. (2007). Radicalization in the west: The homegrown threat. New York: Police Department.

Simi, P., Windisch, S., & Sporer, K (2016). Recruitment and Radicalization among US Far Right Terrorists." College Park, MD: START, 2016. Geraadpleegd via [https://www.start.umd.edu/pubs/START\\_RecruitmentRadicalizationAmongUSFarRightTerrorists\\_Nov2016.pdf](https://www.start.umd.edu/pubs/START_RecruitmentRadicalizationAmongUSFarRightTerrorists_Nov2016.pdf).

Sitler-Elbel, F. (2021). "From Swiffers to Swastikas: How the #tradwife movement of conventional gender roles became synonymous with white supremacy". Senior Projects Spring 2021, 131. [https://digitalcommons.bard.edu/senproj\\_s2021/131](https://digitalcommons.bard.edu/senproj_s2021/131).

Smith, D. (2011). *Less than human: why we demean, enslave, and exterminate others*. New York: St. Martin's Press.

Southern Poverty Law Center (g.d.). Asatru Folk Assembly. Geraadpleegd via: <https://www.splcenter.org/fighting-hate/extremist-files/group/asatru-folk-assembly>.

Spaaij, R. (2012). *Understanding Lone Wolf Terrorism: Global Patterns, Motivations and Prevention*. New York: Springer.

Spee, I., Reitsma, M. (2015). Puberaal, lastig of radicaliserend. Geraadpleegd via <http://www.schoolenveiligheid.nl/po-vo/kennisbank/puberaal-lastig-of-radicaliserend/>.

Steizinger, J. (2018). "The significance of dehumanization: Nazi ideology and its psychological consequences," *Politics, Religion & Ideology*, volume 19, nr. 2, pp. 139–157. doi: <https://doi.org/10.1080/21567689.2018.1425144>.

Stephens, W. Sieckelinck, S. & Boutellier, H. (2021) Preventing Violent Extremism: A Review of the Literature, *Studies in Conflict & Terrorism*, 44:4, 346-361.

Sterkenburg, N. (2019). *Far-Right Extremism. A Practical Introduction*. Radicalisation Awareness Network Factbook. Geraadpleegd via [https://home-affairs.ec.europa.eu/system/files/2019-12/ran\\_fre\\_factbook\\_20191205\\_en.pdf](https://home-affairs.ec.europa.eu/system/files/2019-12/ran_fre_factbook_20191205_en.pdf).

Sterkenburg, N. (2021). Maar dat mag je niet zeggen: De nieuwe generatie radicaal- en extreemrechts in Nederland. *Das Mag*.

Stinson, C. (2022). Algorithms are not neutral. *AI Ethics*, vol. 2, p. 763-770. <https://doi.org/10.1007/s43681-022-00136-w>.

TikTok (g.d.). Our approach to content moderation. Geraadpleegd via (januari, 2023): <https://www.tiktok.com/transparency/en-us/content-moderation/>.

TikTok (2022, 11 maart). Hoe Tik Tok video's aanbeveelt #VoorJou. Geraadpleegd via: <https://newsroom.tiktok.com/nl-nl/hoe-tiktok-videos-aanbeveelt-voorjou>.

TikTok (2023, maart). Communityrichtlijnen. Geraadpleegd via: <https://www.tiktok.com/community-guidelines>.

Torney-Purta, J. (2004). Adolescents' political socialization in changing contexts: An international study in the spirit of Nevitt Sanford. *Political Psychology*, 25(3), 465–478. <https://doi.org/10.1111/j.1467-9221.2004.00380.x>.

Jones, C., Trott, V., & Wright, S. (2019). "Sluts and soyboys: MGTOW and the production of misogynistic online harassment." *New Media & Society*, 00(0), p. 1-19. DOI: 10.1177/1461444819887141.

Twitter Help Center (2022, 10 februari). *Hateful Conduct*. Geraadpleegd via: <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>.

Valentini, D., Lorusso, A. & Stephan, A. (2020). Onlife Extremism: Dynamic Integration of Digital and Physical Spaces in Radicalization. *Front. Psychology*, 11(524). Geraadpleegd via: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00524/full#B38>.

Valk, I. van der, & Wagenaar, W. (2010). *Monitor Racisme & Extremisme: in en uit extreemrechts*. Amsterdam: Anne Frank Stichting / Amsterdam University.

Van den Berg, G. (2022). *Wat kun je doen aan polarisatie en radicalisering? Begrippen, modellen, factoren en interventies*. Nederlands Jeugdinstituut. Geraadpleegd via <https://www.nji.nl/sites/default/files/2022-02/Wat-kun-je-doen-aan-polarisatie-en-radicalisering.pdf>.

Van der Sanden, E. (2020, 12 oktober). *QAnon in Nederland: een paraplu-terme waaronder steeds meer complotdenkers zich verzamelen*. RTL Nieuws. Geraadpleegd via: <https://www.rtlnieuws.nl/nieuws/nederland/artikel/5189782/qanon-aanhangers-complotdenken-theorie-nederland>.

Van Hiel, A., Onraet, E., Crowson, H.M., & Roets, A. (2016). The relationship between rightwing attitudes and cognitive style: A comparison of self-report and behavioural measures of rigidity and intolerance of ambiguity. *European Journal of Personality*, 30(6), 523-531.

Van Prooijen, J.-W., Ligthart, J., Rosema, S. and Xu, Y. (2022), The entertainment value of conspiracy theories. *Br. J. Psychol.*, 113: 25-48. <https://doi.org/10.1111/bjop.12522>.

Van Puffelen, M. (2021). *Rechtsextremisme: Geweld met een rechtsextremistisch motief*. Amsterdam: DSP.

Venhaus, J.M. (2010). *Why youth join al-Qaeda*. Washington: United States Institute of Peace.

View, T. (2019). *I Was Looking At The "Likes" On The YouTube Page Of Buckey Wolfe, The Qanon Follower Who Is Accused Of Killing His Brother. Wolfe Also Told [...]*. Thread Reader App. <https://threadreaderapp.com/thread/1083>.

Von Behr, I., Reding, A., Edwards, C. & Gribbon, L. (2013). *Radicalisation in the digital era: the use of the internet in 15 cases of terrorism and extremism*. RAND Europe. Geraadpleegd via: [https://www.rand.org/content/dam/rand/pubs/research\\_reports/RR400/RR453/RAND\\_RR453.pdf](https://www.rand.org/content/dam/rand/pubs/research_reports/RR400/RR453/RAND_RR453.pdf).

VPRO Tegenlicht (2020). *Alfamannen*. Geraadpleegd via: <https://www.vpro.nl/programmas/tegenlicht/kijk/afleveringen/2019-2020/alfamannen.html>.

Wagenaar, W. (2021). *Extreemrechts in Nederlandse Gemeentes*. Factsheet 13, Anne Frank Stichting.

Wall Street Journal Staff (2021). *Inside TikTok's Algorithm: A WSJ Video Investigation*. The Wall Street Journal. Geraadpleegd via: <https://www.wsj.com/articles/tiktok-algorithm-video-investigation-11626877477>.

Ward, C. & Voas, D. (2011). *The Emerge of Conspirituality*. *Journal of Contemporary Religion*, vol. 26, nr. 1, p. 103-121. Geraadpleegd via: <https://www.unifr.ch/webnews/content/108/file/The%20Emergence%20of%20Conspirituality.pdf>.

Watts, M. W. (1999). Are there typical age curves in political behavior? The "age invariance" hypothesis and political socialization. *Political Psychology*, 20(3), 477-499. <https://doi.org/10.1111/0162-895X.00153>.

Weenink, A. W. (2015). Behavioral problems and disorders among radicals in police files. *Perspectives on Terrorism*, 9, 17-33.

Welzen, A. van, Caarels, M., & Potters, P. (2017). *Vanzelf Mediawijs? 2017. The Choice marktonderzoek & advies*.

Whittaker, J. (2022). Recommendation Algorithms and Extremist Content: A Review of Empirical Evidence. GIFCT Transparency Working Group. Geraadpleegd via: <https://gifct.org/wp-content/uploads/2022/07/GIFCT-22WG-TR-Empirical-1.1.pdf>.

Whittaker, J. (2022). Rethinking Online Radicalization. *Perspectives on Terrorism*, 16(4), 27–40. <https://www.jstor.org/stable/27158150>.

Whittaker, J. & Herath, C. (2021). Online Radicalisation: Moving Beyond a Simple Dichotomy. *Global Network on Extremism & Technology*. Geraadpleegd via: <https://gnet-research.org/2021/12/20/online-radicalisation-moving-beyond-a-simple-dichotomy/>.

Whittaker, J., Looney, S., Reed, A., & Votta, F. (2021). "Recommender systems and the amplification of extremist content." *Internet Policy Review*, 10(2). <https://doi.org/10.14763/2021.2.1565>.

Wiktorowicz, Q. (2004). Joining the cause: Al-Muhajiroun and radical Islam. Paper presented at the Roots of Islamic Radicalism conference, Yale University, Cambridge, MA, 8 – 9 May 2004.

Wiktorowicz, Q. (2005). *Radical Islam rising: Muslim extremism in the West*. Oxford, United Kingdom: Rowman & Littlefield.

Wilson, J. (2017, mei 23). Hiding in plain sight: how the 'alt-right' is weaponizing irony to spread fascism. *The Guardian*. Geraadpleegd via: <https://www.theguardian.com/technology/2017/may/23/alt-right-online-humor-as-a-weapon-facism>.

Winter, C., Neumann, P., Meleagrou-Hitchens, A., Ranstorp, M., Vidino, L. & Fürst, J. (2020). Online Extremism: Research Trends in Internet Activism, Radicalization, and Counter-Strategies. *International Journal of Conflict and Violence*, vol. 14(2), 1-20. Geraadpleegd via: <https://www.ijcv.org/index.php/ijcv/article/view/3809/3868>.

Wolfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasasi, B. (2019). A Field-Wide Systematic Review and Meta-analysis of Putative Risk and Protective Factors for Radicalization Outcomes. *Journal of Quantitative Criminology*. Geraadpleegd via: <https://prohic.nl/wp-content/uploads/2020/11/2020-08-10-Radicalisation-RiskProtectiveFactorsMeta.2019.pdf>.

Woodbury-Smith, M. R., Loftin, R., Westphal, A., & Volkmar, F. R. (2022). Vulnerability to Ideologically-Motivated Violence Among Individuals With Autism Spectrum Disorder. *Frontiers in psychiatry*, 13, 873121.

Yesilada, M. & Lewandowsky, S. (2022). Systematic review: YouTube recommendations and problematic content. *Internet Policy Review*, 11(1). <https://doi.org/10.14763/2022.1.1652>.

Yonder (2021). *A Brand Guide for Fringe Social Networks*. Yonder by Primer. Geraadpleegd via (9 januari 2023): <https://www.yonder-ai.com/resources/blog/brand-guide-fringe-social-networks/>.

Young, H. (2022). Extremists use video games to recruit vulnerable youth. Here's what parents and gamers need to know. *The Conversation*. Geraadpleegd via: <https://theconversation.com/extremists-use-video-games-to-recruit-vulnerable-youth-heres-what-parents-and-gamers-need-to-know-193110>.

YouTube (webpagina, g.d.). Hoe beschermt YouTube de community tegen haat en intimidatie? Geraadpleegd via: <https://www.YouTube.com/howYouTubeworks/our-commitments/standing-up-to-hate/>.

YouTube Help (g.d. (1)). Communityrichtlijnen van YouTube. Geraadpleegd via (januari, 2023): [https://support.google.com/YouTube/answer/9288567?hl=nl&ref\\_topic=6151248](https://support.google.com/YouTube/answer/9288567?hl=nl&ref_topic=6151248).



YouTube Help (g.d. (2)). Beleid voor externe links. Geraadpleegd via (januari, 2023): <https://support.google.com/YouTube/answer/9054257>.

YouTube Help (g.d. (3)). Beleid tegen aanzetten tot haat. Geraadpleegd via (januari, 2023): <https://support.google.com/YouTube/answer/2801939>.

Zannettou S, Kwak H, De Cristofaro E, et al. (2018) What is gab? A bastion of free speech or an alt-right echo chamber? In: WWW'18 Conference Companion, New York: ACM. Available at: <https://doi.org/10.1145/3184558.3191531>.

Zidan, K. (2022). Fight for the Right: Why So Many MMA Fighters Are Becoming Right-Wing Political Activists. Global Sport Matters. Geraadpleegd via: <https://globalsportmatters.com/opinion/2022/10/20/why-many-mma-fighters-becoming-right-wing-political-activists/>.

Zmigrod, L., Rentfrow, P.J., & Robbins, T.W. (2019). Cognitive inflexibility predicts extremist attitudes. *Frontiers in psychology*. <https://doi.org/10.3389/fpsyg.2019.00989>.

Zuiderveen Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, 5(1). <https://doi.org/10.14763/2016.1.401>.

Zwissler, L. (2018). New Age, Wicca and Paganism. *International Encyclopaedia of Anthropology*, 1-11. Geraadpleegd via: [https://www.academia.edu/download/67678668/New\\_Age\\_Wicca\\_Paganism\\_International\\_Encyclopedia\\_of\\_Anthropology.pdf](https://www.academia.edu/download/67678668/New_Age_Wicca_Paganism_International_Encyclopedia_of_Anthropology.pdf).

4chan (g.d.). Global Rules. Geraadpleegd via: <https://www.4channel.org/rules>.

## COLOFON

Opdrachtgever	Wetenschappelijk Onderzoek- en Documentatiecentrum
Auteurs	Drs. R. van Wonderen D. Burggraaff, MSc Dr. S. Ganpat Drs. G. van Beek (TextGain) O. Cauberghe (TextGain)
Met medewerking van	M. Keijzer
Foto omslag	Adobe Stock, YakobchuckOlena, 470017365
Uitgave	Verwey-Jonker Instituut Giessenplein 59 C 3522 KE Utrecht T (030) 230 07 99 E <a href="mailto:secr@verwey-jonker.nl">secr@verwey-jonker.nl</a> I <a href="http://www.verwey-jonker.nl">www.verwey-jonker.nl</a>

De publicatie kan gedownload worden via onze website:  
<http://www.verwey-jonker.nl>

ISBN 978-94-6409-250-9

© Verwey-Jonker Instituut, Utrecht, augustus 2023.

Het auteursrecht van deze publicatie berust bij het Verwey-Jonker Instituut. Gedeeltelijke overname van teksten is toegestaan, mits daarbij de bron wordt vermeld. The copyright of this publication rests with the Verwey-Jonker Institute. Partial reproduction of the text is allowed, on condition that the source is mentioned.