Instituut VERWEY JONKER

Online geweld tussen burgers

Schaduwzijde van digitalisering

Marjolein Odekerken | Esther de Weger | Maxime Yenga | Dirk Woertink

Inhoud

De wereld van **online geweld**

Marjolein Odekerken, Esther de Weger, Maxime Yenga & Dirk Woertink

Wat online geweld is

Online geweld is **wijdverspreid** en **diffuus** — het kent vele gezichten, maar herkenbare patronen. Het is hybride: online en offline lopen naadloos in elkaar over. Het is **dubbel**: anoniem én persoonlijk. En het groeit **snel**, tussen burgers onderling, met gevolgen die verder reiken dan het scherm. Online geweld **ontwricht**. Het tast welzijn aan, polariseert gemeenschappen en heeft echte gevolgen in de echte wereld.

Online geweld is écht geweld

Individuele impact
→ Stress, isolement, schaamte, zelfcensuur

Sociale impact
→ Hele groepen worden weggezet, waardoor polarisatie ontstaat

Maatschappelijke impact
→ Ondermijning van de democratie

"Wat je wel ziet is dat heel veel niet illegaal is, maar wel schadelijk. En dus niet lawful, maar harmful."

Hoe groot het is

Er is geen eenduidige definitie of volledig beeld van hoe groot online geweld is. Aangiftes, registraties en onderzoek lopen achter. Wat we wél weten: **het probleem is groot en groeiende**.

Voorbeeld – jongeren:

→ **12,9%** is één of meerdere keren pleger

→ **26,2%** is één of meerdere keren slachtoffer

→ **9,7%** kreeg kwetsende berichten over identiteit of achtergrond

*"Als je het vaak doet, wordt het **normaal**."*

Wie erbij betrokken zijn

Slachtoffers:
Vooral jongeren, vrouwen, LHBTI+personen en mensen met een migratieachtergrond.
Zij ervaren schaamte, victim blaming en het gevoel niet serieus genomen te worden.

Plegers:
Voornamelijk mannen. Vaak binnen online (sub)culturen of georganiseerde groepen. Soms uit frustratie of groepsdruk.Veel plegers normaliseren of bagatelliseren hun gedrag.

VERWEY JONKER Instituut

Waarom aanpak uitblijft

- **Versnippering**: geen duidelijke regie of verantwoordelijkheden
- **Beperkte informatie** en juridische drempels
- **Onvoldoende kennis** bij professionals
- **Schaamte en taboe** bij slachtoffers
- **Weinig moderatie** op grensoverschrijdende platforms

Kortom: het systeem is niet ingericht op online geweld.

Wat er nodig is

- **Heldere regie en samenwerking** – publiek, privaat en internationaal
- Verantwoordelijkheid van **techbedrijven**
- Investeren in kennis, onderzoek en monitoring
- Inzet van vroegtijdige interventies (vroegsignalering)
- Bewustwording en preventie
- **Erkenning: online geweld is echt geweld**

"Je zit in onwetendheid. Elke man die me aankeek dacht ik: ken jij mij van online?"

Sociale media-analyse

Bekijk de uitlegvideo

Het doel: De aard en omvang cijfermatig in beeld brengen.

Methoden: **Natural Language Processing** (NLP) om tienduizenden reacties onder Nu.nl-artikelen en posts op X te analyseren.

Resultaten:

- Meeste berichten worden 'goedgekeurd' of zijn 'acceptabel'.
- Problematische taal komt vaker voor in berichten over gemarginaliseerde groepen. Vooral discriminerende woorden worden gebruikt.

"Waarom heb je de politie erbij gehaald? Die verkrachting was maar een grapje hoor."

Het onderzoek

Deze factsheet is gebaseerd op onderzoek van het Verwey-Jonker Instituut, in opdracht van het Ministerie van Justitie en Veiligheid, en hoort bij het rapport **Online geweld tussen burgers onderling: de schaduwzijde van digitalisering**.

Download het rapport

Contact

dr. Marjolein Odekerken
modekerken@verwey-jonker.nl

VERWEY JONKER Instituut

Samenvatting

Online geweld tussen burgers is een onderbelicht vraagstuk
Online geweld tussen burgers kent veel kanten. Het raakt juridische, technologische en sociale grenzen en vraagt om een breed gedragen aanpak. Om zo'n aanpak te kunnen ontwikkelen is beter inzicht nodig in de aard en omvang, de onderliggende factoren en impact van dit type geweld. Tegelijkertijd is online geweld tussen burgers onderling een *'nieuw' vraagstuk* voor zowel onderzoek/wetenschap, beleid als de uitvoering.

In opdracht van het Ministerie van Justitie en Veiligheid voert het Verwey-Jonker Instituut onderzoek uit naar (strafbare) vormen van online geweld tussen burgers. Dit rapport presenteert de belangrijkste bevindingen over de aard en omvang van dit fenomeen, de betrokken burgers (plegers en slachtoffers) en professionals/experts en hun perspectieven en ervaringen, en mogelijke interventies en beleidsmaatregelen om dit vraagstuk aan te pakken.

Drie vormen van online geweld
Online geweld is een brede term voor misdrijven en geweld die plaatsvinden in het digitale domein—binnen de online leefwereld van mensen—en gericht zijn op burgers en publieke personen. Omdat online geweld zo breed is, is het belangrijk de **focus** te beperken. In dit onderzoek kijken we naar de volgende drie vormen van *online geweld tussen burgers onderling*, omdat het Ministerie hier haar aanpak2 op wil richten en er nog weinig onderzoek naar is:

1. Online bedreiging en intimidatie – gericht op het zaaien van angst, vaak via directe dreigementen.
2. Haatzaaien en discriminatie – het aanzetten tot haat of discriminatie tegen specifieke groepen.
3. Kleinschalige ordeverstoringen en opruiing – het oproepen tot gewelddadig of strafbaar gedrag, vaak in groepsverband.

Probleemstelling
Het onderzoeksdoel is om meer inzicht te krijgen in de verschillende uitingsvormen van online geweld, de onderliggende maatschappelijke context en de profielen van plegers en slachtoffers. Daarnaast gaan we in op de impact van online geweld op slachtoffers en de bredere samenleving, evenals de behoeften van slachtoffers aan ondersteuning.

De deelvragen die centraal staan, zijn:

1. Welke vormen van online geweld zijn het meest voorkomend in Nederland?
2. Wie zijn de plegers van deze vormen van geweld, wat zijn hun motieven en kenmerken?
3. Wie zijn de slachtoffers van deze vormen van geweld, welke impact heeft het online geweld, welke behoeften hebben zij en wat zijn hun kenmerken?
4. Is er sprake van overlap tussen slachtofferschap en plegerschap?
5. Wat zijn de modus operandi van plegers?
6. Wat zijn de weerbaarheidsstrategieën/copingmechanismen van slachtoffers?
7. Hoe ziet de interactie tussen plegers en slachtoffers eruit?
8. Hoe hangt online geweld samen met fysiek geweld, welke stadia kunnen er worden onderscheiden? Hoe draagt de online wereld/dynamiek bij aan offline geweld?
9. Welke preventieve en repressieve interventies kunnen mogelijk ontwikkeld en geïmplementeerd worden om (a) plegers te stoppen en (b) slachtoffers te beschermen en ondersteunen? En welke belemmeringen zijn er in het huidige beleid om online geweld aan te kunnen pakken?

Praktijkgericht onderzoek
Dit onderzoek richt zich op online geweld tussen burgers, een relatief nieuw aandachtsgebied. In dit onderzoek hebben we ons voornamelijk laten voeden door de verhalen en ervaringen uit de praktijk. Het onderzoek is praktijkgericht waarbij we de perspectieven van burgers en professionals/experts centraal stellen. Er is gebruik gemaakt van een multimethod design waarin we de volgende kwalitatieve- en kwantitatieve onderzoeksmethoden integreren:

1. Literatuurstudie naar wat al bekend is over online geweld tussen burgers onderling.
2. Kwantitatieve analyse van zelfgerapporteerd pleger- en slachtofferschap van online geweld onder jongeren.
3. Social media analyse waarin meer dan tienduizenden reacties onder nieuwsberichten van Nu.nl (NU - Het laatste nieuws het eerst op NU.nl) en posts (voorheen tweets) op X zijn geanalyseerd.
4. Interviews praktijk- en wetenschappelijke experts.
5. Interviews plegers en slachtoffers - casusonderzoek.
6. Twee focusgroepen met professionals/experts onder andere werkzaam in de strafketen en slachtofferhulp.
7. Online reflectiesessie voor disseminatie van kennis en het reflecteren met elkaar over de bevindingen en conclusies.
8. Expertsessie in samenwerking met het Ministerie van Justitie & Veiligheid met stakeholders – o.a. met politie, OM, social media platform, gemeenten en slachtofferhulp organisaties.

Belangrijkste bevindingen

Online geweld is diffuus maar heeft herkenbare patronen

Online geweld kent vele verschijningsvormen en is moeilijk af te bakenen, maar laat wel duidelijke patronen zien. Slachtoffers worden vaak geraakt op basis van identiteit of publieke positie, en maatschappelijke contexten (zoals coronacrisis of actuele debatten) versterken dit. Het geweld richt zich doelbewust op specifieke groepen en wordt beïnvloed door factoren als anonimiteit, groepsdruk en de dynamiek van online gemeenschappen.

Online geweld is wijdverspreid

De platforms waarop online geweld zich manifesteert, opereren veelal grensoverschrijdend en de verspreiding van problematische content stopt niet bij landsgrenzen. Iets wat online plaatsvindt, ook als het verwijderd wordt van een platform, kan jarenlang iemand blijven achtervolgen (bijvoorbeeld als het al gedownload en gedeeld is voordat het eraf wordt gehaald).

Online geweld is een continuüm

Dat betekent dat het niet alleen bestaat uit extreme uitingen zoals doodsbedreigingen maar ook subtiele, alledaagse vormen kan aannemen. In dit onderzoek is een schaal ontwikkeld van online gedrag, variërend van acceptabel tot strafbaar, om het fenomeen online geweld beter te begrijpen.

Online geweld heeft een hybride karakter

De werelden van online en offline geweld zijn niet meer te onderscheiden waardoor we beter over hybride vormen van geweld kunnen spreken. In verschillende casussen blijken meerdere vormen van online en offline geweld door elkaar heen plaats te vinden.

Online geweld laat zich kenmerken door een escalatieladder en als katalysator van ‘nieuw’ geweld

De escalatieladder is een model dat geweld beschrijft als een proces dat stapsgewijs toeneemt in intensiteit. Er is sprake van een hybride geweldspiraal waarin verschillende vormen en fases elkaar in ‘no tempo’ kunnen opvolgen. Iets wat eerst onaardig is, kan razendsnel uitmonden in een gewelddadige geweldssituatie. Tegelijkertijd kan iets wat online begint, uitmonden in fysiek geweld of omgekeerd.

Online geweld is dubbel

De dubbelheid – persoonlijk én anoniem, offline én online – maakt online geweld moeilijker te bestrijden. Bovendien worden de verschillende vormen van online geweld niet alleen ingezet als uitlaatklep, maar ook strategisch gebruikt om iemand persoonlijk te raken, individuen of groepen uit te sluiten en/of uit de samenleving en het publieke debat te weren.

Online geweld is moeilijk in beeld te brengen: er ontbreekt een eenduidig beeld

Momenteel ontbreekt het aan een gedeeld, gedragen en voldoende uitgewerkt beeld om online geweld als complex en multidimensioneel fenomeen goed te begrijpen en effectief aan te pakken.

Online geweld komt veelal laat in beeld

Wanneer incidenten van online geweld bij partijen in de uitvoering in beeld komen (bijvoorbeeld bij politie, gemeenten of scholen) is de situatie veelal geëscaleerd. Juist zicht op het voorstadium waarin de eerste spanningen plaatsvinden, blijft nog achterwege.

Online geweld is een groeiend maatschappelijk probleem dat urgentie vereist

Online geweld is een groeiend maatschappelijk probleem dat urgentie vereist. Experts, professionals maar ook burgers onderkennen dat we in een tijdperk terecht zijn gekomen waarin online geweld ver is doorgedrongen in onze leefwereld en *de schaduwzijde ervan niet meer is weg te denken.*

Online geweld heeft enorme impact op (het welzijn van) individuen maar ook op gehele groepen (met specifieke kenmerken). Het werkt ontwrichtend en zet aan tot polarisatie

Online geweld heeft niet alleen directe gevolgen voor individuele slachtoffers, maar tast ook de fundamenten van de samenleving aan. Het ondermijnt de sociale cohesie, belemmert het vrije publieke debat en kan leiden tot uitsluiting, polarisatie en uiteindelijk tot verzwakking van de democratische rechtsstaat.

Online geweld kan iedereen overkomen maar raakt ook juist kwetsbare groepen

Specifieke kenmerken maken sommige individuen en groepen kwetsbaar voor online geweld. Te denken valt aan vrouwen, jongeren (met een licht verstandelijke beperking), LHBTQIA+ personen, mensen met een migratieachtergrond, activisten etc. Daarbij kan online geweld ook iedereen treffen.

Online geweld wordt (nog) niet gezien als ‘echt’ geweld

Zowel aan de kant van plegers als slachtoffers is te zien dat ze strategieën en copingmechanismen inzetten waarbij online geweldsincidenten worden gerelativeerd, gebagatelliseerd, genegeerd of gerationaliseerd.

Online geweld kenmerkt zich nog door het gebrek aan adequate ondersteuning, serieuze opvolging van meldingen/aangiften en effectieve opsporing

Er is bij online geweld vaak sprake van schaamte of andere emoties die ervoor zorgen dat er nog niet openlijk over wordt gesproken of dat er niet wordt aangeklopt voor hulp.

Online geweld is nog geen gedeeld en gedragen vraagstuk

Aan de kant van de wetenschap, beleid en de uitvoering (waaronder politie en gemeenten) zijn nog verschillende drempels die een effectieve aanpak in de weg staan: zoals een gebrek aan kennis, gebrek aan eigenaarschap, signalen die blijven liggen, gebrek aan capaciteit, onwetendheid over wat wel mag en kan en het gebrek aan een domeinoverstijgende en structurele (inter)nationale samenwerking.

Online geweld is nog “nieuw”

Het complexe vraagstuk van online geweld is nog nieuw en men is zoekende naar de juiste inzichten, antwoorden én oplossingen. Er is gebrek aan onderzoek naar de wereld achter online geweld, welke interventies kansrijk én effectief zijn (wat zijn de werkzame elementen, geleerde lessen en randvoorwaarden) en (nieuwe of aangepaste) wet- en regelgeving.

Online geweld is toe aan een inhaalslag

Als we kijken naar hoe complex het vraagstuk van online geweld is, naar de impact ervan, het gebrek aan specifieke wetgeving en het ontbreken van het nemen van verantwoordelijkheid van allerlei verschillende actoren waaronder burgers zelf, professionals en experts, dan liggen er nog genoeg kansen voor verbetering om dit met elkaar te doen. We staan aan het begin van een noodzakelijke stevige maatschappelijke en juridische inhaalslag.

Social media analyse

Uit de analyse van 37.061 reacties op Nu.nl blijkt dat bijna een vijfde van de reacties niet door de moderatie komt, met opvallend veel afkeuring bij berichten over transgender personen en LHBTIAQ+, wat wijst op een verhoogd risico op online geweld richting deze groepen. Aanvullend onderzoek op X toont dat bij termen als *moslim*, *wappie* en *tokkie* regelmatig beledigende en discriminerende taal voorkomt; zo werd in een derde van de berichten over moslims ook het woord “tuig” gebruikt. Hoewel het merendeel van de berichten technisch gezien acceptabel is, komt problematisch taalgebruik – vooral richting publieke figuren – duidelijk naar voren. De beperkingen van automatische classificatie onderstrepen dat dit onderzoek een eerste verkenning is van de schaal en aard van online geweld in Nederland.

Aanbevelingen

Dit onderzoek onderscheidt vijf aanbevelingen die niet op prioriteit zijn gerangschikt. Om tot een effectieve aanpak van online geweld te komen is het van belang ze in samenhang in te zetten. Onder de verschillende aanbevelingen benoemen we concrete acties en stappen:

1. **Samenwerking en gedeelde verantwoordelijkheid** – Online geweld vraagt om een integrale en holistische aanpak met duurzame, domeinoverstijgende samenwerking tussen publieke en private partijen, nationaal én internationaal. Slachtoffers moeten kunnen rekenen op ondersteuning, opvolging en opsporing.
2. **Regie en coördinatie** – Het Ministerie van Justitie & Veiligheid kan de centrale regierol nemen, in samenhang met andere ministeries en afdelingen. De aanpak moet zowel repressie (achterkant) als preventie en vroegsignalering (voorkant) omvatten.
3. **Verantwoordelijkheid van techbedrijven** – Techbedrijven hebben een directe verantwoordelijkheid: transparant beleid, actieve handhaving en platformoverstijgende afspraken zijn noodzakelijk.
4. **Kennis, onderzoek en monitoring** – Structureel investeren in systematische monitoring, onderzoek en interventies is essentieel om inzicht te krijgen in de aard, omvang en dynamiek van online geweld. Ook moet een eenduidige definitie worden ontwikkeld en expertise van onderzoekers en professionals versterkt.
5. **Preventie en bewustwording** – Stel duidelijk de norm dat online geweld óók geweld is. Investeer in educatie, digitale geletterdheid en weerbaarheid, ondersteun professionals en rolmodellen, en stimuleer een inclusieve online cultuur en veilige publieke ruimte.

Extra informatie

Aanvullend op dit rapport zijn een factsheet en video gemaakt.

- Downloadbare factsheet: De wereld van online geweld.
- Video: Online geweld: wat is het en hoe vaak komt het voor?

1 Inleiding

De online wereld is een onmisbaar onderdeel geworden van het dagelijks leven. Sociale media en online platforms bieden kansen voor ontmoeting, (creatieve) expressie, dialoog en maatschappelijke betrokkenheid. Maar diezelfde platforms zijn ook de plek waar spanningen oplopen, grenzen vervagen en geweld tussen burgers nieuwe vormen aannemen. Online geweld; van bedreigingen en intimidatie tot het aanzetten tot haat en geweld is een *complex maatschappelijk probleem*. De snelheid waarmee berichten zich verspreiden, de vaak anonieme aard van communicatie en het publieke karakter van online platforms maken dit geweld ontwrichtend. Niet alleen voor slachtoffers die te kampen hebben met bijvoorbeeld depressie, angst, stress of reputatieschade, maar ook voor de samenleving als geheel. Het vertrouwen in de online leefwereld daalt, polarisatie neemt toe en het (online) maatschappelijke debat verhardt.

Online geweld tussen burgers kent veel kanten. Het raakt juridische, technologische én sociale grenzen, en vraagt om een breed gedragen aanpak. Om die te kunnen ontwikkelen is beter inzicht nodig in de aard en omvang, de onderliggende factoren en impact van dit type geweld. Tegelijkertijd is online geweld tussen burgers onderling een *'nieuw' vraagstuk* voor zowel onderzoek/wetenschap, beleid als de uitvoering. De nieuwsberichten laten zien dat online geweld niet meer is weg te denken. Denk bijvoorbeeld aan de berichten over Roblox en een aflevering van BOOS.[1] De wereld van online en offline geweld is vervlochten, dat wordt ook zichtbaar bij de gebeurtenissen in Zwijndrecht en Beverwijk.[2] Maar alledaagse verhalen of ervaringen over online geweld tussen burgers onderling blijven nog achterwege. Hoe ziet die wereld van online geweld eruit, wat doet het met burgers en hoe gaan hulpverleners/professionals hiermee om? Allemaal vragen waar nog geen zicht op is.

In opdracht van het Ministerie van Justitie en Veiligheid voert het Verwey-Jonker Instituut onderzoek uit naar (strafbare) vormen van online geweld tussen burgers. Dit rapport presenteert de belangrijkste bevindingen over de aard en omvang van dit fenomeen, de betrokken burgers ('plegers' en 'slachtoffers') en professionals/experts en hun perspectieven en ervaringen, en mogelijke interventies en beleidsmaatregelen om dit vraagstuk aan te pakken.

1 'Nederlandse kinderen actief in chatgroepen voor seksuele afpersing en geweld' Inzichten in online haatspraak en discriminatie; Hoe bedreigingen jouw vrijheid in gevaar brengen (met Dilan Yeşilgöz) | BOOS S12E12.

2 Jongerengeweld Zwijndrecht en Beverwijk 'wordt online extra aangezwengeld'; Angst onder jongeren in Beverwijk: 'Ik ga 's avonds liever niet naar buiten'.

1.1. Drie vormen van online geweld

Online geweld is een overkoepelende term voor delicten en vormen van geweld die plaatsvinden in het digitale domein—binnen de online leefwereld van mensen—en gericht zijn op burgers en 'publieke figuren'. Verschillende vormen van online geweld worden vaak in samenhang of door elkaar heen gebruikt, kunnen in combinatie worden ingezet met fysieke delicten en escaleren vaak in intensiteit en impact. Vanwege de brede reikwijdte van online geweld, is het van belang de scope te beperken. In dit onderzoek ligt de focus op de volgende drie vormen van *online geweld tussen burgers onderling*, aangezien het Ministerie hier haar aanpak[3] op wil richten en onderzoek hiernaar nog schaars is (zie ook hoofdstuk 2):

1. Online bedreiging en intimidatie – gericht op het zaaien van angst, vaak via directe dreigementen.
2. Haatzaaien en discriminatie – het aanzetten tot haat of discriminatie tegen specifieke groepen.
3. Kleinschalige ordeverstoringen en opruiing – het oproepen tot gewelddadig of strafbaar gedrag, vaak in groepsverband.

We begrijpen dat er meerdere vormen van online geweld zijn die ook in elkaars verlengde liggen. Bijvoorbeeld *doxing*[4], waar we in dit rapport ook hier en daar op ingaan wanneer het samengaat met de bovengenoemde drie vormen van online geweld. Door de beperking van geweld tot het publieke en semipublieke domein worden huiselijk en seksueel geweld (zoals stalking, sextortion[5],

3 In het geschetste beleid heeft een aantal geweldthema's extra prioriteit en zijn in lijn met het regeerprogramma toezeggingen gedaan voor een intensivering van de aanpak. Het gaat hier om de volgende zes geweldthema's: aanslagen met explosieven, straatroven, online geweld, geweld in de sport, geweld in het openbaar vervoer en geweld tegen werknemers (Brief Tweede Kamer Aanpak geweld, 5 april 2025).

4 Het verzamelen of delen van persoonsgegevens, zoals een adres of telefoonnummer, om iemand te intimideren is sinds 1 januari 2024 strafbaar. Dit wordt ook wel doxing genoemd (Doxing | Privacy en persoonsgegevens | Rijksoverheid.nl).

5 Sextortion is iemand chanteren met naaktbeelden. Dit wordt ook wel afpersing genoemd. Het gaat bijvoorbeeld om naaktfoto's of video's die van je zijn gemaakt met een webcam. Het kan ook gaan om nepfoto's of nepvideo's (deepfakes) (Hulp na sextortion - Slachtofferhulp Nederland).

grooming[6], etc.) in dit onderzoek niet meegenomen. Ook cultureel en politiek geweld blijven buiten beschouwing. Verder gaat het om online geweld gericht op personen en valt online geweld tegen dieren of goederen (vandalisme) buiten de scope van het onderzoek.

1.2. Probleemstelling en deelvragen

Het onderzoeksdoel is om meer inzicht te krijgen in de verschillende uitingsvormen van online geweld, de onderliggende maatschappelijke context en de profielen van plegers en slachtoffers. Daarnaast gaan we in op de impact van online geweld op slachtoffers en de bredere samenleving, evenals de behoeften van slachtoffers aan ondersteuning.

De deelvragen die centraal staan, zijn:

1. Welke vormen van online geweld zijn het meest voorkomend in Nederland? (o.a. aard, omvang, gebruikte platforms, typologieën)
2. Wie zijn de plegers van deze vormen van geweld, wat zijn hun motieven en kenmerken? (o.a. leeftijd, online gedrag, antecedenten, opleiding, gezondheid, beweegredenen)
3. Wie zijn de slachtoffers van deze vormen van geweld, welke impact heeft het online geweld, welke behoeften hebben zij en wat zijn hun kenmerken? (o.a. leeftijd, online gedrag, ondersteuningsbehoeften en copingmechanismen)
4. Is er sprake van overlap tussen slachtofferschap en plegerschap? (wanneer, binnen welke maatschappelijke context)
5. Wat zijn de modus operandi van plegers?
6. Wat zijn de weerbaarheidsstrategieën/copingmechanismen van slachtoffers?
7. Hoe ziet de interactie tussen plegers en slachtoffers eruit?
8. Hoe hangt online geweld samen met fysiek geweld, welke stadia kunnen er worden onderscheiden? Hoe draagt de online wereld/dynamiek bij aan offline geweld?
9. Welke preventieve en repressieve interventies kunnen mogelijk ontwikkeld en geïmplementeerd worden om (a) plegers te stoppen en (b) slachtoffers te beschermen en ondersteunen? En welke belemmeringen zijn er in het huidige beleid om online geweld aan te kunnen pakken?

6 Grooming is digitaal kinderlokken. Er is sprake van grooming als een volwassene via ICT contact legt met een kind, met de intentie om dat kind te ontmoeten met het doel om seksueel misbruik te plegen of kinderpornografische afbeeldingen te produceren (Wat is grooming? | politie.nl).

1.3. Onderzoeksaanpak

De focus van dit onderzoek naar *online geweld tussen burgers onderling* is nieuw. In dit onderzoek hebben we ons daarom ook voornamelijk laten voeden door de verhalen en ervaringen uit de praktijk. Het onderzoek is *praktijkgericht* waarbij we de perspectieven van burgers en professionals/experts centraal stellen. In dit rapport maken we volop gebruik van *levendige voorbeelden* waaronder citaten en casussen. De citaten geven kleur aan de bevindingen die we bij meerdere respondenten hebben opgehaald. Wanneer iets maar eenmalig is gehoord, zullen we dat ook bij de bevindingen weergeven.

Om goed zicht te krijgen op de verschillende deelvragen, maken we gebruik van verschillende onderzoeksmethoden. We hebben gekozen voor een *multimethod design* waarin we kwalitatieve- en kwantitatieve onderzoeksmethoden integreren.

Hieronder beschrijven we achtereenvolgens de verschillende onderzoeksactiviteiten:

1.3.1. Literatuurstudie

De literatuurstudie vormt het eerste onderdeel van het onderzoek en is gebaseerd op een brede verkenning van (inter)nationale bronnen. Daarbij is gebruik gemaakt van zowel wetenschappelijke publicaties als grijze literatuur, zoals rapporten van kennisinstituten en beleidsdocumenten. Voor de selectie van relevante studies is uitsluitend gekeken naar publicaties uit de periode 2015–2025. De focus lag specifiek op strafbare vormen van online geweld, met bijzondere aandacht voor de drie vormen van online geweld waar we in dit onderzoek op focussen: (1) *bedreiging en intimidatie* (2) *haatzaaien en discriminatie* en (3) *kleinschalige ordeverstoringen*.

De literatuurstudie is niet systematisch van opzet, maar richt zich doelgericht op relevante thema's rond deze drie vormen van online geweld tussen burgers. De geselecteerde literatuur is thematisch geanalyseerd met behulp van de kwalitatieve analysetool ATLAS. Ti. Tijdens dit proces is een codeboom ontwikkeld, die op iteratieve wijze tot stand is gekomen. Hierbij is gewerkt met een combinatie van inductieve en deductieve benaderingen: enerzijds zijn codes afgeleid uit bestaande theorieën en onderzoeksvragen, anderzijds zijn nieuwe codes ontstaan uit patronen en inzichten die tijdens de analyse naar voren kwamen. Deze werkwijze maakt het mogelijk om zowel vooraf gedefinieerde als spontaan opkomende thema's systematisch te verkennen en een verdiepend inzicht te verkrijgen in de aard, omvang en maatschappelijke context van online geweld.

De volledige literatuurstudie 'Online geweld tussen burgers onderling: wat weten we al? is hier te vinden: Stramienversie Verwey-Jonker publicatie.

1.3.2. International Self-Report Delinquency Study (ISRD) Online geweld onder jongeren

Het gaat hier om een kwantitatieve analyse van zelfgerapporteerd pleger- en slachtofferschap van online geweld onder jongeren. De vormen van online geweld die hier naar voren komen, zijn breder dan de focus van dit onderzoek. Hierbij gaat het om de volgende delicten:

- Versturen, delen of reposten van een intieme foto of video van iemand tegen diens wil in.
- Versturen van kwetsende berichten of reacties op sociale media over iemands etnische achtergrond, nationaliteit, religie, genderidentiteit, seksuele voorkeur, of een andere soortgelijke reden.
- Gebruiken van internet, e-mail of sociale media om anderen te bedreigen en/of te misleiden om geld te verdienen.
- Hacken van een computer of account om data te verkrijgen, dan wel te vernietigen.

Aangezien de data over aard en omvang van online geweld zeer schaars is, is ervoor gekozen om deze analyse mee te nemen in dit onderzoek, om zo meer zicht te krijgen op de aard en omvang van online geweld onder jongeren. Wij hebben een kwantitatieve analyse uitgevoerd op basis van data uit de vierde wave van de *International Self-Report Delinquency Study* in Nederland. Deze data zijn in 2023 en 2024 verzameld onder scholieren in het voortgezet onderwijs (regio's Den Haag en Amsterdam). Deze data zijn niet representatief voor de doelgroep, maar indicatief. De kwantitatieve analyse hebben we vormgegeven in de factsheet Online geweld onder jongeren. In de factsheet wordt enerzijds de reguliere beschrijvende statistiek weergegeven (absolute aantallen, percentages en gemiddelden). Anderzijds is er bij de achtergrondkenmerken getoetst in hoeverre de gemiddelde scores per groep (bijvoorbeeld meisjes/jongens of leeftijdscategorie) significant verschillen. Hiertoe is in het statistische programma SPSS gebruikgemaakt van zogeheten t-toetsen (analyse om gemiddelde scores van twee groepen te vergelijken en te toetsen op significantie) en one-way ANOVA-toetsen (analyse om gemiddelde scores van meerdere groepen te vergelijken en te toetsen op significantie). Naast deze analysen is er op basis van bestaande literatuur over offline jeugdcriminaliteit gekeken naar de mate van samenhang tussen uit de literatuur gebleken relevante variabelen en online pleger- en slachtofferschap. Dit is gedaan met behulp van Pearson- en Spearman rho-correlatietoetsen. Om de correlatiewaarden (ofwel de mate waarin de ene variabele samenhangt met de andere variabele) te interpreteren, is de volgende interpretatie gebruikt: groter dan 0,5 duidt op een sterke samenhang; tussen 0,3 en 0,5 duidt op een matige samenhang en tussen de 0 en 0,3 duidt op een zwakke samenhang.

1.3.3. Social media analyse

In dit onderzoek hebben we met de methode van Natural Language Processing (NLP) meer dan tienduizenden reacties onder nieuwsberichten van Nu.nl (NU - Het laatste nieuws het eerst op NU.nl) en *posts* (voorheen *tweets*) op X geanalyseerd om inzicht te krijgen in het gebruik van potentieel discriminerend, bedreigend, beledigend of onbeschoft taalgebruik of desinformatie. Deze indeling is deels gebaseerd op de scope van ons onderzoek en anderzijds op basis van de literatuurstudie en methodologische literatuur van sociale media scrapings. We hebben hierbij naar gevoelige en gepolariseerde onderwerpen gekeken om te kijken hoe mensen hier online over praten. In Bijlage 1 beschrijven we de uitgevoerde social media analyses stap voor stap. Hieronder volgt een verkorte samenvatting.

Voor de dataverzamelingsfase hebben we twee bronnen gebruikt. Ten eerste hebben we van Nu.nl (NU - Het laatste nieuws het eerst op NU.nl) een steekproef ontvangen van reacties onder een aantal nieuwsartikelen uit 2024 (in totaal 37.061 reacties). We ontvingen zowel reacties die door het moderatieteam waren afgekeurd op basis van de huisregels van Nu.nl, alsook geplaatste/goedgekeurde reacties. Deze nieuwsartikelen hebben we gegroepeerd op basis van zes overkoepelende thema's: LHBTQIA+, Klimaat, Buitenland, Immigratie/Asiel, Transgender personen en Overig. Deze thematische indeling komt niet overeen met de standaard indeling van artikelen op Nu.nl. Zo beoogden we reacties te analyseren die samenhangen met het bredere maatschappelijk debat.

Ten tweede is met behulp van het scrapingprogramma Zeeschuimer alle Nederlandstalige posts op X verzameld van 1 juli tot 31 december 2024 die de zoektermen "Wappie", "Tokkie", "Moslim" of "Fascist" bevatten (in totaal 236.230 posts).

Zeeschuimer[7]

Zeeschuimer is een browserextensie ontworpen voor onderzoekers die systematisch content op sociale mediaplatforms willen bestuderen, vooral op platforms die conventionele scrapingmethoden of API-gebaseerde dataverzameling bemoeilijken. De extensie is ontwikkeld door Stijn Peters voor de Digital Methods Initiative, wat onderdeel is van de Universiteit van Amsterdam.

7 Peters, S. (z.d.). Zeeschuimer [Computer software]. GitHub. https://github.com/digitalmethodsinitiative/zeeschuimer

Momenteel kan het data verzamelen van de volgende platformen:

- TikTok (posts en reacties)
- Instagram (alleen posts)
- X/Twitter
- LinkedIn
- 9gag
- Imgur
- Douyin
- Gab
- Truth Social
- Pinterest
- RedNote/Xiaohongshu

Voordat we begonnen aan de analyses hebben we alle reacties en posts opgeschoond. Alle gebruikersnamen, hashtags, URL's en speciale tekens zijn verwijderd om ruis te voorkomen. Vervolgens zijn leestekens, cijfers en overbodige witruimtes weggehaald, en zijn alle woorden omgezet naar kleine letters. Tot slot is een uitgebreid stopwoordenfilter toegepast, inclusief algemeen geaccepteerde lijsten en een context specifieke aanvulling voor termen die in onze data te frequent en niet-informatief bleken.

Voor de analyses gebruikten we de lexicons van The Weaponized Word[8]: woordenlijsten voor discriminerende, beledigende, bedreigende termen en zogenaamde watchwords. Hierbij een overzicht van de inhoud van deze lexicons:

1. Discriminerende woorden: woorden die specifiek gericht zijn op de ontvanger zijn nationaliteit, etniciteit, religie, gender, seksuele oriëntatie, beperking en/of chronische ziekte of sociaaleconomische klasse.
2. Beledigende woorden: woorden die beledigend worden gebruikt, maar niet direct betrekking hebben op de sociale identiteit van de ontvanger.
3. Bedreigende woorden: woorden die (direct of indirect) dreigingen of geweld impliceren.
4. Watch words: signalerende woorden die samenhangen met haatspraak, of beledigend of dreigend taalgebruik. In de tekst hieronder zullen we spreken van 'risicovolle' woorden.

Per reactie of post is geteld hoe vaak lexicontermen voorkwamen en welk percentage berichten ten minste één term bevatte. Bij Nu.nl koppelden we reacties aan de zes thematische categorieën om per onderwerp te kunnen vergelijken hoe vaak schadelijke taal voorkomt en hoe moderatiestatus (geaccepteerd versus afgekeurd) daarmee samenhangt. Voor X maakten we verkennende wordclouds per zoekterm, keken we welke lexicons het vaakst voorkomen per zoekterm en rangschikten we de tien meest voorkomende lexiconwoorden.

8 Lexicographic data courtesy of The Weaponized Word (weaponizedword.org).

The Weaponized Word[9]

The Weaponized Word is een organisatie die publieke en private organisaties helpt om een gezonde online gemeenschap te beschermen tegen gebruikers die schadelijke content verspreiden. The Weaponized Word maakt gebruik van dynamische woordenlijsten die voortdurend worden bijgewerkt: van bekende haatwoorden en bedreigingen tot phishing en bronnen van misinformatie. Daarnaast 'leert' de techniek welke woordcombinaties en zinspatronen typisch zijn voor negatieve of schadelijke uitingen. Zo kan het snel herkennen wanneer iemand in een bericht discrimineert, beledigt, bedreigt of verkeerde informatie verspreidt, zonder dat er naar persoonlijke gegevens wordt gekeken. Op die manier helpt The Weaponized Word platformen gezond te houden: het beperkt de verspreiding van schadelijke content en geeft moderators de kans om onmiddellijk in te grijpen en de dialoog veilig te houden.

Dankzij deze aanpak kunnen beheerders van online gemeenschappen razendsnel zien welke bijdragen extra aandacht nodig hebben. In tientallen talen signaleren ze woorden en uitdrukkingen die vallen onder de volgende lexicons:

1. **Discriminerende woorden:** woorden die specifiek gericht zijn op de ontvanger zijn nationaliteit, etniciteit, religie, gender, seksuele oriëntatie, beperking en/of chronische ziekte of sociaaleconomische klasse.
2. **Beledigende woorden:** woorden die beledigend worden gebruikt, maar niet direct betrekking hebben op de sociale identiteit van de ontvanger.
3. **Bedreigende woorden:** woorden die (direct of indirect) dreigingen of geweld impliceren.
4. **Watch words:** signalerende woorden die samenhangen met haatspraak, of beledigend of dreigend taalgebruik.

Aanvullend hebben we in dit onderzoek de posts op X geclassificeerd om de aard en omvang van schadelijke of problematische uitingen betrouwbaar in kaart te brengen. Tekstclassificatie is een van de fundamentele toepassingen van NLP (Rayinto et al., 2023). In de kern is tekstclassificatie het proces waarbij we ongestructureerde data toewijzen aan vooraf gedefinieerde categorieën. Deze categorieën kunnen sentimentlabels zijn ("positief", "negatief"), thematische labels ("politiek", "gezondheid") of, zoals in ons onderzoek, vormen van schadelijke inhoud, o.a.: acceptabel, discriminatie, haatspraak en belediging.

9 Idem.

Om sociale mediaberichten systematisch te analyseren, hebben we een indeling gemaakt in verschillende categorieën. Deze classificaties zijn gebaseerd op wetenschappelijke inzichten en juridische definities. Hieronder lichten we per categorie toe wat we eronder verstaan en op basis waarvan berichten zijn gecodeerd (een volledige toelichting staat in Bijlage 1.3):

- Acceptabel
- Smaad en laster
- Discriminatie, haatspraak & haatzaaien
- Bedreiging en intimidatie
- Belediging
- Desinformatie
- Kleine ordeverstoringen en opruiing
- Counterspeech
- Publieke Figuren
- Uit te filteren berichten

Om een classificatiemodel te bouwen, beginnen we met een gelabelde dataset: een steekproef van posts waar wij elke post beoordelen volgens een codeboek. Die labels vormen de basis waarvan het model leert. Daarom hebben we een willekeurige sample getrokken van vijfhonderd posts voor elke zoekterm. Deze samples gebruikten we om vijfhonderd Nederlandstalige berichten handmatig te labelen. Drie onderzoekers labelden elk bericht aan de hand van het codeboek (zie bijlage). Vervolgens splitsten we die gelabelde data in een trainingsset (80 %) en een testset (20 %). Hiermee kunnen we kijken hoe goed het model presteert op data die het nog niet eerder heeft gezien. In de training leert het model patronen in woordgebruik, zinsbouw en context te associëren met de labels. Daarna toetsen we op de testset hoe vaak het model correct voorspelt (nauwkeurigheid), hoe vaak het relevante uitingen herkent (recall) en hoe vaak de voorspellingen zuiver zijn (precisie).

Onderzoek toont telkens aan dat BERT-modellen hun traditionele tegenhangers verslaan op metrics als F1-score of accuracy, vooral bij korte, context-afhankelijke teksten of in meertalige omgevingen (Yu et al., 2019; Wu & Xu, 2024; Nishigaki et al., 2023). In dit onderzoek vergelijken we daarom drie transformer-modellen: de Nederlandse modellen RobBERT v2 (Delobelle, Winters & Berendt, 2020) en BERTje (De Vries et al., 2019), een meertalig XLM-T model (Barbieri et al. (2021). In onze analyses trainen en evalueren we daarom deze drie individuele modellen op exact dezelfde 80/20-split van de sample. Per model vergelijken we de F1-scores per zoekterm én voor wanneer we de gelabelde tweets voor de zoektermen samen nemen. Naast deze individuele modellen bouwen we een ensemble dat de voorspellingen van alle individuele modellen combineert. Hierbij leunen we op het fenomeen *majority voting*, wat inhoudt dat het label wordt gekozen die door de meerderheid van de BERT-modellen wordt gekozen. In andere woorden, als twee van de drie modellen aangeven een bepaalde post als discriminatie en haatspraak labelen, dan krijgt deze post dit label. Op deze manier kunnen we een robustere labelling van de posts krijgen omdat de modellen elkaar aanvullen.

Na fine-tunen van meerdere modellen selecteren we het model met de beste balans tussen accuracy, precisie, recall en F1 score (zie kader). Dat model passen we vervolgens toe op de volledige dataset met posts. We hebben hierbij zowel gekeken of er een model is dat voor de 1500 handmatig gelabelde posts samen werkt, of dat sommige modellen beter werken voor een bepaalde zoekterm dan voor een andere.

Hoe beoordelen we of het model goed werkt?

Om te controleren of een model goede voorspellingen doet, kijken we naar vier veelgebruikte cijfers: accuracy, precision, recall en de F1-score. (1) Accuracy laat zien hoeveel van de voorspellingen helemaal correct zijn. Maar dat getal zegt niet alles. Het kan zo zijn dat één categorie veel vaker voorkomt dan de andere, dan kan het model "goed scoren" door alleen die ene categorie te kiezen. In andere woorden, het is alsof er altijd A ingevuld wordt op een meerkeuzetoets. Als toevallig veel antwoorden ook echt A zijn, dan is de uitkomst best goed, maar dit komt meer door toeval dan door kennis. Daarom kijken we ook naar precision (hoe vaak het model gelijk had als het een bepaalde categorie voorspelde) en recall (hoe goed het model alle voorbeelden van die categorie weet te vinden). De F1-score is een combinatie van precision en recall en geeft een gebalanceerd beeld van hoe goed het model per categorie presteert. We letten vooral op de F1-score per categorie, omdat dat het beste laat zien waar het model goed of slecht in is. Accuracy alleen kan misleidend zijn.

1.3.4. Interviews praktijk- en wetenschappelijke experts

Aanvullend op de voorlopige inzichten uit de literatuurstudie hebben we n=12 interviews afgenomen met praktijk-/wetenschappelijke experts (zie bijlage 2). Deze interviews waren gericht op het verdiepen van het inzicht in de aard en omvang van online geweld, de kenmerken van plegers en slachtoffers, bestaande en mogelijke interventies en beleidsmaatregelen, en de relevantie van (inter)nationale kennis voor de Nederlandse context.

Bij de selectie van de experts hebben we bewust gestreefd naar een evenwichtige samenstelling van wetenschappelijke en praktijkgerichte deskundigen. Daarbij is specifiek gelet op expertise in verschillende vormen van online geweld, zoals haatberichten, intimidatie, bedreiging en opruiing, evenals ervaring met diverse doelgroepen, waaronder jongeren, LHBTQIA+-personen en mensen met een migratieachtergrond. We hebben onder meer experts benaderd die betrokken zijn bij recent onderzoek, beleidsontwikkeling of praktijkinitiatieven vanuit verschillende relevante domeinen, zoals wet & regelgeving, veiligheid, onderwijs, slachtoffer- en plegersondersteuning en techbedrijven.

Opvallend is dat de expertise over de vormen van online geweld waar wij ons in dit onderzoek op richten (haatzaaien/discriminatie, intimidatie/bedreiging, kleinschalige opruiing), *zeer schaars* is. Daarbij zijn wel genoeg aanliggende terreinen zoals cybercriminaliteit, cyberpesten, oproepen tot grootschalige ordeverstoringen, eergerelateerd geweld, politiek geweld, seksueel geweld etc. die ook een groot online geweld component hebben en die inzichten kunnen bieden. De schaarste aan expertise op dit specifieke thema van online geweld tussen burgers beschrijven we later ook in hoofdstuk 5.

Voor de interviews is een reflexieve benadering gehanteerd (Pessoa et al., 2019). Dit betekent dat we niet alleen feitelijke informatie hebben verzameld, maar de experts ook actief hebben uitgenodigd om samen met ons stil te staan bij de betekenis van bestaande kennis, lacunes in onderzoek en aandachtspunten voor beleid en praktijk. De topic guide en analysemethode zijn ontwikkeld op basis van inzichten uit de literatuurstudie en bestaande wetenschappelijke kaders. Tegelijkertijd werkten we iteratief: bevindingen uit de eerste interviews zijn gebruikt om vervolginterviews verder aan te scherpen en de analysecategorieën waar nodig bij te stellen.

De interviews zijn getranscribeerd en vervolgens geanalyseerd met behulp van een thematic analysis, zoals beschreven door Braun en Clarke (2006). Hierbij hebben we systematisch gezocht naar terugkerende thema's en patronen binnen de interviews, met oog voor zowel overeenkomsten als nuanceverschillen tussen de perspectieven van verschillende experts. Voor deze analyse is een codeboom gebruikt die iteratief is ontwikkeld en aangescherpt tijdens de onderzoeksfase, op basis van de literatuur, de eerste interviews en tussentijdse reflecties binnen het onderzoeksteam. Door deze werkwijze konden we flexibel inspelen op nieuwe inzichten en ontwikkelingen binnen dit relatief jonge onderzoeksveld en tegelijkertijd zorgen voor een consistente, transparante analyse.

1.3.5. Interviews plegers en slachtoffers - casusonderzoek

Om naast de inzichten van experts ook de perspectieven van direct betrokkenen verder te verdiepen, hebben we n=12 diepte-interviews afgenomen met personen die online geweld hebben ervaren als slachtoffer en/of als pleger: alle geïnterviewden waren slachtoffer, waarbij 3 geïnterviewden ook pleger waren. Deze interviews waren gericht op het verder uitdiepen van verschillende vormen van online geweld, de dynamiek tussen pleger en slachtoffer, de ervaringen en perspectieven van betrokkenen, de impact van het geweld, achterliggende motivaties, copingmechanismen en de ondersteuningsbehoeften van burgers.

Bij de samenstelling van de groep slachtoffers en plegers hebben we bewust gezocht naar variatie in typen online geweld, zoals haatberichten, intimidatie, bedreigingen en opruiing, evenals naar diversiteit in de achtergrond van de betrokkenen. Op die manier hebben we geprobeerd recht te doen aan de brede waaier aan ervaringen en motieven die binnen dit thema spelen. Voor de werving van deelnemers zijn verschillende strategieën ingezet. We hebben een wervingsvideo opgenomen waarin we burgers hebben opgeroepen om mee te doen aan ons onderzoek Heb jij te maken gehad met online geweld?. We hebben vervolgens de hulp ingeschakeld van representatieve en ondersteunende organisaties, zoals belangenverenigingen, hulpverleningsorganisaties en partners uit het zorg-, onderwijs- en veiligheidsdomein. Deze partij hebben bijvoorbeeld intern een oproep uitgezet (bijvoorbeeld door de wervingsvideo te delen) maar ook door mondeling collega's op ons onderzoek te wijzen. Daarnaast hebben we actief geworven via online platforms, waaronder Reddit en TikTok en via meer activistische organisaties. De werving via sociale media en de meer activistische of belangen vertegenwoordigende organisaties bleken uiteindelijk het meest effectief om een diverse groep slachtoffers en plegers te bereiken, mede vanwege de laagdrempelige toegang en de mogelijkheid tot anonimiteit die deze kanalen bieden.

Bij de uitvoering van de interviews hebben we aangesloten bij de wensen, veiligheid en anonimiteit van de deelnemers. Zo konden deelnemers bijvoorbeeld zelf aangeven of zij het gesprek online of fysiek wilden voeren, en op welk moment dit voor hen het beste uitkwam. Ook hielden we rekening met persoonlijke voorkeuren, bijvoorbeeld of zij liever met een mannelijke of vrouwelijke interviewer spraken. Daarnaast zijn we in de gespreksvoering zorgvuldig omgegaan met de gevoeligheid van het onderwerp en hebben we deelnemers expliciet de ruimte gegeven om grenzen aan te geven. Voor de interviews hebben we, net als bij de expertgesprekken, een reflexieve methode toegepast, waarbij deelnemers actief werden uitgenodigd om samen met ons stil te staan bij de betekenis van hun ervaringen en de bredere context ervan (Pessoa et al., 2019). Waar gewenst, boden we ondersteuning via de storyboard-methode, een visuele, arts-based aanpak waarbij deelnemers hun verhaal konden structureren met behulp van iconen, tijdslijnen en andere verbeeldende elementen (de Jager et al., 2017; Macken et al., 2021; Drew et al., 2010).

In de praktijk gaven de meeste deelnemers echter de voorkeur aan een meer natuurlijke, vrije vertelvorm, waarbij zij zonder visuele ondersteuning hun ervaringen deelden. Alle interviews zijn getranscribeerd en vervolgens geanalyseerd met behulp van een thematische analyse. Hierbij is gebruik gemaakt van dezelfde codeboom als die voor de analyse van de expertinterviews. Deze codeboom is iteratief (door)ontwikkeld: op basis van inzichten uit de literatuurstudie, de eerste interviews en tussentijdse analyses zijn de categorieën en codes waar nodig aangescherpt en verfijnd. Op deze manier konden we systematisch en consistent de rode draden, verschillen en nuance in de ervaringen van zowel slachtoffers als plegers in kaart brengen.

Op basis van de interviews hebben we de casussen uitgewerkt. Deze casussen worden ook door het rapport heen gebruikt ter illustratie van de verschillende elementen van online geweld.

Tabel 1.1. Pleger en slachtoffer deelnemers

Interview nr.	Pleger / Slachtoffer	Type online geweld
01	Slachtoffer/pleger	Bedreiging + doxing
02	Slachtoffer	Bedreiging
03	Slachtoffer/pleger	Haatspraak, discriminatie
04	Slachtoffer/pleger	Haatspraak, discriminatie
05	Slachtoffer	Bedreiging (intimidatie)
06	Slachtoffer	Haatspraak, discriminatie
07	Slachtoffer	Discriminatie, haatspraak, bedreiging
08	Slachtoffer	Bedreiging
09	Slachtoffer	Smaad, laster en doxing
10	Slachtoffer	Doxing, intimidatie
11	Slachtoffer	Bedreiging
12	Slachtoffer	Bedreiging

1.3.6. Focusgroepen

We hebben twee digitale focusgroepen georganiseerd met professionals/experts onder andere werkzaam in de strafketen en slachtofferhulp. Denk aan beleidsmakers en uitvoerend praktijkexperts binnen politie, OM en slachtofferhulp, gemeenten en social media platformen (zie bijlage voor de lijst met deelnemende organisaties per focusgroep). Tijdens de focusgroepen stond de interactie en dialoog met de deelnemers voorop. We hebben casussen – afkomstig uit de gesprekken met plegers en slachtoffers - gebruikt om praktijkervaringen te analyseren en te bespreken. Er is gekozen voor verschillende casussen die onze bevindingen goed weerspiegelden. Hierbij hebben we rekening gehouden met de verschillende thema's, gebaseerd op onze deelvragen: achtergrond en kenmerken plegers/slachtoffers, maatschappelijke context, motivatie, coping mechanismen/ strategieën, belemmeringen in de huidige aanpak en ideeën voor een effectieve aanpak van online geweld. Op basis van de casussen hebben we met elkaar het gesprek gevoerd en hebben deelnemers op onze bevindingen kunnen reflecteren maar ook nieuwe input kunnen geven. Ook hier hebben we weer gebruik gemaakt van een thematische analyse in AtlasTi, gebaseerd op dezelfde codeboom als bij de overige onderzoeksactiviteiten. Hierbij hebben we de codeboom ook weer iteratief (door)ontwikkeld.

1.3.7. Online reflectiesessie

Vervolgens hebben we een online reflectiesessie georganiseerd waarin we de belangrijkste onderzoeksresultaten hebben besproken. Het doel van de sessie was om de opgedane kennis op een interactieve manier te delen en hierop te reflecteren zodat belanghebbenden de kans hebben om te reflecteren op hoe de resultaten succesvol vertaald en benut kunnen worden in de praktijk. Daarbij hebben we de deelnemers ook om feedback gevraagd om onze aanbevelingen met betrekking tot een effectieve aanpak Online geweld verder aan te scherpen.

1.3.8. Expertsessie

Dit rapport zal voedingsbodem zijn voor de aanpak Online geweld vanuit het Ministerie van Justitie. Op basis van de inzichten van dit rapport is een fysieke expertsessie bij het Ministerie van Justitie en Veiligheid georganiseerd. Hierbij aanwezig waren verschillende afdelingen van het Ministerie van Justitie & Veiligheid, het Ministerie van Onderwijs, Cultuur en Wetenschap, het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, het Ministerie van Volksgezondheid, Welzijn en Sport, het Openbaar Ministerie, de Politie, Slachtofferhulporganisaties, de Vereniging van Nederlandse Gemeenten en het Centrum voor Criminaliteitspreventie en Veiligheid. Tijdens deze interactieve sessie zijn we met elkaar aan de slag gegaan met een holistische, integrale en domein-overstijgende aanpak van Online geweld tussen burgers onderling. Het doel hiervan was om verbinding en draagvlak te creëren en om ook al de eerste concrete stappen te zetten.

1.4. Leeswijzer

Deze rapportage is opgebouwd op basis van onze onderzoeksvragen. In **hoofdstuk 2** 'Wat weten we al'? staan we kort stil bij de belangrijkste bevindingen van de literatuurstudie. **Hoofdstuk 3** gaat in op de aard, dynamiek en omvang van online geweld. Vervolgens komen in **hoofdstuk 4** de achtergrondkenmerken, motieven, modus operandi en copingmechanismen. **Hoofdstuk 5** beschrijft de huidige interventies, belemmeringen en verbeterpunten voor een effectieve aanpak. In **hoofdstuk 6** sluiten we af met de belangrijkste bevindingen en presenteren we onze aanbevelingen. Vervolgens presenteren we de **Literatuurlijst** en **Bijlagen** waarin we de respondenten uiteenzetten en de sociale media analyse uitgebreid toelichten.

2 Wat weten we al?

In de reeds gepubliceerde literatuurstudie zijn waardevolle inzichten verkregen over de aard en omvang van online geweld, de betrokken burgers en de impact ervan. Daarnaast zijn er duidelijke kennishiaten geïdentificeerd. In dit hoofdstuk vatten we de belangrijkste bevindingen kort samen. Voor de volledige literatuurstudie verwijzen we naar: Onderzoek naar strafbare vormen van online geweld - Verwey-Jonker Instituut.

2.1. Aard en omvang

Het is lastig om een samenhangend beeld te vormen van de aard en omvang van online geweld in Nederland. Dit komt voornamelijk door het gebrek aan eenduidige definities en meetmethoden, onderrapportage door slachtoffers en beperkte toegang tot data van technologiebedrijven. Desondanks biedt de (inter)nationale literatuur een eerste indicatie van de aard en schaal van het probleem.

Aard: online geweld verwijst naar agressief en strafbaar gedrag dat plaatsvindt via digitale platforms en sociale media. Het richt zich vaak op individuen of groepen en kent vele vormen, waaronder bedreigingen, discriminatie en haatzaaien, doxing (het verspreiden van privégegevens), en het aanzetten tot fysiek geweld of opruiing. Vooral gemarginaliseerde groepen zoals vrouwen, LHBTQIA+ personen en mensen met een migratieachtergrond zijn vaak doelwit van dit soort geweld (Obemaier et al 2024; Wright & Wachs 2019; Castano-Pulgarin et al 2021; KIS 2022; Peterson & Densley 2017; Mukred et al 2014). De juridische kaders in Nederland definiëren verschillende vormen van online geweld als strafbaar. Zo vallen bedreigingen onder artikel 285 van het Wetboek van Strafrecht, haatzaaien onder artikel 137d, en opruiing onder artikelen 131 en 132. In de praktijk uiten deze vormen zich via sociale media in de vorm van berichten, memes, video's en livestreams.

Omvang: Hoewel er steeds meer cijfers beschikbaar zijn over strafbaar online geweld tussen burgers in Nederland, blijft het beeld onvolledig en inconsistent. Slechts een klein deel van de slachtoffers doet melding of aangifte, vaak uit angst voor vergelding, gebrek aan vertrouwen in instanties of onduidelijkheid over wat strafbaar is. Daarnaast hanteren verschillende organisaties uiteenlopende definities en meetmethoden, wat het vergelijken van cijfers bemoeilijkt.

Ook gegevens van techbedrijven blijven vaak gesloten, wat verdere analyse belemmert. In 2022 gaf 4% van de Nederlanders aan slachtoffer te zijn geweest van online bedreiging of intimidatie, met jongeren, jonge vrouwen en LHBTQIA+ personen als risicogroepen (CBS 2022). Online haatzaaien en discriminatie zijn eveneens wijdverspreid, maar worden zelden gemeld. In 2023 registreerde MiND 3.305 meldingen van online discriminatie, vooral op basis van herkomst, huidskleur en seksuele gerichtheid. Het CIDI rapporteerde een sterke toename van online antisemitisme, mede veroorzaakt door internationale conflicten. Cijfers over het online aanjagen van geweld en opruiing zijn schaars. Wel zijn er signalen dat online uitingen bijdragen aan maatschappelijke onrust, zoals bij voetbal gerelateerde incidenten.

Ook ten aanzien van de aard en omvang van online geweld onder jongeren, is de data schaars. Bestaand onderzoek naar de aard en omvang is er wel, maar richt zich meer op specifieke vormen van online geweld in plaats van de drie vormen waar we in dit onderzoek op focussen. Onder jongeren is online geweld wijdverspreid: ruim een kwart is slachtoffer geweest van online delicten zoals bedreiging, shamesexting of discriminatie (Burggraaff, Odekerken & Steketee, 2024). Jongeren zijn ook vaker pleger van online grensoverschrijdend gedrag. Meisjes zijn vaker slachtoffer van seksueel getinte berichten en stalking, terwijl jongens vaker betrokken zijn bij het verspreiden van intiem beeldmateriaal.

Ondanks de groeiende hoeveelheid data blijven er aanzienlijke kennishiaten bestaan. Zo is er geen volledig en eenduidig beeld van de omvang van online geweld, mede door onderrapportage, uiteenlopende definities en het ontbreken van gedeelde data van techbedrijven. Cijfers over het online aanjagen van geweld en opruiing zijn bijzonder schaars, waardoor het lastig is om de impact hiervan goed te duiden. Vooral de drie vormen van online geweld waar we ons op richten, ontbreken in bestaand onderzoek.

2.2. Burgers: plegers & slachtoffers

Er is nog relatief weinig bekend over de (achtergrond)kenmerken van burgers die betrokken zijn bij online geweld. Dit komt deels door de manier waarop online geweld wordt gerapporteerd, gedefinieerd en gemeten, maar ook doordat het om een breed en veelzijdig fenomeen gaat. Als we naar de literatuur kijken, dan zien we dat ook hier weer online geweld in brede zin wordt opgenomen en niet met de afbakening van de drie vormen waar we in dit onderzoek op focussen. Online geweld gaat dan ook over stalking, seksuele intimidatie etc. Ook wordt weer duidelijk dat inzicht ten aanzien van de kenmerken over plegers en slachtoffers van de drie vormen van online geweld ontbreekt. Als we naar de bestaande literatuur kijken, dan is er wel wat bekend over de achtergrond van andere vormen van online geweld. Daar zullen we hieronder met name op ingaan.

Zo verschillen de kenmerken en motieven van plegers van bijvoorbeeld online stalking of wraakporno van die van plegers van bijvoorbeeld opruiing of haatzaaien. Deze diversiteit maakt het lastig om algemene conclusies te trekken over 'de' pleger of 'het' slachtoffer van online geweld. Daarbij gaan veel onderzoeken nog steeds uit van een strikte scheiding tussen plegers en slachtoffers, terwijl in de praktijk vaak sprake is van een wederzijdse dynamiek. Slachtoffers kunnen zelf ook (re)ageren met grensoverschrijdend gedrag, en sommige plegers hebben eerder zelf online geweld ervaren (Mukred et al., 2024; Naranjo-Pou et al., 2025; Obermaier et al., 2024). Deze wisselwerking tussen plegerschap en slachtofferschap vraagt om een meer genuanceerde benadering dan in veel bestaand onderzoek wordt gehanteerd (zoals we in ons empirisch hoofdstuk 4 beschrijven).

"Slachtoffers": Slachtoffers van online geweld (in brede zin) vertonen vaak specifieke achtergrondkenmerken die hun "kwetsbaarheid" vergroten. Demografische factoren zoals geslacht, leeftijd, etniciteit en sociaaleconomische status kunnen hierbij een belangrijke rol spelen. Vrouwen lopen bijvoorbeeld een groter risico op seksuele intimidatie, online stalking en bedreigingen, terwijl jongens vaker te maken krijgen met fysieke bedreigingen en vernedering. Jongeren zijn over het algemeen gevoeliger voor cyberpesten, haatspraak en het ongewenst delen van intieme beelden, mede doordat sociale acceptatie in deze levensfase cruciaal is. Ook jongeren met een migratieachtergrond of een LHBTQIA+-identiteit worden vaker slachtoffer van respectievelijk racistische of homofobe haatspraak. Daarnaast spelen psychosociale factoren een grote rol. Jongeren met een laag zelfbeeld, beperkte sociale vaardigheden of weinig sociale steun zijn extra kwetsbaar. Sociale isolatie en onzekerheid over de eigen identiteit maken het moeilijker om effectief te reageren op online aanvallen, wat het risico op herhaald slachtofferschap vergroot. Ook mensen met een lagere sociaaleconomische status lopen meer risico, onder andere door beperkte digitale vaardigheden en minder toegang tot hulpbronnen, wat hen gevoeliger maakt voor online oplichting en manipulatie. Online geweld (in brede zin) is dus geen willekeurig verschijnsel, maar hangt nauw samen met bredere sociale en structurele ongelijkheden (Castellanos et al., 2023; ISRD Online geweld onder jongeren Factsheet; Mukred et al., 2024; 't Hoff-de Goede & Leukfeldt;). Inzicht in deze complexe wisselwerking is essentieel voor effectieve preventie en aanpak.

"Plegers": Plegers van online geweld (in brede zin) vertonen uiteenlopende kenmerken die zowel demografisch als psychosociaal van aard zijn. Hoewel jongeren vaak centraal staan in onderzoek, blijkt uit recente studies dat ook volwassenen en ouderen zich schuldig maken aan online geweld. Zo gaf een aanzienlijk deel van de 66-plussers aan recent online agressief gedrag te hebben vertoond. Dit laat zien dat online plegerschap niet beperkt is tot één leeftijdsgroep, maar voorkomt bij mensen met verschillende achtergronden en motieven (Pabian & Vandenbosch, 2024). Psychologische factoren spelen een belangrijke rol in het uiten van online geweld. Plegers ervaren vaak verhoogde niveaus van stress, angst of depressie, wat kan bijdragen aan agressief gedrag. Ook persoonlijke omstandigheden, zoals een negatieve thuissituatie of een gebrek aan emotionele steun, vergroten het risico. Mensen die zich onbegrepen of machteloos voelen, kunnen online agressie gebruiken als uitlaatklep of als manier om controle te herwinnen. Daarnaast zijn sociale factoren van invloed. Personen die zich sociaal geïsoleerd voelen of weinig steun ervaren van hun omgeving, hebben een grotere kans om zich online agressief te uiten. De anonimiteit van het internet versterkt dit gedrag, doordat plegers zich minder verantwoordelijk voelen en minder empathie tonen voor hun slachtoffers. Ook blijkt dat mensen met een sterke behoefte aan sociale dominantie – het streven naar superioriteit over anderen – vaker online haatspraak gebruiken. Dit gaat vaak gepaard met een lager empathisch vermogen en een grotere morele afstand tot hun gedrag, waardoor zij de impact van hun acties minder erkennen. Zulke plegers ontwikkelen bovendien vaker positieve opvattingen over geweld, wat hun bereidheid tot online agressie vergroot (Sincek et al., 2017; Elsaesser et al., 2017; Wright & Wachs, 2019; Barlett et al., 2019; Pabian et al., 2015; Wright, 2014).

2.3. Impact

Online geweld (in brede zin) heeft diepgaande gevolgen op zowel individueel als maatschappelijk niveau. Slachtoffers ervaren psychologische klachten zoals angst, stress, depressie en posttraumatische stress, vaak verergerd door de blijvende zichtbaarheid van online berichten. Deze psychologische schade kan ook leiden tot fysieke gezondheidsproblemen, zoals slaapproblemen en verhoogde bloeddruk. Daarnaast voelen slachtoffers zich sociaal geïsoleerd en verliezen zij het vertrouwen in digitale platforms en online gemeenschappen (Sincek et al 2017; Leukveldt et al 2018; Rathenau 2021; Castaño-Pulgarin et al., 2021). Vooral vrouwen, mensen van kleur en LHBTQIA+ personen worden vaker en zwaarder getroffen, mede door intersectionele vormen van discriminatie. Op maatschappelijk niveau ondermijnt online geweld het publieke debat en het vertrouwen in democratische processen.

Door haat en intimidatie trekken mensen zich terug uit het publieke discours, wat leidt tot zelfcensuur en een verschuiving naar extremere standpunten (Castaño-Pulgarin et al., 2021). Sociale media-algoritmes versterken dit effect door polariserende content te belonen met meer zichtbaarheid. Het gebrek aan effectieve handhaving van online misdrijven draagt bij aan het gevoel van straffeloosheid en wantrouwen in instituties (Rathenau 2021; Eddebo 2022; Castaño-Pulgarin et al., 2021; Leukveldt et al 2018)

Hoewel de literatuur belangrijke inzichten biedt, blijven er ook aanzienlijke kennishiaten bestaan. Eerder werd al aangegeven dat bestaand onderzoek niet gericht is op de drie vormen waar we in dit onderzoek op focussen. Daarnaast is er nog weinig bekend over de langetermijneffecten van online geweld op de mentale en fysieke gezondheid, met name bij jongeren. Ook is er beperkt inzicht in de effectiviteit van bestaande interventies en hulpverlening: het is onduidelijk welke vormen van ondersteuning het meest effectief zijn voor verschillende doelgroepen. Daarnaast is de precieze rol van algoritmes in het versterken van online geweld nog onvoldoende empirisch onderbouwd. Tot slot ontbreken internationale vergelijkingen vaak, waardoor het lastig is om te leren van succesvolle aanpakken in andere landen. Deze kennishiaten benadrukken de noodzaak voor verder onderzoek en beleidsontwikkeling.

2.4. Interventies en beleidsmaatregelen

Er is weinig bekend over de effectiviteit van meldingssystemen en de mate waarin slachtoffers daadwerkelijk geholpen worden. Verder ontbreekt diepgaand inzicht in de ervaringen van specifieke groepen, zoals mensen met een beperking of mensen met een migratieachtergrond. Daarbij is er nog onvoldoende kennis over de langetermijneffecten van online geweld, zowel op individueel als maatschappelijk niveau, en over de rol van algoritmes in het versterken van haat en geweld online. Deze kennishiaten maken het moeilijk om gericht beleid te ontwikkelen en passende interventies te ontwerpen.

Als we kijken naar interventies, dan worden die toegepast op meerdere niveaus: *individueel, gemeenschapsgericht, beleidsmatig en technologisch*. Dit hoofdstuk geeft een beknopte samenvatting van de interventies op deze niveaus die in de (inter)nationale literatuur worden omschreven. Voor meer informatie over mogelijke interventies en beleidsmaatregelen zie Onderzoek naar strafbare vormen van online geweld - Verwey-Jonker Instituut en het empirische hoofdstuk 4.

2.4.1. Individueel niveau

Interventies op individueel niveau kunnen zich richten op het vergroten van bewustzijn, digitale weerbaarheid en het stimuleren van actief ingrijpen door omstanders. Jongeren zijn vaak passief als zij online haat of intimidatie waarnemen, mede uit angst voor represailles of het gevoel dat hun tussenkomst weinig verschil maakt. Weerbaarheidstrainingen kunnen hen ondersteunen bij het herkennen van schadelijk gedrag en hen handvaten bieden om effectief op te treden (Obermaier, 2024). Ook blijkt uit onderzoek dat jongeren nauwelijks op zoek gaan naar hulp bij online geweld, vaak uit schaamte of onbekendheid met het aanbod. Hulp wordt meestal gezocht bij informele bronnen (zoals familie en vrienden), die door jongeren als waardevol worden ervaren. Jongeren geven de voorkeur aan informatie via sociale media, onderwijs en chatgebaseerde hulpkanalen (Rosielle et al., 2024).

2.4.2. Gemeenschapsniveau

Online haatzaaien en bedreigingen vormen niet alleen een aanval op individuele slachtoffers, maar ook een bedreiging voor de sociale cohesie en democratische waarden in de samenleving. Haatzaaiende uitlatingen hebben een verstikkend effect op het publieke debat: ze kunnen mensen ervan weerhouden deel te nemen aan discussies en ondermijnen zo een open uitwisseling van ideeën. Algoritmes en echokamers versterken dit proces. Er is daarom behoefte aan digitale omgevingen waar diversiteit van stemmen wordt gestimuleerd en haatspraak actief wordt tegengegaan (Pérez-Escolar & Noguera-Vivo, 2022).

Cruciaal hierbij is samenwerking in lokale en (online) gemeenschappen om negatieve dynamieken tijdig te signaleren en in te grijpen. (Greijdanus et al., 2023) benadrukken dat het voorkomen en verminderen van online aangejaagd geweld vraagt om een gecoördineerde inzet van diverse netwerkpartners, zoals gemeentelijke instanties, politie, jongerenwerkers, scholen en mediaplatforms. Door deze actoren samen te laten werken kan men escalaties sneller opsporen en tegengaan (Greijdanus et al., 2023). Dergelijke samenwerkingsinitiatieven kunnen zorgen voor een breder maatschappelijk draagvlak voor de norm dat online geweld onacceptabel is, wat bijdraagt aan een veerkrachtiger online klimaat.

2.4.3. Beleids- en wetgevingsniveau

Hoewel nieuwe Europese wetgeving zoals de Digital Services Act (DSA) belangrijke stappen zet richting verantwoordelijkheid bij platformen, blijft effectieve handhaving lastig. Nationale overheden missen vaak kennis, capaciteit en juridische instrumenten om daadkrachtig op te treden. Ook samenwerking tussen overheden en techbedrijven is vaak nog beperkt. Leeftijdsgrenzen voor sociale media blijken in de praktijk moeilijk te handhaven. Er is behoefte aan meer bestuurlijke middelen zoals digitale gebiedsverboden en aan grensoverschrijdende samenwerking om handhaving internationaal te versterken.

2.4.4. Technologische niveau

Techbedrijven spelen een sleutelrol in de aanpak van online geweld. Publiek-private samenwerking is cruciaal voor het delen van signalen over dreigende escalaties en het verbeteren van meldprocedures. Technologische oplossingen zoals automatische detectie en contentmoderatie zijn waardevol, maar niet zonder beperkingen. Context blijft moeilijk te beoordelen, waardoor onbedoelde fouten ontstaan. De ‘redirect method’, waarbij gebruikers worden omgeleid naar de-escalerende content, toont veelbelovende resultaten, maar de effectiviteit op de lange termijn is nog onvoldoende bewezen. Ook het idee van een ‘online paspoort’ (digitale identiteit) is controversieel, omdat het vragen oproept rond privacy en vrijheid van meningsuiting.

3 Aard, dynamiek en omvang van online geweld

In dit hoofdstuk beschrijven we op basis van de verschillende onderzoeksactiviteiten (ISRD-data, interviews met praktijk- en wetenschappelijke experts, interviews met plegers en slachtoffers, twee focusgroepen en de online reflectiesessie) de bevindingen. Hierin gaan we in op de aard (3.1), dynamiek (3.2) en omvang (3.3 en 3.4) van online geweld.

3.1. Aard: geen eenduidig fenomeen

Uit de interviews met betrokken burgers komt een breed en veelzijdig beeld naar voren van de aard van online geweld tussen burgers onderling. De casussen variëren in intensiteit, motieven, contexten en platformen.

> *“Alles hangt met alles samen, er zitten heel veel kanten aan.”*
>
> Deelnemer focusgroep

Duidelijk is dat online geweld zich niet *manifesteert als een eenduidig fenomeen*, maar als verzameling praktijken met verschillende doelen, dynamieken en gevolgen.

Patronen

Ondanks dat online geweld tussen burgers onderling veelzijdig is, is hierin wel een aantal patronen te onderscheiden. Een belangrijk deel van het online geweld in de interviews was gericht op iemands **identiteit of hun (publieke) profilering in het maatschappelijk debat**. Denk hierbij aan mensen uit de lhbtiqa+-gemeenschap, activisten, of burgers die deelnemen aan protestbewegingen. Zo werd een respondent, die in transitie is gegaan, bedreigd via Threads, omdat ze een transvlaggetje in haar biografie had staan. Bij dezelfde respondent was een persoon systematisch haar vriendinnengroep afgegaan via Instagram om hen te bedreigen na een eerdere confrontatie op een demonstratie. De bedreiging leek vooral gericht op haar identiteit als trans-persoon. Ook andere geïnterviewden kregen te maken met doelgerichte aanvallen. Een man die zich uitsprak voor homorechten werd door grote groepen jongeren bedreigd en geïntimideerd. De haatspraak bleef niet alleen beperkt tot het online domein, er werden ook pogingen gedaan om eigendommen van hem te vernielen. Deze bevinding wordt ook erkend door de verschillende experts in de focusgroepen en online sessie. Eerder liet de literatuurstudie ook zien dat bepaalde groepen op basis van bijvoorbeeld hun etniciteit, seksuele geaardheid en gender vaker worden geconfronteerd met online geweld zoals bedreiging, intimidatie en uitsluiting (Online geweld tussen burgers onderling).

Ook de **maatschappelijke context** kan een aanjager zijn voor het uitten van online geweld. Een respondent die we hebben gesproken was bedreigd via een anonieme e-mail, ondertekend door ‘rechts verzet’, na deelname aan een actie van een “linkse/progressieve” protestgroep. Ook een boerin was slachtoffer van lasterlijke filmpjes over haar bedrijf, geplaatst door een dierenrechtenactivist. Deze voorbeelden tonen hoe persoonlijke kenmerken (zoals genderidentiteit of seksuele oriëntatie) of maatschappelijke stellingname of positie (zoals activisme of een bepaald beroep) aanleiding kunnen vormen voor gerichte online intimidatie.

Daarnaast horen we van verschillende respondenten terug dat de coronatijd ook veel invloed heeft gehad. Er lijkt sprake van een verharding van het maatschappelijke debat en een toename van complottheorieën en wantrouwen richting overheid en wetenschapsinstanties, wat zich ook uit in online geweld. Verschillende deelnemers aan ons onderzoek vertelden over hun ervaringen met online intimidatie, vooral in het licht van corona-gerelateerde complottheorieën. Een respondent beschreef hoe hij geconfronteerd werd met online dreigementen binnen groepen die dergelijke complottheorieën deelden:

> *“Dan heb je het echt over de extreme complottheorieën, zoals ‘5G veroorzaakt corona’. Dat zijn van die radicale ideeën die ook de technologie raken. Daar kon ik in het begin nog over meepraten… Sommige mensen gingen er toen nog in mee, maar je werd al snel uitgemaakt voor overheidstrol, of er werd gezegd dat je voor iemand werkt. Je merkte ook dat het ‘wij-tegen-hen’-gevoel heel sterk werd. Je belandde al gauw in twee kampen. Ik had daar liever wat meer afstand van gehouden, maar uiteindelijk word je toch een bepaalde kant op geduwd. Zo ontstonden die gesprekken. Op een gegeven moment heb ik besloten, zeker op Twitter, om dat anoniem te gaan doen, omdat ik me er niet prettig bij voelde. Er kwamen ook echt bedreigingen binnen zoals: ‘we zoeken je wel op’, ‘we komen wel langs’, of ‘als ik je tegenkom…’. Dat soort dingen.”* Burger

Casus 1 Telegramgroepen

Tom had jarenlang een rustige online aanwezigheid en beperkte zich tot het delen van informatie over zijn hobby's. Toen de coronacrisis uitbrak en hij door de lockdowns meer tijd had, raakte hij betrokken bij online discussies over complottheorieën, met name over 5G en vaccins. Aanvankelijk probeerde hij misinformatie te ontkrachten onder zijn eigen naam, maar nadat hij meer online aandacht kreeg, besloot hij anoniem verder te gaan. Omdat hij zich anoniem op het internet begaf en zich goed kon beveiligen, was hij niet bang voor consequenties. Zijn activiteit verschoof van Twitter naar Telegram, waar hij zich aansloot bij een groep die complotten analyseerde en soms ook troll-acties uitvoerde. Daarnaast observeerde hij anoniem verschillende complotgroepen.

Op een gegeven moment werd een online kennis van hem bedreigd. Vanwege zijn technische vaardigheden werd Tom gevraagd informatie te zoeken over de dader en ontdekte hij binnen enkele uren diens woonadres. Vervolgens stuurde zijn kennis een foto van de voordeur van de dader naar de dader, met een subtiele dreiging.

Daarnaast belandde Tom in extremere Telegramgroepen, waaronder een besloten extreemrechtse groep. Daar zag hij hoe nazisympathieën openlijk gedeeld werden. Nadat er in de groep een openbare activiteit werd georganiseerd, besloot Tom deze informatie op zijn anonieme Twitteraccount te delen. Later werd hij door een bekende benaderd: door zijn actie waren twee onschuldige mensen onterecht aangezien voor hem en bedreigd door leden van de extreemrechtse groep. Dit was een schok. "Ik dacht dat ik alleen maar mensenonline te kijk zette, maar dit ging veel verder."

Deze ervaring deed Tom beseffen hoe groot de impact van online acties kon zijn. Hij stopte met doxing en het openbaar maken van informatie over individuen, maar bleef anoniem online aanwezig als observator. Hij maakte meldingen bij de politie, wat soms leidde tot stopgesprekken of zelfs arrestaties. Zijn kijk op complotdenkers veranderde ook: waar hij hen eerst enkel ridiculiseerde, kreeg hij later meer begrip voor de sociale en economische omstandigheden die mensen vatbaar maken voor complottheorieën.

Ook andere respondenten benoemden dat het maatschappelijke debat sinds de coronapandemie is verhard. Mensen voelen zich sneller gerechtvaardigd om anderen online aan te vallen, zeker wanneer iemand zich publiekelijk uitspreekt over gevoelige onderwerpen. Zoals een respondent het verwoordde:

> *"Sinds corona is het veel harder geworden. Mensen voelen zich sneller gerechtvaardigd om anderen online af te branden, zeker als je je ergens over uitspreekt."* Praktijk expert

De bevindingen sluiten aan bij bredere inzichten uit dit onderzoek. Zo laat de literatuurstudie zien dat maatschappelijke polarisatie, het publieke debat over thema's zoals klimaat, corona of gender, en het groeiende wantrouwen richting instituties online geweld tussen burgers onderling kunnen aanwakkeren. Daarbij blijft het niet bij verbaal geweld; ook doxing (het online verspreiden van privégegevens), intimidatie en lastercampagnes komen geregeld voor. Experts zien bovendien duidelijke parallellen met andere vormen van online geweld, zoals seksuele intimidatie en misogynie. Met name vrouwen die zich publiekelijk uitspreken over maatschappelijke thema's zijn relatief vaak het doelwit van seksistisch getinte bedreigingen of vernederingen. Een expert verwoordde dit als volgt:

> *"Het patroon is herkenbaar: zodra iemand — en zeker een vrouw — zich uitspreekt over maatschappelijke kwesties, wordt het ineens heel persoonlijk gemaakt. Dan gaat het opeens over uiterlijk, seksualiteit of zelfs familie."* Praktijk expert

Online geweld staat dan ook zelden op zichzelf, maar is verweven met bredere maatschappelijke spanningen. Discussies over thema's zoals corona, klimaat, stikstof, immigratie of complottheorieën kunnen zowel online als offline escaleren, waarbij sociale media een katalyserende rol spelen in het snel verspreiden van haat, intimidatie, bedreigingen (en misinformatie). Daarbij is er sprake van een duidelijke wisselwerking: spanningen in de online wereld voeden de spanningen in de offline samenleving — en andersom. Wat zich online afspeelt, blijft niet beperkt tot het scherm, maar werkt door in de dagelijkse omgang en het maatschappelijke klimaat.

> *"Ik weet dat het buiten de scope van dit onderzoek valt, maar tijdens de coronatijd zagen we natuurlijk al dat er een verschuiving was van offline naar online gedrag. Die trend zag je deels ook terug bij seksuele misdrijven."* Wetenschappelijk expert

Ook de context van gamen en de invloed van de manosphere[10] komen tijdens de gesprekken vaak terug:

> *“ Ik heb gewoon het gevoel dat ik heel sterk ben geconfronteerd met een soort manosphere, waarin het volkomen normaal is om extreem gedetailleerde doodsbedreigingen te uiten. Dat wordt ook nog eens aangemoedigd en vaak geliket.”* Burger

In het verlengde hiervan is in meerdere casussen naar voren gekomen dat online geweld doelbewust en strategisch wordt ingezet om anderen (met name *bepaalde groepen mensen*) te beschadigen, af te schrikken en/of uit te sluiten. Het voorbeeld van bedreigingen richting activisten is al genoemd, als voorbeeld om deelnemers van demonstraties af te schrikken om deel te nemen. Ook het plaatsen van een filmpje van dieren van een boer kan als strategische actie worden gezien om eigen ideeën te verspreiden. Of bijvoorbeeld de casus van hierboven (casus 1 Telegramgroepen) waarin een man vertelde hoe hij tijdens de coronatijd in zijn vrije tijd bij Telegramgroepen infiltreerde en samen met anderen complotdenkers probeerde te “exposen” door gevoelige informatie te delen en hen op ludieke manieren belachelijk te maken. Daarnaast belandde hij in een extreemrechtse groep, waar hij uiteindelijk ook informatie over heeft gedeeld. Eveneens had hij een vriend geholpen om een man te intimideren, nadat deze man zijn vriend had bedreigd. Deze voorbeelden geven aan dat digitale middelen als strategisch instrument ingezet kunnen worden om een ander te schaden, te laten schrikken of onder druk te zetten.

Naast ideologisch of strategisch gebruikt geweld, komt online geweld ook voort uit persoonlijke relaties en conflicten. In verschillende casussen startte het geweld offline of in een sociale context, om vervolgens digitaal te escaleren. Een moeder vertelde hoe haar tienerdochter werd gepest, wat vervolgens via snapchat explodeerde waarbij ze werd bedreigd met verkrachting, en haar vriendje bedreigd werd met de dood. De moeder vertelt:

> *“ De kinderen waren volop aan het schelden. Mijn dochter vroeg of ze normaal konden doen. Toen kreeg ze te horen: “ik hoop dat je verkracht wordt”. En dan kreeg ze een PM (personal message): “en anders doe ik het”. Degene die de screenshots van de dreigementen maakte, kreeg vervolgens: “ik hoop dat je met één oog open slaapt.” En er werden foto’s gestuurd van geweren en dat soort dingen.”* Burger

10 De manosphere is een verzamelnaam voor een online netwerk van websites, fora, sociale media-accounts en podcasts waarin mannelijkheid en de positie van mannen centraal staan. Deze beweging is actief op platformen zoals YouTube, TikTok, Reddit en X. Binnen de manosphere delen mannen veel content over onderwerpen als gezondheid, zelfontwikkeling, daten, mannelijkheid, identiteit en carrière. Deze content lijkt op het eerste gezicht positief. Maar hieronder liggen vaak schadelijke overtuigingen. Wie zich meer verdiept in de manosphere, wordt vaker geconfronteerd met vijandige ideeën over vrouwen, feminisme en maatschappelijke gelijkheid. Vrouwenhaat, of misogynie, komt in verschillende vormen voor (De impact van de manosphere | Nederlands Jeugdinstituut).

Casus 2 School

Het begon met een opmerking in een groep van snapchat. Haar dochter werd gepest en vroeg of klasgenoten alsjeblieft normaal wilden doen. Het escaleerde. Wat volgde waren bedreigingen ‘dat ze zou worden verkracht’ en foto’s van wapens. Haar vriendje, die screenshots maakte van de berichten, werd ook bedreigd. Toen de moeder de screenshots zag, kon ze nauwelijks geloven dat zulke teksten onder jonge tieners rondgingen. Ze stapte naar school en stuurde alles door. En het antwoord was: “misschien had haar dochter zelf iets gezegd wat dit had uitgelokt.” Het voelde alsof het slachtoffer verantwoordelijk werd gehouden. De jongen die haar bedreigd had, werd de volgende dag naast haar geplaatst in de klas. Ze verstijfde. Thuis kwamen de paniekaanvallen. Ze durfde niet meer terug.

De moeder schakelde alles in: school, politie, jeugdhulp. De communicatie met de politie verliep stroperig, de jeugdagent kwam bij de eerste afspraak niet opdagen. Voor hulpverlening stond ze op lange wachtlijsten. Er kwam uiteindelijk een politieagent in de klas, maar dat maakte weinig indruk op de leerlingen. Een dag later kreeg haar dochter een anoniem sms›je: “Waarom heb je de politie erbij gehaald? Die verkrachting was maar een grapje hoor”. De leerplichtambtenaar wilde dat het slachtoffer weer naar school ging, maar de moeder hield haar thuis. Als haar dochter naar school ging, zag ze langzaam het licht uit haar ogen verdwijnen.

Na maanden kreeg ze plek op een school voor speciaal onderwijs. Kleiner, rustiger, veilig. Niet vanwege een diagnose, maar omdat het nergens anders meer ging. Haar moeder heeft niet de illusie dat ouders grip hebben op het onlinegedrag van hun kinderen. Wat ze probeert, is een open gesprek te voeren en laten weten dat haar dochter altijd bij haar terecht kan. “Als ze niet bij mij was gekomen zodat ik kon ingrijpen,” zei ze, “dan had ze er misschien zelf een einde aan gemaakt.”

In een ander geval leidde het afwijzen van een Tindercontact tot systematisch lastigvallen en het doxen van gegevens via een Telegramkanaal, wat leidde tot nog meer ongewenste en anonieme berichtjes en intimidatie.

Bij jongeren blijkt ook de groepsdruk- en dynamiek van invloed op het ontstaan en escaleren van online geweldsituaties.

> *Ik weet niet of ze dat dan echt heel bewust doen. Ik denk dat het gewoon zo werkt. Op het moment dat je die jongens apart van elkaar neerzet, zijn ze veel meer hun eigen individu en niet de rol die zij hebben binnen de groep. Je ziet dat zodra ze weer in die groep komen, ze zich anders gaan gedragen. Er zijn echt voorbeelden van jongens die bij elkaar op school zitten, maar uit rivaliserende bendes komen. Dat gaat altijd goed tot het moment dat hun vrienden erbij komen dan loopt het uit de hand. Hetzelfde geldt als ze elkaar met z’n tweeën tegenkomen in het openbaar vervoer. Dan kijken ze misschien even boos, maken een foto, zetten die op Instagram en beledigen of bedreigen elkaar daar. Maar het blijft dan bij woorden. Zodra er anderen bij zijn, escaleert het wél. Dat komt gewoon door die groepsdynamiek.”* Praktijk expert

In het volgende citaat komt een respondent aan het woord die vertelt hoe tijdens het gamen het gedrag steeds wat meer escaleerde.

> *Het werd steeds erger. Je neemt het spel ook serieuzer als je beter wordt en dan bouwt het zich op. Eerst speel je veel met vrienden en op een gegeven moment ga je het ook weleens alleen doen. Nou ja, dan is het eigenlijk alleen maar erger. Want als je met onbekende mensen bent, ben je toch sneller geneigd om — nou ja, het klinkt stom — te gaan schelden of iets dergelijks.”.* Burger

3.2. Dynamiek: hybride aard

Een belangrijk kenmerk van online geweld is *hybride aard*.

“ Die verwevenheid tussen online en offline; die werelden zijn er eigenlijk niet meer. Dat is gewoon één. Je kunt nauwelijks meer over verwevenheid spreken.”

Deelnemer focusgroep

“ Online, offline, het gaat door elkaar heen.”

Praktijk expert

Intimidatie, haatspraak en bedreiging lopen vaak in elkaar over. En haatspraak en desinformatie gaan regelmatig gepaard met het aanzetten tot geweld. In specifieke subculturen, zoals de drill-rapscene, zien we dat online posts – vaak op Snapchat of Instagram – leiden tot echte, fysieke confrontaties. Beelden van vechtpartijen of rellen worden online gedeeld en zorgen juist voor verdere escalatie. Dit wijst op een zorgwekkende overlap tussen online opruiing en offline geweld.

> *“Op Snapchat wordt veel gedeeld. Instagram wordt nog steeds veel gebruikt— zeker als het gaat om drillrap. Juist omdat je dat natuurlijk in de openheid wilt doen. Je wil dat de hele wereld het kan zien en niet alleen die specifieke groep waarmee je het deelt. Het hangt ook van het type dreigement af. We zien dat ze elkaar online uitdagen. Drillrappers uiten soms directe bedreigingen. Maar we zien ook steeds vaker vernederingsvideo’s: video’s waarin jongeren sorry moeten zeggen, op hun knieën moeten zitten of in elkaar geslagen worden.”* Praktijk expert

Veelal is er sprake van een onlosmakelijke relatie tussen online en offline geweld die elkaar over en weer opvolgen:

> *In mijn beleving gaat het dwars door elkaar heen. Het is vaak zoiets als: 'Oh, staat je gezicht me niet aan? Dan maken we een groepje op Snapchat en gaan we lelijke foto's van elkaar maken.' Dat is vaak wel echt het begin. Die besloten groep wordt alleen maar groter en groter en uiteindelijk krijg je pestgedrag. Ze kunnen iemand echt heel klein maken. Daar hebben ze zelfs Snapgroepen voor: als iemand in elkaar wordt geslagen, maken ze er een filmpje van en gooien dat erin. En dan staat het hele centrum vol met jongeren. Ik kan me niet voorstellen dat dit alleen hier gebeurt… Maar het kan ook zijn dat er eerder ruzie is geweest. Soms is het gewoon een woordenwisseling op school. Of een nare opmerking onder een reactie op Snapchat. Het gaat echt alle kanten op — helaas".* Deelnemer focusgroep

Ook bij incidenten blijkt bij traditionele criminaliteit dat de oorsprong vaak in de online wereld ligt:

> *Wij zien het best wel veel. Vaak missen we natuurlijk ook veel informatie online en die komt dan in de offline wereld bij ons terecht. Bijvoorbeeld bij de McDonald's in het centrum in één keer staan er vijftig gasten op fatbikes tegen elkaar aan te rossen en dan blijkt dus dat die oorsprong natuurlijk weer ligt in de online wereld. Of de ene school tegen de andere school, dan hebben ze onderling weer ergens in een bepaalde groep op een social media kanaal afgesproken. En dan doordat je dat niet ziet en niet meemaakt, krijg je dus alleen het fysieke gedeelte mee, maar de oorsprong ligt in het online gebeuren."* Deelnemer focusgroep

> *Er is online via de Playstation afgelopen zaterdag ruzie geweest onderling bij jongeren. Uiteindelijk is daar een actie op gekomen. Ze hebben als reactie (op de ruzie online) iemand rondom de woning lastiggevallen. Anderhalve dag geleden hebben ze besloten om twee Cobra's via Snap, ik vermoed via Snap,] te kopen wat heel makkelijk is. Voor een paar euro heb je een paar cobra's, wat plakband bij de Action en een aansteker bij de Action en ze hebben dat gegooid. En uiteindelijk gaat het nu weer online verder. Dus het komt nu in de school. Het lastige aan het hele verhaal is, is dat dader en slachtoffers gewoon bij elkaar op school zitten en ze zien geen oorzaak en gevolg."* Praktijk expert

De wisselwerking tussen online en offline geweld manifesteert zich op verschillende manieren binnen diverse vormen van online geweld. Een duidelijk voorbeeld is het vastleggen en verspreiden van fysiek geweld, zoals gevechten of rellen. Zulke beelden worden vaak gedeeld via sociale media als Snapchat en Instagram. Dit vergroot het bereik van het geweld en kan leiden tot verdere escalatie of vergeldingsacties. Zo circuleren er bijvoorbeeld filmpjes van straatgevechten die later aanleiding blijken voor nieuwe confrontaties tussen betrokken groepen. Online zichtbaarheid fungeert hier als *brandstof voor herhaald fysiek geweld*. Ook bij intimidatie en bedreiging is deze dynamiek zichtbaar: handelingen die offline plaatsvinden—zoals het opzoeken en intimideren van personen—worden regelmatig doelbewust online gedeeld. Dit vergroot niet alleen de zichtbaarheid, maar versterkt ook de *impact en dreiging van het geweld*. Bij discriminatie en haatspraak is eveneens sprake van een wederzijdse relatie. Online haatdragende uitingen kunnen voortkomen uit fysieke gebeurtenissen, maar ook andersom aanleiding vormen voor offline spanningen of incidenten. Een concreet voorbeeld hiervan zijn de Southport rellen, waarbij online haatspraak bijdroeg aan het ontstaan van grootschalige rellen. Online haat creëert zo een klimaat waarin discriminatie en agressie in de fysieke wereld als gerechtvaardigd worden ervaren, met het risico op toenemende maatschappelijke polarisatie.

> *"De online wereld komt ook naar de offline wereld. En dat gaat super, super, super snel. Ik zit er nu middenin, dus ik hoor van alles aan alle kanten."*
>
> Praktijk expert

3.2.1. Glijdende schaal en escalatie

Naast dat online geweld zich lastig laat afbakenen en er sprake is van hybride vormen geweld waarin offline en online elkaar over en weer opvolgen, is er bij online geweld ook sprake van een *'glijdende schaal'*. Daarmee wordt bedoeld dat er verschillende gradaties zijn die elkaar ook in hoog tempo kunnen opvolgen.

> *Die schaal is gewoon heel erg vaag of het gebied is heel grijs. Je kan zo van haatzaaien tot discriminatie gaan. Het is niet heel goed af te bakenen, omdat het overal mee te maken heeft".* Deelnemer focusgroep

Veel berichten die online geplaatst worden zijn niet direct strafbaar, maar kunnen wel diep beledigend of intimiderend zijn.

> *Wat je wel ziet is dat heel veel niet illegaal is, maar wel schadelijk. En dus niet lawful, maar harmful".* Deelnemer focusgroep

De impact op slachtoffers is aanzienlijk: ook niet-strafbare berichten kunnen leiden tot gevoelens van angst, onveiligheid en sociale uitsluiting (zie verder hoofdstuk 4 over de impact). In sommige gevallen ontstaat er een 'glijdende schaal' waarbij 'hufterig gedrag' snel omslaat in serieus strafbaar gedrag.

> *Mensen denken natuurlijk online makkelijk mensen te bedreigen of erop te reageren. Dat heb ik zelf wel eens gehoord: "Het is maar online." Zij hebben niet altijd het besef, dat een bedreiging op social media een strafbaar feit is".* Praktijk expert

Daarbij kunnen de ogenschijnlijk onschuldige uitingen snel escaleren naar strafbare gedragingen zoals bedreiging of opruiing. Iets wat eerst nog heel onzichtbaar is, kan in 'no tempo' uitmonden in een gewelddadige situaties waarbij soms duizenden mensen betrokken zin.

3.2.2. Anoniem en bekenden

Online geweld speelt zich niet altijd in de anonieme sfeer af. Dat komt zowel naar voren uit de interviews met plegers en slachtoffers maar ook de gesprekken met experts. In familieverbanden, romantische relaties of jeugdgroepen worden dreigingen en pesterijen vaak door bekende personen geuit. Tegelijkertijd biedt anonimiteit voor anderen juist een dekmantel om ongehinderd en grootschalig digitaal geweld te plegen.

> *Het komt ook een beetje omdat ik die mensen niet ken. Ik heb niet echt iemand in mijn omgeving of zo aangevallen. Het voelt een beetje afstandelijk dan zeg maar. Het voelt een beetje ver weg."* Burger

> *Die zitten graag in hun eigen bubbel op de verschillende social media. En die komen dan ineens in een groep terecht op Telegram bijvoorbeeld, dat zag je gebeuren. Mensen die dus eigenlijk een kleine online wereld hadden, komen in een grote online wereld. En die hebben ineens geen filter. Want die denken van 'oh, niemand kent mij hier.' Of 'Oh, ik heb niks te vrezen.' En dat soort zaken. En het omgekeerde is waar. De meeste mensen hebben totaal geen idee hoe makkelijk mensen online te vinden zijn."* Burger

3.3. Omvang

Ondanks dat onze empirische bevindingen laten zien dat online geweld een groeiend maatschappelijk probleem is, ontbreken hiervoor nog harde cijfers en ontbreekt er een genuanceerd beeld van de omvang. Zo blijkt uit CBS onderzoek dat online discriminatie lijkt te zijn verdubbeld: van 2% in 2022 % naar 4 % in 2024. Via de sociale media analyse hebben we geprobeerd om grip te krijgen op de omvang van online geweld tussen Nederlandse burgers onderling. Alle experts en professionals zoals politie, maatschappelijke organisaties en academici (zie bijlage 2 voor de respondentenlijst) bevestigen dat cijfers over online geweld nog ontbreken.

> *Je ziet het heel veel. Ik zit zelf in de gemeentepolitiek en ook daar zien we het heel veel gebeuren. Ik denk dat er ook nog een dark number is in deze problematiek."* Deelnemer focusgroep

Online geweld wordt nog niet geregistreerd als aparte categorie. Daarbij zijn er *nauwelijks aangiftes/meldingen* van online geweld tussen burgers onderling (zie ook 5.1.1 waarin verschillende redenen worden beschreven). Ook bureau Halt krijgt nauwelijks meldingen van online geweld, ondanks dat ze zien dat scholen er bijvoorbeeld wel mee te maken krijgen. Dit krijgen we natuurlijk ook in het nieuws mee, waarbij scholen soms ook gesloten worden om de situatie veilig te houden/maken.

Weliswaar hebben we door middel van de ISRD data zicht gekregen op online geweld onder jongeren. Hieruit blijkt dat *12,9%* van de jongeren wel eens één of meerdere keren pleger is geweest en *26,2%* is daar wel eens één of meerdere keren slachtoffer van geweest:

- Bedreigd via sociale media (3,2%)
- Verstuurd, gedeeld of gerepost van intieme foto of video van jou (3,3%)
- Kwetsende berichten of reacties ontvangen op sociale media over etnische achtergrond, nationaliteit, religie, genderidentiteit, seksuele voorkeur, of een andere soortgelijke reden (7,6%)
- Online bedreigd met geweld of gewelddadig behandeld vanwege etnische achtergrond, nationaliteit, religie, genderidentiteit, seksuele voorkeur, of een andere soortgelijke reden (zie ook: Factsheet online geweld onder jongeren).

Ook hebben we met onze sociale media analyse zicht gekregen op de omvang van online geweld (zie 3.4). Andere cijfermatige onderbouwing van de omvang van online geweld ontbreekt helaas (zie ook hoofdstuk 6 bij de aanbevelingen).

Groeiend maatschappelijk probleem

Ondanks dat cijfers ontbreken, laten de kwalitatieve onderzoeksbevindingen zien dat het op grote schaal voorkomt. De interviews met wetenschappelijke en praktijk experts benadrukken dat online geweld een *groeiend maatschappelijk probleem* is. Professionals houden hun hart vast voor wat mogelijk nog komen gaat. Daarbij zien we ook nieuwe ontwikkelingen opspelen zoals nieuwe vormen van online geweld tussen burgers onderling. Nu wordt er bijvoorbeeld onder jongeren al veel gebruik gemaakt van gifjes in de bedreigingen en intimidatie via social media. Maar professionals maken zich zorgen als jongeren de wereld van deepfaken, porno duplicate, gaan ontdekken. Dan worden er straks geen gifjes meer verspreid, maar intimidatie en bedreiging gaat dan ook echt gebeuren door gebruik te maken van 'echte' filmpjes. Die filmpjes zijn niet van echt te onderscheiden.

> *"Ik hou mijn hart vast wanneer jongeren de apps ontdekken waarmee je op hele andere manieren mensen kan exporteren, weet je. Dat zijn nu de meeste meldingen. En de jeugd heeft hem nog niet helemaal, denk ik, ontdekt. Maar als dat komt dan denk ik echt wat gaan we daar dan weer tegen beginnen. Dus het voelt soms echt als een rat race van technologie versus waar mensen aankomen te zitten. En wat je er tegenover moet stellen de hele tijd om het te kunnen reguleren. Ik weet ook niet waar dat eindigt..."* Praktijk expert

Ook zien ze al signalen dat in gemeenten burgers zich gaan voordoen als 'jongerenwerkers' en dit gaat zelfs zo ver dat er nepaccounts worden aangemaakt met alle gevolgen van dien.

3.4. Social media analyse

Doordat er weinig zicht is op de aard en omvang van online geweld in bestaande data en literatuur, hebben wij in dit onderzoek (naast de analyse van de ISRD data, zie hierboven bij 3.2) een vernieuwende onderzoeksmethode gebruikt om meer zicht hierop te krijgen. Dit hebben we gedaan door het verrichten van een social media analyse. In paragraaf 1.3 Onderzoeksaanpak en bijlage 1 staat meer over wat deze methode inhoudt. Hieronder beschrijven we samenvattend de resultaten van de sociale media analyse. In bijlage 1 is de uitgebreide beschrijving terug te vinden waarin we eerst de resultaten voor *Nu.nl* berichten die onder artikelen worden geplaatst door (vaak anonieme) gebruikers (NU - Het laatste nieuws het eerst op NU.nl) beschrijven en vervolgens de resultaten per zoekterm van X. Zoals eerder beschreven hebben we gebruik gemaakt van de lexicons van The Weaponized Word. Hierbij nog een kort overzicht van de lexicons:

1. Discriminerende woorden: woorden die specifiek gericht zijn op de ontvanger zijn nationaliteit, etniciteit, religie, gender, seksuele oriëntatie, beperking en/of chronische ziekte of sociaaleconomische klasse.
2. Beledigende woorden: woorden die beledigend worden gebruikt, maar niet direct betrekking hebben op de sociale identiteit van de ontvanger.
3. Bedreigende woorden: woorden die (direct of indirect) dreigingen of geweld impliceren.
4. Watch words: signalerende woorden die samenhangen met haatspraak, of beledigend of dreigend taalgebruik. In de tekst hieronder zullen we spreken van 'risicovolle' woorden.

3.4.1. Resultaten per zoekterm

Platform Nu.nl

- We hebben in totaal 37.061 reacties geanalyseerd die zijn geplaatst op Nu.nl. Deze reacties waren geplaatst onder nieuwsartikelen. Deze nieuwsartikelen gingen over verschillende thema's. Iets meer dan 77% van de reacties onder deze nieuwsartikelen was door de moderatie goedgekeurd. Dat betekent dat een kwart van de reacties als ongepast werd gemarkeerd en verwijderd.
- Toch zijn er duidelijke verschillen tussen de thematische categorieën. Zo blijkt dat in de categorie 'Buitenland' meer dan een derde van de reacties werd afgekeurd. Ook in de categorieën 'LHBTIAQ+' en 'Transgender personen' zien we relatief veel afkeuringen: bijna een derde van de reacties werd hier verwijderd. In de reflectie kijken we hier meer uitgebreid op terug.
- Wanneer we specifiek kijken naar het taalgebruik, dan zien we dat discriminerend, beledigend of bedreigend taalgebruik vaker in de afgekeurde reacties terugkomt dan in de geaccepteerde reacties. Discriminerend taalgebruik kwam het vaakst voor in de reacties op Nu.nl: in goedgekeurde reacties komt dit soort taal in 1,9% van de gevallen voor, terwijl dat bij afgekeurde reacties 4,5% is.
- Verder zagen we dat discriminerende woorden het vaakst voorkomen in reacties onder artikelen over LHBTQIA+ onderwerpen. In maar liefst 35,6% van deze reacties werd een discriminerend woord aangetroffen.

Platform X

- Voor het tweede deel van de analyse hebben we gekeken naar publieke posts op X. We verzamelden grote hoeveelheden berichten rond vier zoektermen die vaak in maatschappelijke discussies opduiken: 'Moslim', 'Fascist', 'Wappie' en 'Tokkie'.
- Voor de term 'Moslim' verzamelden we 121.645 posts. Binnen deze posts komt dat bedreigende taal hier het vaakst voor. In bijna 8% van de berichten werd een bedreigend woord gebruikt. Daarnaast kwam discriminerend, bedreigend en beledigend taalgebruik vaker voor bij deze zoekterm dan bij de andere zoektermen.
- Een specifiek voorbeeld hiervan is het woord 'tuig', dat opvallend vaak samen met 'Moslim' werd gebruikt. In meer dan een derde van de berichten waarin 'Moslim' voorkwam, werd ook 'tuig' genoemd.
- Bij de zoekterm 'Fascist' vonden we een ander patroon. Hier kwamen discriminerende woorden het vaakst voor, gevolgd door bedreigende termen. In totaal bevatte 3,5% van de berichten discriminerende taal.
- Ook bij de termen 'Wappie' en 'Tokkie' zagen we het vaakst discriminerende woorden terug in de berichten. Woorden als 'idioten', 'idioot', 'sukkel', 'achterlijke' en 'achterlijk' kwamen opvallend vaak voor.

3.4.2. Reflectie resultaten sociale media analyse

In dit onderzoek hebben we met behulp van Natural Language Processing (NLP) tienduizenden reacties onder Nu.nl-artikelen en posts op X geanalyseerd op potentieel discriminerend, beledigend en onbeschoft taalgebruik. We verzamelden 37.061 reacties onder Nu.nl-berichten en 236.230 posts op X met de zoektermen "Moslim", "Fascist", "Wappie" en "Tokkie" in de periode juli–december 2024. Bij Nu.nl groepeerden we reacties bovendien in zes thematische clusters om te kunnen vergelijken hoe moderatie of schadelijke taal per onderwerp samenhangen. Ook hebben we gebruik gemaakt van de lexicons van The Weaponized Word om te kijken in welke mate discriminerende, beledigende, bedreigende en risicovolle woorden ("watchwords") in de reacties voorkomen. Voor X maakten we per zoekterm verkennende wordclouds per zoekterm en identificeerden we de frequentie van de woorden uit de lexicons in de posts.

Seksuele oriëntatie

Bij de Nu.nl-reacties zagen we dat reacties onder artikelen in de categorie 'Buitenland' het vaakst worden afgekeurd. Daarnaast zijn het aantal afgekeurde reacties onder artikelen over transgender personen en LHBTQIA+ opvallend. Dit suggereert dat *seksuele oriëntatie een vaak voorkomende grond van online geweld kan zijn, en dat personen die zich identificeren met de LHBTIAQ+ gemeenschap online een groter risico lopen.* Eerder onderzoek heeft aangetoond dat haat richting de LHBTQIA+ gemeenschap op sociale media is toegenomen in de afgelopen jaren (Van Dijk et al., 2023). Ook discriminatiecijfers van de politie laten zien dat seksuele oriëntatie een veel voorkomende grond is voor online discriminatie.

Discriminerende woorden

Verder zagen we dat discriminerende woorden het vaakst in de reacties voorkomen, ongeacht het onderwerp. Discriminerende woorden kwamen het vaakst voor in reacties onder artikelen die gingen over LHBTQIA+ gerelateerde onderwerpen. In meer dan een derde van de reacties kwamen discriminerende woorden voor.

Moslim

Op X zagen we bij de zoekterm "Moslim" in 7,9% van de posts bedreigende woorden terug en in 6,8% discriminerende termen, wat neerkomt op bijna één op de twaalf berichten met expliciet gewelddadige of uitsluitende taal. De frequentie van deze woorden was een stuk hoger dan bij de andere zoektermen: voor "Fascist" bevatte 3,5% van de posts discriminerende woorden en 2,8% bedreigingen; bij "Wappie" en "Tokkie" kwamen bijvoorbeeld beledigende woorden in 2,3% van de posts voor. Dit suggereert dat het woord 'Moslim' meer negatieve sentimenten oproept dan andere termen. Voor vervolgonderzoek is het dan ook interessant om relaterende woorden zoals 'Moslims', 'Islam' of 'Islamitisch' ook mee te nemen in analyses.

Tuig

In alle zoektermen is "tuig" de meest voorkomende discriminerende term: in berichten over "Moslim" zelfs in meer dan een derde van alle posts. In het licht van eerdere studies is dit gegeven niet heel opvallend. Zo blijkt uit onderzoek over *fake* Facebookgroepen dat moslims worden verbonden aan gewelddadigheid en gebrek aan empathie (Farkas et al., 2018). In een onderzoek in Nederland over moslims met verschillende etnische achtergronden is aangetoond dat 'criminelen' het meest voorkomende stereotype was voor Nederlanders met een Turkse of Marokkaanse achtergrond (Wiemers et al., 2024). Ook de politiek speelt een rol in de labeling van moslims als 'tuig' of criminelen, aangezien politici belangrijke bronnen kunnen zijn van sociale categorieën en labels voor de rest van de samenleving (Brubaker & Cooper, 2000). Roggebrand en van der Haar (2017) vonden dan ook dat 'tuig' een begrip was die politici gebruiken om Nederlandse jongeren met een Marokkaanse achtergrond te categoriseren.

Het effect van het onderwerp

Daarnaast zagen we dat er voor verschillende onderwerpen en zoektermen in verschillende mate problematisch taalgebruik werd gebruikt. Zo zien we dat posts over 'Moslims' relatief vaak bedreigende taal bevatten, terwijl bij 'Fascist' en 'Wappie/Tokkie' juist discriminerende of denigrerende termen domineren. Daarnaast zagen we dat onderwerpen rondom LHBTQIA+ vaker problematisch taalgebruik met zich meebrengt dan onderwerpen zoals immigratie of klimaat. Dit wijst erop dat het onderwerp zelf invloed heeft op de aard van de reacties die het oproept. Deze verschillen roepen de vraag op of mensen zich online consequent problematisch gedragen, ongeacht het onderwerp, of dat hun gedrag meer onderwerp-specifiek is. Met andere woorden: zijn er gebruikers die zich over meerdere thema's negatief uitlaten, of zijn er groepen die zich uitsluitend mengen in discussies over één specifiek onderwerp dat hen raakt of interesseert? Dit zou een interessante richting kunnen zijn voor vervolgonderzoek.

De toon van het taalgebruik

Hoewel de analyses waardevolle inzichten bieden in het gebruik van problematische taal op sociale media, zijn er ook beperkingen. Een belangrijk punt is dat we ons in dit onderzoek hebben gericht op de frequentie van woorden uit vooraf gedefinieerde lexicons. Dit geeft een indicatie van hoe vaak bepaalde termen voorkomen, maar zegt nog niets over het sentiment waarmee ze worden gebruikt. Een woord als 'idioot' kan bijvoorbeeld in een ironische, kritische of zelfreflectieve manier gebruikt worden. Zonder sentimentanalyse weten we niet of de uitingen daadwerkelijk negatief, neutraal of positief zijn. De verdieping die zich richt op classificatie – zie kopje tekstclassificatie - zou hier mogelijk meer licht op kunnen werpen, maar het zou ook waardevol zijn om in toekomstig onderzoek expliciet te kijken naar sentiment en toon van de berichten. Een voorbeeld hiervan is het onderzoek van de Groene Amsterdammer en de Data School van de Universiteit Utrecht (Van Dijk et al., 2023).

Hoe wijdverspreid is het taalgebruik?

Daarnaast is het belangrijk om te erkennen dat we in deze analyse niet hebben gekeken naar de verspreiding van problematisch taalgebruik onder gebruikers. We weten hoe vaak bepaalde woorden voorkomen in berichten, maar niet hoeveel verschillende gebruikers deze woorden gebruiken. Dit maakt het lastig om uitspraken te doen over de schaal van het probleem. Is het een kleine groep die herhaaldelijk problematische taal gebruikt, of is het een breder gedragen fenomeen? Het zou daarom interessant zijn om netwerkanalyses te doen om meer zicht te krijgen van de spreiding van online geweld en hoe daders met elkaar interacteren.

Daders

Ook de context van gebeurtenissen in de fysieke wereld speelt een rol. We weten dat nieuwsgebeurtenissen, politieke ontwikkelingen of maatschappelijke spanningen invloed kunnen hebben op hoe mensen zich online uiten (bijv. Menshikova & Van Tubergen, 2022; Czymara & Gorodzeisky, 2024). Sommige uitingen kunnen voortkomen uit ideologische overtuigingen, terwijl andere meer impulsief of terloops zijn. We weten dus nog niet wat de motieven zijn van daders en wat de triggers zijn voor hun online gedrag.

Platforms

In dit onderzoek hebben we gekeken naar Nu.nl en X, en er zijn waarschijnlijk verschillen tussen de twee platformen. Zo trekken Nu.nl en X waarschijnlijk verschillende gebruikersgroepen aan, met uiteenlopende demografische kenmerken, interesses en communicatiestijlen. Dit kan zich vertalen in verschillen in taalgebruik, maar ook in de frequentie en aard van problematische uitingen. Op Nu.nl zagen we dat het percentage problematische taal in goedgekeurde berichten lager ligt dan op X. Dit kan wijzen op effectievere moderatie, maar ook op verschillen in platformcultuur. Omdat we op X geen informatie hebben over afgekeurde berichten, kunnen we de moderatie daar niet goed vergelijken.

Daarnaast weten we niet of gebruikers actief zijn op meerdere platformen, en zo ja, hoe hun gedrag zich daar ontwikkelt. Het zou bijzonder interessant zijn om cross-platform gedrag te analyseren: laten gebruikers op verschillende plekken dezelfde patronen zien? Zijn er overeenkomsten in hoe ze zich uiten, of passen ze hun toon aan per platform? Om dit te onderzoeken zou het nodig zijn om data van meerdere platformen te combineren en gebruikers over die platformen heen te volgen.

3.4.3. Tekstclassificatie

Onze aanvullende social media analyse (zie ook 1.3 in bijlage 1) is een eerste, verkennende schets van de schaal waarop online geweld tussen Nederlandse burgers zich op sociale media voordoet. Daarbij lag de focus op drie veelbesproken zoektermen: "moslim", "fascist", en "wappie/tokkie". Door gebruik te maken van geavanceerde taalmodellen (BERT-gebaseerde classificatiemodellen) is onderzocht welk type berichten online circuleren rond deze termen, en in hoeverre die als problematisch kunnen worden beschouwd. De taalmodellen presteerden niet optimaal op de posts, wat betekent dat de resultaten in het onderstaande stuk moeten worden gezien als een grove inschatting van de aard en omvang van online geweld, maar niet als een (volledig) betrouwbare representatie van de online leefwereld. In de Reflectie gaan we dieper in op deze beperkingen en bespreken we aanbevelingen om hier in de toekomst beter mee om te gaan.

Voor deze analyses hebben we gebruik gemaakt van de data die we eerder in dit onderzoek van X hebben verzameld via Zeeschuimer. Dit zijn alle Nederlandstalige posts die tussen 1 juli 2024 en 31 december 2024 zijn geplaatst waarin de termen (1) “Wappie” en “Tokkie”, (2) “Moslim” of (3) “Fascist” voorkwamen. In totaal resulteerde dit in 236.230 verzamelde posts.

In onze analyses trainen en evalueren we daarom drie individuele modellen (RobBERT v2, BERTje en XLM-RoBERTa) op exact dezelfde 80/20-split van de sample. Per model vergelijken we de F1-scores per zoekterm én voor wanneer we de gelabelde tweets voor de zoektermen samen nemen. Naast deze individuele modellen bouwen we een ensemble dat de voorspellingen van alle individuele modellen combineert. Hierbij leunen we op het fenomeen *majority voting*, wat inhoudt dat het label wordt gekozen die door de meerderheid van de BERT-modellen wordt gekozen. In andere woorden, als twee van de drie modellen aangeven een bepaalde post als discriminatie en haatspraak labelen, dan krijgt deze post dit label. Op deze manier kunnen we een robustere labelling van de posts krijgen omdat de modellen elkaar aanvullen.

We hebben voor elke zoekterm drie trainingsrondes uitgevoerd op alle beschikbare gelabelde tweets binnen die zoekterm, én drie ronden op de volledige dataset (“ALL”). Voor elke zoekterm keken we welke trainingsronde de hoogste F1score gaf op de bijbehorende testset. Bij het classificeren van de ongelabelde posts laden we vervolgens voor elke zoekterm die specifieke trainingsronde met de hoogste F1-score om de nieuwe data te labelen. Hierdoor gebruiken we telkens de best presterende versie van het model per zoekterm.

Het ensemblemodel voor de specifieke zoektermen levert in alle drie de steekproeven de hoogste F1scores en is daarmee duidelijk superieur aan de individuele modellen. Het Ensemble-model behaalt met een F1 van 0,79 op de moslimsubset de hoogste score. Dit is ‘acceptabel’ in de zin dat we met bijna 80 % van de relevante berichten correcte labeling verwachten. Voor de steekproef voor de zoektermen “wappie” en “tokkie” scoort het ensemble oké met een F1 van 0,67, ondersteund door een hoge accuracy van 0,71 en een goede precision van 0,82, wat betekent dat het merendeel van de inschattingen tijdens het trainen van het model goed lukte. Hoewel de ensemblebenadering consequent de beste F1scores oplevert, liggen de absolute waarden voor precisie en recall nog beneden het gewenste niveau.

Uit onze classificatie van posts over de thema’s “fascist”, “moslim” en “wappie en tokkie” komt duidelijk naar voren dat het merendeel van de berichten valt in de categorie Acceptabel. Daarmee bedoelen we dat deze berichten weliswaar soms grof van toon, onbeschoft of polariserend kunnen zijn, maar niet voldoen met de gedefinieerde grens voor strafbare of grensoverschrijdende inhoud. Tegelijkertijd hebben we in mindere mate posts gevonden, die wel in potentie problematisch taalgebruik omvatten, zoals belediging. Daarnaast vonden we relatief veel problematisch taalgebruik richting publieke figuren. Dit ligt buiten de scope van het onderzoek, maar dit bevestigt nogmaals dat prominente figuren een hoger risico vormen voor het ervaren van belediging of haat.

Bij het automatisch classificeren van de posts voor de zoekterm “moslim” viel op dat álle berichten door de modellen als Acceptabel werden gezien; er werden geen tweets in categorieën als Discriminatie, haatspraak & haatzaaien of misinformatie en desinformatie teruggevonden. Dit is bijzonder omdat bij onze handmatige annotatie wél voorbeelden van deze vormen van online geweld en misleidende informatie werden geïdentificeerd (zie Tabel 1). De taalmodellen zelf presteerden redelijk met F1-scores rond 0,79 voor het ensemble en 0,70 voor de individuele BERT-varianten, maar die prestaties hebben er niet toe geleid dat deze problematische categorieën automatisch werden herkend. Het contrasteert sterk met onze verwachtingen op basis van de trainingsdata en onderstreept dat zelfs goed scorende modellen in de praktijk belangrijke categorieën kunnen missen.

Bij de posts met zoekterm “fascist” viel op dat, in tegenstelling tot de andere zoektermen, een substantieel deel van de posts in de categorie Belediging terechtkwam. Deze bevinding sluit aan bij onze handmatige labels: het woord fascist fungeert in de praktijk vaak als scheldwoord, en dit werd niet vaker door een specifieke groep gebruikt. De term kwam vaak in politieke discussies voor en het gebruik van fascist was niet beperkt tot één kant van het politieke spectrum. Zowel voor- als tegenstanders in de online discussies hanteerden de term als persoonlijke aanval. Gezien het gebruik in politieke online discussies, en het breed gebruik van de term, wordt het meer gebruikt als een retorisch instrument om mensen in een bepaald beeld te plaatsen.

3.4.4. Reflectie tekstclassificatie

Ongelijke verdeling in training data

Na het opschonen en verkennen van de handmatig gelabelde data zagen we dat er sprake was van een ongelijke verdeling in de training data. Dit is een belangrijke beperking van dit onderzoek aangezien het invloed heeft op de betrouwbaarheid van de automatische classificatie: het model lijkt beter in het herkennen van dominante categorieën, zoals acceptabele posts, terwijl zeldzamere, subtielere, of overlappende vormen van online geweld mogelijk minder goed zijn opgemerkt. Gezien deze beperking is het belangrijk om te benadrukken dat dit onderzoek slechts een eerste inschatting geeft van de omvang van online geweld.

Bovendien, doordat sommige categorieën in de gelabelde steekproef zeldzamer waren dan andere (bijv. smaad en laster), ontstond er mogelijk bij de training een bias in de richting van de meerderheid. We hebben dit deels ondervangen door meerdere BERT-modellen met elkaar te vergelijken en samen te laten werken in een ensemble model. Door te kijken naar de meest voorkomende voorspellingen van RobBERT v2, BERTje en XLM-RoBERTa voor elke post, verminderden we de variantie in de voorspellingen. Hiermee creëerden we een vangnet voor de voorspellingen waarin de individuele modellen niet goed presteerden. Hoewel deze aanpak niet geheel de ongelijke verdeling kan opheffen, levert het ensemble een robuustere labelling op en maakt het de resultaten beter bestand tegen de dominantie van de meerderheidscategorieën.

Vervolgonderzoek zou deze uitkomsten kunnen gebruiken om in de toekomst accurater te zijn op de minder voorkomende vormen van online geweld. Idealiter wordt een grotere hoeveelheid data gelabeld, zodat minderheidsklassen voldoende voorbeelden bevatten om het model tijdens training expliciet mee te nemen. Daarnaast kunnen technieken als stratified sampling, gerichte oversampling of semi supervised learning helpen om de vertekening te verkleinen zonder blind te vertrouwen op random sampling. Door deze stappen ¬te combineren met onze ensemble aanpak, kunnen we in vervolgonderzoek betere generalisatie en hogere betrouwbaarheid bereiken.

Een andere optie voor vervolgonderzoek omvat het uitlichten van één specifieke categorie per zoekterm en daarop te focussen. Als er veel categorieën zijn om te classificeren, dan kan dit leiden tot aanzienlijke ruis voor de taalmodellen. De kans om hiermee zeldzamere vormen van online geweld te detecteren neemt hiermee af. In het geval van de zoekterm “moslim” zou dat betekenen dat het model uitsluitend traint en evalueert op discriminatie en haatspraak.

Door deze focus vermindert de overlap met andere categorieën en kan de modelperformance op juist die vormen van gewelddadige of grensoverschrijdende taal worden geoptimaliseerd. Hierdoor ontstaat niet alleen een nauwkeurigere detectie van haatvolle uitingen, maar neemt ook de interpretatie van de resultaten toe.

Overlap en grijze gebieden

Een belangrijk inzicht is dat veel categorieën in de praktijk overlappen. Belediging, discriminatie en haatspraak zijn niet altijd strikt te scheiden, en ook counterspeech kan qua toon of inhoud lijken op de uiting waartegen het zich keert. Deze overlap maakt het classificeren van berichten niet alleen technisch uitdagend, maar roept ook vragen op over hoe we online gedrag duiden en adresseren. Wat als een vorm van verdediging is bedoeld, kan in de praktijk alsnog bijdragen aan verhitte en polariserende discussies.

Dit grijze gebied en de overlap tussen categorieën viel ons ook op tijdens het handmatig labelen van de steekproef. Zo kunnen posts waarin er sprake is van discriminatie of haatspraak tegelijkertijd beledigende of intimiderende uitingen bevatten, en vormen van misinformatie kunnen bijvoorbeeld gepaard gaan met agressieve taal. In onze eerdere publicaties wezen wij al op het gebrek aan breed gedragen definities voor vormen van online geweld. We baseerden ons in dit onderzoek vooral op strafrechtelijke kaders en internationale wetenschappelijke literatuur, maar de prestatie van de tekstmodellen suggereert dat deze kaders niet altijd aansluiten op de dynamische en contextafhankelijke aard van socialmedia berichten in Nederland.

Ook dit kan een reden zijn waarom bepaalde prestatiemetrics niet goed uitvielen Mogelijk hadden BERTmodellen moeite om deze samengestelde uitingen consequent te scheiden, omdat zij per definitie één label per post toewijzen. Ook kan het zo zijn dat het de subtiele varianten van haatspraak of intimidatie moeilijk kon herkennen omdat bepaalde vormen van online geweld erg genuanceerd zijn. Dit gegeven kan leiden tot ruis in de voorspellingen en F1scores voor de zoektermen. In vervolgonderzoek kan er in meer detail naar online uitingen van online geweld gekeken worden zodat de classificatie van de verschillende vormen nog scherper kan worden geformuleerd. Dit kan bijdragen aan een breder begrip van de vormen van online geweld, en een nauwkeurigere afbakening van deze vormen.

Een schaal van online gedrag: van acceptabel naar strafbaar

Desalniettemin blijkt de schaal die we hebben gehanteerd – van acceptabel gedrag tot aan strafbare uitingen – waardevol om het fenomeen online geweld beter te begrijpen. Het helpt om niet alleen te kijken naar de uitersten, maar juist ook naar het brede tussengebied. Hier bevindt zich het zogenaamde 'onbeschoft' gedrag: uitingen die niet strafbaar zijn, maar wel bijdragen aan een verharding van het publieke debat—waardoor bepaalde meer kwetsbare groepen zich juist meer gaan terugtrekken uit de online wereld en het publieke debat dat zich online afspeelt. Deze categorie verdient meer aandacht in beleid en preventie. Door tijdiger te interveniëren in dit segment, kan mogelijk worden voorkomen dat online interacties escaleren naar vormen van strafbaar geweld en kan ervoor zorgen dat iedereen, met name kwetsbare groepen zich veiliger voelen online.

De balans tussen vrijheid en bescherming

Deze eerste verkennende schaal onderstreept nogmaals de fundamentele uitdaging van de balans tussen vrijheid van meningsuiting en bescherming tegen online schade. Wanneer wordt scherpe kritiek of satire grensoverschrijdend? En hoe voorkomen we dat de norm vervaagt tussen polariserende maar legitieme opinies en haatdragende communicatie? In dit onderzoek hebben we geprobeerd de grenzen van online geweld scherp te definiëren aan de hand van juridische en maatschappelijke kaders, maar de interpretatie hiervan blijft deels contextueel en subjectief. Daarom is aanvullend onderzoek nodig, met name op andere sociale mediaplatforms. X wordt door slechts een selecte groep Nederlandse burgers gebruikt en heeft een eigen ecosysteem; voor een volledig beeld is het belangrijk om deze analyse aan te vullen met data van andere platforms zoals TikTok, Facebook en Instagram.

4 Betrokken burgers: ”plegers” en “slachtoffers” van online geweld

4.1. Achtergrond

De achtergronden van plegers en slachtoffers zijn divers en moeilijk eenduidig te typeren. Toch zijn er op basis van de interviews met experts, burgers en twee focusgroepen met professionals voorzichtig enkele patronen te onderscheiden.

4.1.1. Plegers

Plegers van online geweld zijn overwegend man en variëren in leeftijd. Vooral jongere plegers gebruiken met name platforms zoals *Snapchat en Discord*, oudere plegers maken vaker gebruik van “klassiekere” sociale media zoals *Instagram, X, of Facebook*. Hoewel *anonimiteit* een grote rol speelt, blijkt uit de interviews dat een aanzienlijk aantal plegers de bedreigende of haatdragende reacties onder hun *eigen naam of met herkenbare* gegevens plaatsen, zeker op publieke platforms. Deze digitale plegers zijn soms actief in *georganiseerde (sub)culturen*, zoals de drill-rapscene, waarin jongeren online worden geronseld (dit gebeurt soms ook door ‘ouderen’ of andere jongeren) en waarbij ze aangemoedigd worden tot online/offline escalatie. Zo is er een voorbeeld van mensen die soms onder naam- en toenaam reageerden onder een YouTube-filmpje, waar een respondent die we spraken in was geknipt:

> “*...en op elke van die media zijn er echt enórm veel reacties gekomen waarin er werd opgeroepen mij te vermoorden, mij te verkrachten op alle mogelijke manieren die je maar kan verzinnen.”* Burger

Anderen plegen hun daden juist individueel, ad hoc uit frustratie, zoals tijdens het gamen veel voorkomt. Anderen plannen hun bedreigingen langer van tevoren. Zo vertelde een respondent dat een vriend van hem werd bedreigd, waarop die vriend vervolgens aan hem vroeg of hij de adresgegevens van die man kon opzoeken.

> “*En toen die persoon die dus die doosbedreiging kreeg, heeft toen dus een foto teruggestuurd van zijn voordeur. En die heeft gezegd tegen die persoon van, ‘Hé, is dat een beetje een leuke wijk om te wonen? Want ik wil nog een huis kopen.’ Dus hij werd bedreigd door iemand van, hé, ik ga je afmaken. En hij stuurde een foto terug van die persoons voordeur met een beetje de algemene vraag van, hé, is dat een beetje een leuke wijk? Wat natuurlijk een één-op-één bedreiging terug is.”* Burger

4.1.2. Slachtoffers

Slachtoffers van online geweld vormen eveneens een brede groep. Ze zijn *niet altijd intensief online*, hoewel een sterke online aanwezigheid het risico op slachtofferschap kan vergroten (zie ook de inzichten uit literatuurstudie). *Specifieke kenmerken* maken mensen kwetsbaarder, zoals seksuele oriëntatie, genderidentiteit, politieke of maatschappelijke zichtbaarheid of afwijkende meningen binnen gepolariseerde debatten. Ook *vrouwen en meisjes* zijn vaak doelwit, waarbij in sommige gevallen het online geweld samenvalt met seksisme.

> “*Ik denk uiteindelijk... dat het iedereen kan overkomen. Zeker als je gaat kijken, in die puberleeftijd. Je hebt toch een aantal aspecten. Je wil leuk gevonden worden en je wil meedoen met de groep. Maar als je die groep gaat splitsen, een beetje gaat trechteren dan speelt toch wel een beetje de zwakkere en LVB jongeren erbij mee. Alleen uiteindelijk kan het iedereen overkomen. Want ik kom ook op het vwo of op het vmbo, mavo, praktijkschool. Dus wat dat betreft maakt dat niet zo heel veel uit. Alleen de grotere groep is wel de kwetsbare. En zeker wel met een LVB”.* Deelnemer focusgroep

> “*Ik heb het gevoel dat je als vrouw iets op het internet doet, je een schietschijf bent voor allerlei haat en geweld. Er zijn weinig manieren waarop je iets kan doen zonder negativiteit over je heen te krijgen. Behalve wanneer je niet uitgesproken bent, of schattig. Maar ook weer niet te sexy want dan krijg je ook weer comments. Er is elke keer wel weer een nieuwe vrouw”.* Burger

Jongeren en kinderen worden ook veelvuldig als slachtoffer (maar ook als pleger) genoemd.

> “*Ze hebben een groepsapp aangemaakt. Die kids. En dat gaat buiten de schoolperken om. Wij hebben daar geen zicht op, maar we horen allerlei bedreigingen via Playstation, spelletjes gaan. Ik krijg daar screenshots van in mijn mailbox, kijk nou dit is gewoon groep 6, kinderen die elkaar de meest erge dingen toe wensen.”* Deelnemer focusgroep

4.2. Overlap plegers en slachtoffers

Uit de interviews blijkt dat er in bepaalde contexten overlap is tussen pleger en slachtoffer, hoewel sommige respondenten ook alleen slachtoffer zijn. De dynamiek en overlap tussen pleger en slachtoffer komt met name naar voren bij wederkerige reacties: iemand wordt beledigd of bedreigd, slaat online terug en wordt vervolgens zelf weer beschuldigd van grensoverschrijdend gedrag. Ook komt het voor dat offline gebeurtenissen, zoals een confrontatie op een demonstratie, zich verplaatsen naar online, of andersom. Zo werd de hele vriendengroep van een vrouw bedreigd nadat een vriendin van haar een foto online had gezet van een man tijdens een demonstratie:

> *“Het was iemand die zij bij een demo gezien had waar zij tegendemonstrant van was, en waar zij een foto van op haar Instagram Stories had geplaatst. En die is vervolgens een soort van geobsedeerd door de hele vriendengroep gaan zitten kijken of die iets kon vinden.”*

Experts vertellen ook dat ze soms zien dat bijvoorbeeld iemand slachtoffer is geworden van online geweld en vervolgens ook zelf online geweld gaat plegen. Ze zijn er dan bijvoorbeeld mee in aanraking gekomen. Dit gebeurt vaak in, bijvoorbeeld, bepaalde subculturen zoals de drillrap scene of online multi-player games (zie casus 3).

4.3. Modus operandi plegers

De werkwijze van plegers van online geweld is veelzijdig en varieert van spontane reacties tot goed voorbereide campagnes. Sommige daden zijn impulsief, bijvoorbeeld beledigingen of bedreigingen tijdens een online game of onder een sociaalmediabericht. Het wordt van kwaad tot erger:

> *“In het begin had ik mijn twijfels, zo van dat moet ik eigenlijk niet zeggen. Maar op een gegeven moment, ik weet het niet, dan gaat er een knopje om en het boeit niet meer of iets dergelijks”* Burger

Andere acties worden doelgericht uitgevoerd, zoals doxing, het aanmaken van nep- en/of haatprofielen, of het systematisch stalken en bedreigen van iemand, vaak vrouwen, via meerdere kanalen.

Er is een tendens zichtbaar van escalatie: online geweld begint vaak met interacties die misschien beledigend of aanstootgevend zijn, maar niet per se strafbaar. Dat kan vervolgens uitgroeien tot intimidatie, bedreiging of het oproepen tot geweld.

In sommige gevallen is er sprake van groepsgedrag: iemand komt in de schijnwerpers te staan, bijvoorbeeld als activist of demonstrant, en wordt dan collectief belaagd.

Ook worden soms jongeren aangezet tot online geweld of geronseld door oudere jongeren:

> *“En dat is nu echt onze prioriteit, omdat daar ook in de hoge leeftijdsegmenten wel grote mannetjes zitten. Die dus kleinere jongeren recruteren, et cetera. Dus er zit een heel piramidesysteem aan vast”.* Praktijk expert

In combinatie met dat de scheidslijn van iets wat mag, eigenlijk niet kan, naar iets wat niet kan en ‘online geweld’ is, zo onduidelijk is én escalatie snel plaatsvindt, maakt dat online geweld naast het soms doelbewust en strategisch wordt ingezet (zie ook hoofdstuk 3), ook soms ‘impulsief’ en ‘spontaan’ plaatsvindt en escaleert. Dit betekent dat de pleger op dat moment impulsief of emotioneel (bijvoorbeeld vanuit een bepaalde frustratie) handelt. Vanuit de literatuurstudie wordt de *moral disengagement* aangedragen als een verklarende factor voor de modus operandi: de normering en de anonimiteit maken dat online geweld gemakkelijk en snel kunnen escaleren.

4.3.1. Niet bewust en niet serieus nemen

Tegelijkertijd maakt online geweld voor sommige respondenten zodanig deel uit van hun digitale ervaring, dat het nauwelijks nog als grensoverschrijdend wordt ervaren. Voornamelijk op bepaalde fora, livestreams of in gamingcontexten zonder moderatie lijkt online agressie sterk genormaliseerd. Er lijkt sprake te zijn van trivialisering van online geweld, waarbij het wordt gebagatelliseerd, geïnternaliseerd of zelfs wordt gezien als vormend. Dat maakt het lastig om slachtofferschap en/of plegerschap te (h)erkennen, zeker wanneer het online geweld ook geen directe gevolgen lijkt te hebben in het fysieke leven.

Plegers geven vaak weinig blijk van bewustzijn van de gevolgen van hun handelen. Het online geweld wordt regelmatig afgewimpeld als ‘maar een grapje’, terwijl het effect op slachtoffers groot kan zijn. Zo werd een tienermeisje gepest (zie casus 2) en nadat ze ook nog online werd bedreigd met verkrachting, kwam er een politieagent langs in de klas. Nadien kreeg ze een anoniem sms’je:

> *“Waarom heb je de politie erbij betrokken? De verkrachting was maar een grapje hoor.”* Burger

Ook werd online geweld gerationaliseerd of gerechtvaardigd, door te benoemen ‘dat het online gebeurde, dus geen fysieke gevolgen heeft’. Anonimiteit speelt een grote rol in de modus operandi van plegers. Het zorgt voor meer vrijheid om te uiten, maar ook om over de schreef te gaan.

> *Je kwam in aanraking met online gelijkgestemden. Dat verplaatste zich ook binnen een maand richting Telegram, waar de verschillende communities op kwamen. Daar zat en zit ik nog steeds ook volledig anoniem op.”* Burger

Ook de platformspecifieke cultuur speelt een rol. In bepaalde online omgevingen zoals specifieke games of fora, is agressief, discriminerend en haatdragend gedrag genormaliseerd. Wanneer moderatie toeneemt op zo’n platform ontstaat vaak een waterbedeffect, waarbij plegers simpelweg uitwijken naar andere, minder gereguleerde kanalen. Ook kan de moderatie relatief makkelijk omzeild worden door berichten met andere tekens te spellen of andere woorden te gebruiken.

> *Er zal altijd iets doorheen komen. Een beetje letters vervangen met cijfers die erop lijken, of andere letters, of buitenlandse tekens. En er zullen altijd mensen komen die zeg maar door die chatfilters heen komen.”* Burger

Ook geven plegers aan (die soms ook slachtoffer worden) dat er ook sprake is van een soort gewenning of ‘afstomping’:

“ Ja, dat vond ik vroeger echt verschrikkelijk. Op een gegeven moment, nou niet dat woord, maar tijdens een spelletje. Als iemand dat tegen je zegt en je zegt het terug, of iets anders, dan voelt het eerst wel kut. Maar als je het heel vaak doet, op een dag bijvoorbeeld, of in de week, dan wordt het genormaliseerd.” Burger

Casus 3 Gaming

Eli was dertien toen hij voor het eerst echt opging in de wereld van online gaming. Counter-Strike, later ook League of Legends. In het begin speelde hij vooral met vrienden, via Discord, waar ze grapten, overlegden, zich opfokten. Maar na verloop van tijd ging hij ook alleen spelen. Dan werd het anders. Heftiger. *Als je alleen bent, dan is de drempel lager om terug te schelden,”* zei hij. *“Dan ken je niemand, en het boeit je minder wat je zegt.”*

Hij had het nooit echt gemeend, die doodsbedreigingen en scheldwoorden. Het hoorde erbij. Zoals hij het zei: *“Je hoort het zo vaak, je voelt het op een gegeven moment niet meer.”* Toch vond hij het raar van zichzelf. Dat hij dingen zei waar hij als kind nog van gruwde. Maar ja, het spel maakte iets in hem los, zeker op slechte dagen.

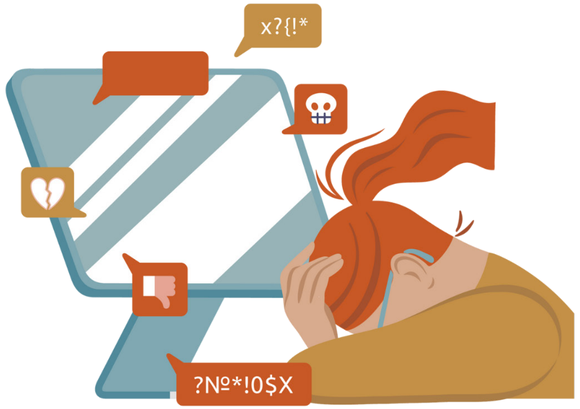
Wat hem hielp, zei hij, was relativeren. Niet alles serieus nemen. Zelfspot. Grenzen kennen. En als anderen die grenzen overschreden? *“Dan blokkeer ik ze. Of ik ga gewoon weg.”*

Hij geloofde niet in een perfect schoon internet. *“Als je iedereen gaat bannen die iets verkeerds zegt, gaan ze gewoon ergens anders verder. Je moet het leren dragen, niet verbieden.”* Hij vond dat moderatie nodig was — zeker als mensen écht anderen bedreigden — maar niet alles moest gladgestreken worden. *“Het internet leert je dat je niet alles serieus hoeft te nemen, daar word je hard van”*.

4.3.2. Sociale media analyses

Op basis van de bevindingen uit de sociale media analyses kunnen we voorzichtig enkele uitspraken doen over de modus operandi van de plegers — oftewel: de manier waarop mensen zich online uitlaten wanneer ze discriminerende, bedreigende of beledigende taal gebruiken. Hoewel we niet over individuele daders spreken, kunnen we wel patronen in het gedrag herkennen die iets zeggen over hoe online geweld zicht manifesteert.

Uit de analyses bleek dat bepaalde groepen consequent worden geassocieerd met negatieve stereotypen. Zo werd de zoekterm *'Moslim'* vaak samen gebruikt met het woord *'tuig'*. Eerder beschreven we in hoofdstuk 3 dat uit eerder onderzoek blijkt dat moslims met negatieve stereotypen geassocieerd worden.



Een mogelijk vergelijkbaar mechanisme zien we bij de termen *'Wappie'* en *'Tokkie'*. In de verzamelde posts worden deze woorden vaak vergezeld door termen als *'idioot', 'sukkel', 'achterlijke'* en *'achterlijk'*. Deze woorden impliceren een gebrek aan intelligentie of irrationaliteit. Ook hier kunnen er dus negatieve stereotypen gebruikt worden om mensen online weg te zetten.

Naast het gebruik van specifieke stereotypen valt op dat verschillende onderwerpen in verschillende mate problematisch taalgebruik oproepen. Zo zien we dat posts over 'Moslims' relatief vaak bedreigende taal bevatten, terwijl bij 'Fascist' en 'Wappie/Tokkie' juist discriminerende of denigrerende termen domineren. Dit kan erop wijzen dat het onderwerp zelf invloed heeft op de aard van de reacties die het oproept. Daarnaast kan het ook zo zijn dat gebruikers mogelijk sneller getriggerd zijn door bepaalde onderwerpen dan bij andere onderwerpen.

4.4. Impact

4.4.1. Individuen en gemeenschappen

Online geweld is daarbij zowel gericht op individuen maar ook hele groepen worden 'weggezet'. Het heeft gevolgen voor de impact die het op iemand of groepen heeft.

> *"Ik denk dat je daar ook weer een verschil moet maken, dus: is het op groepen gericht of echt ook op individuele personen? Ik denk dat daar al verschil in zit. Want haat zaaien, discriminatie, seksisme… Zo'n Tate, die zet hele groepen negatief weg, maar soms is het ook op een bepaald persoon gericht. Van oh, jij zegt dat, maar jij bent moslim of joods. Ik denk dat daar verschil in zit… In het effect... Ik denk, stel nou dat jij persoonlijk wordt aangevallen omdat je moslim bent, je bent joods of je bent niet hetero. Ik denk dat dat voor die ene persoon heel erg bedreigend is. Ik denk dat als je lid bent van een groep die wordt aangevallen, en de aanval is algemeen of een of ander samenzweringsverhaal. Ik denk dat je dat ook enorm kan raken en treffen. Mij lijkt nog ingewikkelder als je ook echt persoonlijk met naam en toenaam genoemd wordt."* Wetenschappelijk expert

Online geweld heeft niet alleen directe gevolgen voor individuele slachtoffers (psychisch welzijn), maar het tast ook de fundamenten van de samenleving aan.

Respondenten geven aan dat de ervaren scheiding tussen online en offline geweld, juist de impact van online geweld versterkt.

> *"De ervaren scheiding van online en offline wereld zorgde bij sommige mensen juist voor meer extra impact. Dus 'ik dacht dat alleen online was, maar het deed me toch meer dan ik van tevoren had kunnen denken'. Die online wereld hoort ook echt in je dagelijks leven in alle aspecten komen die angstgevoelens bijvoorbeeld terug."* Deelnemer focusgroep

In zowel de interviews met experts, slachtoffers en plegers kwam naar voren dat de impact van online geweld nog wordt onderschat ten opzichte van fysiek geweld. Maar door de reikwijdte van de online wereld, kan online geweld juist meer impact hebben op zowel slachtoffers als de bredere samenleving.

> *"Ongeacht waar jij bent. En dat maakt het ook indringender, hè?"* Praktijk expert

> *Ja wat voor effecten denk je dat zo'n vorm van haat zaaien heeft? Die dus ook groter kan zijn, want je hebt toegang tot meer groepen die je misschien in je directe omgeving niet eens altijd hebt. Nou ja, het vergiftigt debatten en ik denk, wat bij dat haat zaaien ook hoort is desinformatie."* Wetenschappelijke expert

Veelal de combinatie van de opmerkingen zelf, de intensiteit, heftigheid en de grote schaal ervan maakt het impactvol.

> *Iedereen krijgt af en toe wel shit over zich heen, weet je wel, of een opmerking. Maar er zit ergens een omslagpunt en dat zit hem in, denk ik, de intensiteit, de heftigheid en hoe massaal het is ofzo. Ja, en wat er dan natuurlijk nog bovenop komt is dat ze weten waar ik werk, waar mijn sportschool is enzo weet je wel, dus dat ze ook nog een keertje echt en werkelijk ín mijn omgeving komen, fysiek. Ik denk dat dat wel echt die optelsom is ja, waardoor het dus heel intens en heel heftig wordt".* Burger

4.4.2. Psychologische en emotionele gevolgen

Veel slachtoffers van online geweld ervaren gevoelens van angst, machteloosheid, woede, verdriet en/of schaamte.

> *Maar het nare van online is gewoon dat je jezelf heel moeilijk kunt verdedigen. Kijk, op het moment dat iemand hier achterom komt en die zegt ik vind dit of dit. Dan kun je uitleg geven, kun je weerwoord geven. En dat is gewoon veel lastiger op het moment dat zoiets online gebeurt. En je weet ook niet wie ziet dit allemaal, wie is hier allemaal getuige van, wat voor eigen leven gaat het leiden. Je weet, er zijn meerdere mensen die hier screenshots van hebben gemaakt. Dus ja, je weet niet waar het stopt en waar het blijft. En dat vind ik natuurlijk met online is dat wel heel vervelend."* Burger

Vooral bij langdurige of grootschalige vormen van online geweld waarbij er ook een offline component is, kan het gevoel van onveiligheid lang doorgaan en in vele aspecten van het leven doordringen:

> *Ik heb heel lang het centrum vermeden en drukke locaties. Zo ook grote evenementen. Op het moment dat ik weer voor de eerste keer naar een café kwam waar ik mezelf altijd thuis voelde en er super vaak kwam, was ik eigenlijk alleen maar bezig met de omgeving scannen, kijken waar de nooduitgang was. Als iemand binnenkwam [die er verdacht uitzag naar mijn mening], raakte ik in paniek. Dus je zit eigenlijk altijd met het gevoel in je hoofd van shit er gaat zo misschien wat gebeuren."* Burger

Een belangrijke factor is de mate van anonimiteit van de pleger. Wanneer slachtoffers niet weten wie er achter de bedreigingen of beledigingen zit, versterkt dat het gevoel van onveiligheid. De ongrijpbaarheid van de bron maakt het lastig om in te schatten hoe serieus het gevaar is, en of de situatie kan escaleren naar offline dreiging. Juist deze onzekerheid zorgt bij veel slachtoffers voor verhoogde stress en angstgevoelens. Zoals verwoord door een participant, die gestalkt en bedreigd werd door een persoon die haar gegevens in een Telegramgroep heeft gedeeld, waardoor ze berichtjes kreeg van vreemde mensen:

> *"Je zit heel erg in onwetendheid. Dus bij elke man die me een beetje aankeek dacht ik van: 'eh, kom jij van die Telegram? Ken je mij ergens van?' Dat was wel een paranoïde fase."* Burger

Ook gaven meerdere respondenten aan dat ze verrast waren over de mate van impact die ze voelden. Een respondent waarbij een filmpje van haar bedrijf werd gedeeld met lasterlijke informatie zegt er het volgende over:

> *Wat ik het gekke vind is, ik ben iemand die best tegen een stootje kan, ik kan ook prima met discussies omgaan, dat vind ik zelfs leuk. Maar ik was best wel verrast over hoe ellendig ik me hierover gevoeld heb. Ik had niet verwacht dat het zo'n impact zou hebben. Dat je eigenlijk geneigd bent om, ja, je zou bij wijze van bij je bedrijf de gordijnen dicht willen doen, de deuren dicht willen doen en alles willen weren. Je voelt je echt te kijk gezet of zo. Je voelt je wel heel kwetsbaar. En dat had ik eigenlijk niet verwacht."* Burger

Een andere respondent herkent dit beeld ook:

> *Ik had nooit gedacht dat die online haat, drek, intimidatie en agressiviteit die je over je heen krijgt, wat dat werkelijk met je doet, met je veiligheidsgevoel doet. Ook met je omgeving, maar ook echt wel met je eigen gevoel. Nu ben ik een best wel een grote vent, sport ik genoeg en heb ik een grote hond. Maar dan nog, als ik ergens buiten liep en een groepje*

mensen zag staan, liep ik de andere kant op. Het doet bewust heel veel met je, maar ook onbewust ben je er continu mee bezig. En dat is iets wat ik heel erg had onderschat, en ik denk wat heel veel mensen enorm onderschatten." Burger

Naast angst noemen enkele slachtoffers ook schaamte en het gevoel dat ze zelf iets verkeerds hebben gedaan. Dit werd soms versterkt door *victim blaming (mogelijk heb je zelf ook iets gedaan waardoor jij slachtoffer bent geworden van')* in hun omgeving, of het uitblijven van serieuze opvolging door instanties. Zo zijn er respondenten die een doodsbedreiging via een sociaal medium ontvingen, vervolgens een melding maakten bij het desbetreffende platform waar daarna niks mee werd gedaan. Dat droeg bij aan het gevoel van onveiligheid:

> *En vervolgens wat mij eigenlijk meer deed schrikken was hoe Facebook daar dan vervolgens op reageerde."* Burger

> *Dus ja, je schrikt ervan. Niet dat ik dacht: 'ik ga nu dood', maar eerder: 'wat is dat een raar figuur zeg'. En wat mij vervolgens verbaasde, was de reactie van Facebook. Je hebt de optie om iets te rapporteren, maar er is nooit iets verwijderd of dat iets serieus is opgepikt."* Burger

Casus 4 Platforms

Sophie scrollde gedachteloos door Threads. Haar profiel was eenvoudig: een lachende profielfoto, een korte bio over haar werk als penningmeester bij een werkgroep, en twee vlaggetjes—een trans-vlag en een regenboogvlag. Een klein gebaar, maar voor haar een bevestiging van wie ze was.

Het bericht kwam onverwacht. *"Jij denkt zeker dat die vlaggen jou veilig maken? Wacht maar, ik maak je af."*

Haar hart sloeg over. Een doodsbedreiging. Zonder aanleiding, zonder reden, behalve haar bestaan. Ze blokkeerde het account en meldde de dreiging direct bij Threads. Weken verstreken. Toen kwam het antwoord: "*We hebben je melding beoordeeld en zullen geen verdere actie ondernemen*."

Sophie staarde verbijsterd naar het scherm. Was het dan zo moeilijk om je aan je eigen gedragsregels te houden?

Het was niet de eerste keer. Op Instagram had ze eerder bedreigingen ontvangen, toen iemand systematisch haar vriendengroep was afgegaan. Een vriendin werd gedoxt. Daar greep Instagram uiteindelijk wél in—maar pas na meerdere meldingen. Op Threads bleef het stil.

Langzaam veranderde haar online gedrag. Ze leerde het verschil tussen echte dreiging en loze woorden, maar de spanning bleef. Online bedreigingen voelden soms abstract, maar in het echte leven werd ze wél eens vastgegrepen op straat. De dreiginghield nooit echt op, alleen de vorm veranderde.

Maar wat haar misschien nog wel het meest verontrustte, was niet de dreiging zelf, maar de onverschilligheid van het platform. Threads werd gepresenteerd als een veilig alternatief voor Twitter. Maar hoe veilig was het echt—voor haar?

4.4.3. Praktische en sociale gevolgen

Veel slachtoffers geven daarnaast aan hun gedrag in het dagelijks leven te veranderen. Zij trekken zich terug uit het publieke of professionele domein, verwijderen sociale media-accounts, passen hun KVK-inschrijving aan ter bescherming van hun woonplaats of vermijden bepaalde fysieke locaties. Zo vertelt een vrouw die onbedoeld in een YouTube-filmpje terecht was gekomen:

> *“Ik ben toen ook wel eens op straat herkend, ik heb al mijn sociale media moeten deactiveren. Ik ben een tijdje niet mijn huis uitgekomen omdat ik bang was om herkend te worden, en ik was ook heel bang om gedoxt te worden omdat dat dat heel makkelijk zal gaan, er staat ook op mijn KVK-inschrijving waar ik woon. Dus ik dacht als deze mensen een steen door de ruiten willen gooien dan kunnen ze dat zo doen.”* Burger

In sommige gevallen betekent dit ook dat slachtoffers meer moeite hadden om hun werk uit te oefenen, of dat ze zich niet liever terugtrekken uit een publiek debat of demonstratie.

> *“En ik ben sindsdien ook niet meer bij die demonstratie geweest. Ik denk dat dat niet voor niets is. Ik ga het zelf niet meer opzoeken. Nou kon ik al die keren niet, maar dat kwam me eigenlijk ook wel goed uit als ik eerlijk ben.”* Burger

> *“Mensen trekken zich terug en allebei is niet goed, want toenemende felheid komt het debat niet ten goede. Als mensen zich terugtrekken, ga je ook waardevolle stemmen verliezen.”* Wetenschappelijke expert

De impact van online geweld op slachtoffers wordt beïnvloed door een aantal samenhangende factoren. Een belangrijke rol hierin speelt de mate van anonimiteit van de pleger. Wanneer iemand wordt belaagd door een onbekende, is het gevoel van controleverlies en onveiligheid vaak veel groter. Het niet weten wie er achter het geweld zit, maakt het moeilijk om in te schatten hoe ernstig de dreiging is en of deze kan escaleren. De onzichtbaarheid van de pleger voedt onzekerheid en wantrouwen, wat het psychisch welzijn van het slachtoffer kan verergeren.

Daarnaast speelt de zichtbaarheid en vindbaarheid van het slachtoffer een grote rol.

> *“Ja, ik heb ook wel een tijdje dat ik dacht, nou liever ga ik mijn huis niet uit. Maar ik voelde me thuis ook onveilig, want ik wist niet zo goed wat ze van me hadden en waar ze achter konden komen. Ik denk, ja, er zijn genoeg slimme mensen die, als je mijn foto’s hebt en mijn nummer, en dan weet je waar ik werk, je weet niet hoe gek de mensen zijn. Straks volgen ze je naar huis, je weet het allemaal niet.”* Burger

Mensen met een open online profiel, of personen die om wat voor reden dan ook in de publieke belangstelling staan (b.v. door hun betrokkenheid bij protest of belangengroepen), worden sneller opgemerkt en zijn daardoor ook makkelijker doelwit van online intimidatie of bedreiging. Een hoge mate van digitale zichtbaarheid kan leiden tot herhaald slachtofferschap, juist omdat plegers eenvoudig toegang hebben tot persoonlijke informatie. Dit vergroot de kans op doxing en versterkt het gevoel van kwetsbaarheid en draagt bij aan langdurige stress. Ook de mate van digitale weerbaarheid van het slachtoffer beïnvloedt de ervaren impact. Mensen die goed weten hoe zij zich online kunnen beschermen, bijvoorbeeld door gebruik te maken van privacy-instellingen, blokkades en veilige wachtwoorden, ervaren doorgaans minder (impact van) online geweld. Zij voelen zich competenter en kunnen sneller en effectiever reageren op digitale dreiging. Een gebrek aan kennis of middelen om zichzelf online te beveiligen kan het gevoel van machteloosheid juist versterken.

Tenslotte speelt de reactie van de sociale omgeving een doorslaggevende rol. Wanneer slachtoffers steun ervaren van vrienden, familie, collega’s of instanties, draagt dat bij aan het gevoel van erkenning en herstel. Emotionele steun biedt een belangrijke buffer tegen de negatieve gevolgen van online geweld. Daarentegen kan afwijzing, bagatellisering of ongeloof uit de omgeving het trauma verergeren. Slachtoffers voelen zich dan niet alleen onveilig, maar ook alleen gelaten.

Tot slot is de impact van online geweld aanzienlijk groter wanneer het gepaard gaat met dreiging van fysiek geweld. In gevallen waarin online agressie overgaat in concrete bedreigingen of acties in de offline wereld – zoals stalking, doxing of intimidatie aan huis – wordt de dreiging zeer tastbaar. De scheidslijn tussen de digitale en fysieke leefwereld vervaagt, waardoor slachtoffers zich ook in hun dagelijkse leven niet meer veilig voelen. Dit kan leiden tot ingrijpende veranderingen in gedrag, sociale isolatie en ernstige psychische klachten.

4.5. Copingmechanismen slachtoffers

Slachtoffers gebruiken verschillende strategieën om met de impact van online geweld om te gaan. Deze copingstrategieën zijn grofweg in te delen in 1) probleemgerichte coping 2) emotiegerichte coping.

4.5.1. Probleemgerichte coping

Dit houdt in dat de slachtoffers zich vooral richten op wat rationele en/of praktische oplossingen zijn om met hun ervaring met online geweld om te gaan. Het gaat dan bijvoorbeeld om het doen van aangifte, steun zoeken (bij vrienden of familie), het offline halen van sociale media profielen, uit een groepsapp stappen en een tegen narratief bieden. Dit deed bijvoorbeeld een respondent door een column te schrijven om een tegengeluid te bieden:

Casus 5 Tegengeluid

Toen de video over een politiek onderwerp online kwam, had Lotte niet meteen door wat het losmaakte. Het was een fragment van een vlogger, die iemand lastigviel met zijn camera. Ze sprak de vlogger aan en had kritiek op zijn werkwijze. En hij zette het gesprek online.

"Op elk platform kwam het voorbij," zei ze. *"Instagram, TikTok, Facebook. Altijd net genoeg tijd ertussen om het opnieuw viral te laten gaan."* Er waren geluidseffecten onder gezet, stukjes uitgeknipt. Het deed het goed op het algoritme, de reacties stroomden binnen. Ze werd belachelijk gemaakt en uitgelachen, maar er waren ook extremere reacties. Doodswensen werden geuit. Er werd opgeroepen haar gegevens te publiceren. Een bekende probeerde haar te chanteren. Ze werd herkend op straat, sliep slecht. Ze verwijderde haar accounts.

Wat haar het meest raakte was niet het filmpje zelf, maar de infrastructuur eromheen: een platform dat bewust dit gedrag en de gevolgen daarvan faciliteerde, zonder enige bescherming voor wie erin belandt. Een melding doen bij het platform was mogelijk, maar dan zouden de makers van het filmpje ook haar account te weten komen. Ondertussen bleven ze het filmpje posten, voor likes en clicks. *"Het is een verdienmodel."*

Ze bleef lang stil omdat ze zich schaamde voor het fragment. Voor vrienden, familie en collega's. Die zouden het toch niet begrijpen. Aangifte doen of andere hulp weigerde ze ook. Ze wilde zich ook geen slachtoffer voelen, daarmee gaf ze de anderen te veel macht.

Tot ze haar verhaal opschreef in een column en zo haar verhaal terug claimde. *"Ik had geen keuze. Het was dat, of altijd in angst leven."*

4.5.2. Emotiegerichte coping

Dit houdt in dat slachtoffers vooral op een emotionele manier (soms bewust soms onbewust) proberen om te gaan met de ervaringen en gevoelens van het online geweld dat zij ervaren/hebben ervaren. Het gaat dan bijvoorbeeld om stil houden/niet over willen praten, negeren/wegdrukken 'alles wat aandacht geeft groeit', relativeren of weglachen. Maar ook het vermijden, ontkennen en zich focussen op hun sociale omgeving zoals zorgen voor familie.

> *" Nou toen ze [slachtofferhulp] me helemaal aan het begin benaderden, zat ik nog helemaal in de modus van overleven en zorgen dat mijn omgeving het goed heeft en oké is. Ik had nog niet helemaal het besef wat het met mijzelf deed."* Burger

Ook spreken we respondenten die zichzelf hebben afgezonderd. Ze willen zich bijvoorbeeld geen slachtoffer voelen of zielig gevonden worden:

> *"Ja ik denk dat ik het ook gewoon heel cringe en gênant vond. En ik wilde ook niet dat mijn vrienden voor mij op zouden komen. Omdat ik denk dat zij helemaal niet begrijpen hoe dit werkt en veel van de hulp die ik aangeboden kreeg, gaf me eigenlijk alleen maar stress."* Burger

> *"Tegelijkertijd had ik helemaal geen zin om me slachtoffer te voelen en zielig gevonden te worden. En had ik het idee dat hoe meer mensen me als zielig zouden zien, hoe meer dat eigenlijk een 'win' is ofzo, voor de ander. Dus ik had helemaal geen zin om slachtofferhulp te krijgen van wie dan ook. En ik ben ook helemaal niet zielig, of wat dan ook. Ik sta nog steeds achter van wat ik gezegd heb. Dus ik hoef niet als een gewond vogeltje opgevangen te worden. Bovendien begrepen mijn vrienden totaal niet wat de grootte was van dit platform en wat de gevolgen van online haat kan zijn, omdat zij dit nog nooit hebben meegemaakt en zij zich niet uitspreken op het internet. Dus ik had niet het idee dat ze dit begrepen. Laat staan mijn ouders, of andere mensen van een andere generatie."* Burger

Daarnaast speelt schaamte ook een rol in het niet willen of durven inschakelen van hulp:

> *"Ik schaamde me ook een beetje voor mijn eigen naïviteit, dat ik iemand op zijn eigen platform kritisch aan zou kunnen spreken. Dat is niet hoe platforms werken. De platforms zijn er om de makers zoveel mogelijk succes te gunnen en de mensen die kritisch zijn belachelijk te maken. Dat is nou eenmaal hoe onze social media aandachtseconomie-infrastructuur werken, dus ik was vooral beschaamd en ik dacht ook van ja ik wilde dit ook gewoon niet met mijn vrienden delen want het is ook gewoon heel dom."* Burger

We zagen al dat door plegers online geweld wordt gerationaliseerd, door te benoemen dat het online gebeurde, dus geen fysieke gevolgen heeft. Dit gebeurt ook aan de kant van de slachtoffers, die het online geweld afzetten tegen fysiek geweld. De gebeurtenis wordt bijvoorbeeld gerelativeerd:

> *"En ja, ik heb ergere dingen in het echte leven meegemaakt. Dus geen [online] doodsbedreiging, maar wel [fysiek] uitgescholden. Maar ik heb gewoon dat ik naar de supermarkt ga en dat iemand me dan opeens vastpakt. Dus zeg maar, de (online doods-) bedreigingen doen wel iets met me, maar als het niet reëel is, dan heb ik heftigere shit meegemaakt."* Burger

> *"Ik denk dat ik zelf dat zo goed in zo'n context kan plaatsen. En ik denk ook dat als mensen roepen ik maak je af online, dat dat 99% woorden zijn waar nooit echt acties aan zitten. Neemt niet weg dat de impact op iemands leven natuurlijk enorm kan zijn, maar ik kan dat wel even relativeren dat het vaak dan toch die toetsenbordhelden zijn die dat doen."* Burger

Of naarmate het vaker voorkomt, wordt het ook normaal en voel je het niet meer:

> *"Nou, aan het begin voelt het niet zo fijn. Stel dat er wel iets is dat ze mij kunnen vinden ofzo. Dat weet je maar nooit hè, in de huidige wereld. Maar op een gegeven moment wordt alles ook weer een beetje genormaliseerd. Dat is heel gek. Dan voel je het gewoon niet meer. Maar ja, dat soort bedreigingen, ik wil niet zeggen dat het normaal is, maar dat gebeurt wel."* Burger

5 Belemmeringen en kansen in de aanpak

Op basis van interviews met wetenschappelijke- en praktijkexperts, slachtoffers en plegers, twee focusgroepen met professionals en de online sessie, is in kaart gebracht welke maatregelen/ interventies momenteel plaatsvinden, waar knelpunten zitten en welke aanbevelingen er vanuit de ervaring worden gedaan om de aanpak van online geweld te verbeteren. Hierbij gaan we ook specifiek in op de behoeften van slachtoffers en plegers van online geweld.

5.1. Interventies & beleidsmaatregelen

Allereerst beginnen we met het onderstrepen dat online geweld tussen burgers onderling en de drie vormen waar we ons op richten nog een 'nieuw' en onderbelicht thema is (dit kwam ook al in hoofdstuk 1 ter sprake). Uit de empirie komt tevens naar voren dat interventies nog in de kinderschoenen staan en er nog niet/nauwelijks bekend is van wat werkt bij online geweld:

> *"Ja, ik denk toch dat we een soort palet of noem het een goocheldoos met interventies hebben die we zouden kunnen inzetten waar we dat nu dus niet hebben. Daar zijn we nu mee bezig om die proberen samen te stellen, zeg maar, maar we hebben zoveel dingen die we in de offline wereld kunnen inzetten en waar we aan kunnen denken. Maar er is zo weinig wat we weten dat we online kunnen inzetten. En dat vraagt natuurlijk ook een soort van gedachte, of ja, vooral verandering in hoe mensen werken en daar zijn we denk ik nu langzaam naartoe aan het werken. En als je dan zo'n trukendoos hebt, dat je je er ook bewust van kan maken van: maar dit zouden we kunnen inzetten. En dat is er gewoon niet voor ons. Ik denk dat er mondjesmaat overal in Nederland wel ergens wat is, maar het komt niet goed bij elkaar. Daar zijn wij nu heel erg druk mee om dat toch beter inzichtelijk te krijgen."* Praktijk expert

In de rest van dit hoofdstuk nemen we de vier niveaus voor interventies/beleidsmaatregelen uit hoofdstuk 2 over om zicht te geven op wat de respondenten ons vertelden over welke interventies er zijn, wat nog belemmeringen zijn en wat tips zij voor een effectieve aanpak van online geweld:

1. Individueel niveau
2. Gemeenschapsniveau
3. Wet- en regelgevingsniveau
4. Technologisch niveau

5.1.1. Individueel niveau

Bewustwording

In het vorige hoofdstuk kwam naar voren hoe slachtoffers en plegers ieder op een eigen manier omgaan met het online geweld. Een terugkerend punt in de gesprekken met slachtoffers, is dat ze zichzelf onder andere niet altijd als slachtoffer (willen) zien of geen hulp zoeken voor hetgeen wat ze hebben meegemaakt (zie ook hoofdstuk 3). Plegers normaliseren, rationaliseren of bagatelliseren het. *Bewustwording* wordt door de respondenten als belangrijk onderdeel van een effectieve aanpak gezien. Meer bewustwording over de impact van online geweld bij plegers, slachtoffers en de gehele samenleving zou een belangrijke stap zijn naar de erkenning van de gevolgen op individueel en maatschappelijk niveau.

Een respondent reflecteert op zijn ervaringen, legt uit dat hij geen hulp heeft geaccepteerd, en concludeert dat hij eigenlijk achteraf gezien toch wel graag had gehad:

> *"...omdat ik toen vooral in een overlevingsmodus zat, was ik vooral bezig met m'n omgeving, zorgen dat zij het oké hadden. En ik dus nog niet het besef had wat het met mezelf deed. En dat besef kwam pas echt één, twee maanden later waarbij het verwerkingsproces begon en toen heb ik gewoon wel twee weken lang nodig had om daar doorheen te komen met slapeloze nachten et cetera dus ja, daar had ik eigenlijk wel iemand voor moeten spreken. Maar goed dat heb ik helaas niet gedaan."* Burger

Onder respondenten waaronder de moeder van een dochter die met online geweld te maken kreeg en professionals/experts is behoefte aan het vergroten van de online weerbaarheid, bijvoorbeeld van kinderen en jongeren.

> *Maar ik zou echt oprecht prima vinden als dingen als social media tot 14, 16 jaar verboden zijn. Kinderen kunnen daar helemaal niet mee omgaan. Hersens zijn daar niet klaar voor. We kunnen er zelf al niet van afblijven van het doelloos scrollen. Hoe moet een kind dat kunnen. Dat kan niet. En als we dan met z'n allen die grens maar gewoon eens trekken, dan wordt het natuurlijk voor ouders ook wel makkelijker om daarop te handhaven. Heel veel ouders hebben ook gewoon geen idee wat een kind online uitspookt. En als er gepest wordt of geïntimideerd wordt of ja als iemand nooit wordt aangesproken op zijn negatieve gedrag en of dat dat nou in online of offline is, als een kind niet wordt aangesproken gaat hij dat gedrag niet veranderen, gaat hij die impulsen niet leren beheersen. Dus ook online moeten kinderen worden aangesproken."* Praktijk expert

> *In ieder geval moet er meer aandacht voor komen. Want kijk, mijn dochter die begon natuurlijk op een jonge leeftijd met een telefoon door corona. Ineens kregen kleine kinderen die eigenlijk te jong waren een telefoon, want alle lessen gingen via het internet. Maar niemand kreeg les in verkeersveiligheid. Ja, internetverkeersveiligheid dus."* Burger

In de praktijk blijkt het lastig om echt grip op bijvoorbeeld jongeren te krijgen en tot hen door te dringen.

> *Wij krijgen dus heel weinig grip op dat online. We proberen heel erg de jongeren individueel en vaak ook wel met familieleden iets te bieden, in hulp en ondersteuning. En dat zijn dan die coaches die dus wel met die jongeren daar ook over in gesprek gaan, wat ze doen en wat ze zien."* Praktijk expert

Melding of aangifte doen

Zowel slachtoffers als praktijkexperts geven aan dat meldingen en aangiften vaak uitblijven, ondanks dat ze meer zouden verwachten op basis van de prevalentie. De redenen hiervoor lopen uiteen. Sommige slachtoffers verwachten geen serieus vervolg op hun melding, in andere gevallen zagen ze hun ervaring niet als 'ernstig genoeg', angst voor represailles, bang voor onbegrip of victim blaming. Voor slachtoffers is het doen van een melding of aangifte soms ook een grote stap. Uit de interviews blijkt dat, wanneer een melding of aangifte niet tot concrete vervolgstappen leidt, het belangrijk wordt gevonden om de betrokkene erkenning te geven en duidelijkheid te bieden over het verdere proces. Dit draagt bij aan het behoud van vertrouwen en het gevoel van rechtvaardigheid in het proces.

> *En ik denk dat je dat ook echt wel ziet in deze vorm van, dat er ook een bepaalde, het werd al wel eerder benoemd, schaamte bij zit. Cyberschaamte. Mensen zullen misschien niet altijd het verhaal willen vertellen wat hun is overkomen. Want misschien denken ze van ja, dan ben ik zwak. Of dan ben ik... Dat zijn wel aspecten die wel meespelen. En ik denk dat ook heel veel mensen nog steeds niet het besef hebben. Dat bepaalde vormen die er zijn, dat die ook strafbaar zijn. En die echt wel de grens overgaan. Dat je echt gewoon strafbare gedragingen hebt".* Wetenschappelijk expert

De respondenten zien nog genoeg kansen liggen om de aangifte- en melding bereidheid te verbeteren. Tegelijkertijd geven ze aan dat het van belang om altijd een afweging te maken, of het doen van aangifte/meldingen wel altijd het juiste middel is om het complexe fenomeen van online geweld aan te pakken:

> *We hebben het ook vaak over strafrechtketen om die te belasten. Dus daar zie je ook vaak data dat je heel goed moet gaan afwegen, al in een vroeg stadium. Is aangifte doen nou het beste middel. Om het probleem wat je hebt op te lossen. En is er niet een alternatieve interventie mogelijk?"* Praktijk expert

Laagdrempelige vormen van hulp tijdens en na incident

Verder hebben (vooral jongere) slachtoffers voorkeur voor laagdrempelige, mogelijk anonieme vormen van het inschakelen van hulp. Een respondent geeft de voorkeur aan het delen van zijn hulpvraag met een organisatie waar hij anoniem zijn verhaal kan doen, boven het vertellen aan zijn ouders:

> *Omdat je dan toch dat vertelt aan iemand die je vertrouwt en naast je staat. En als ze daar een mening over hebben, ja, dat voelt gewoon veel zwaarder dan iemand waarmee je het van een afstandje bespreekt, zeg maar."* Burger

Respondenten geven aan dat jongeren vaak liever niet bellen, maar wel willen chatten.

> *Want ze gaan nou eenmaal niet bellen. Dus je ziet dat ze dan eigenlijk het liefst willen chatten met iemand al dan niet anoniem. Ze zijn op zoek naar iemand bij wie ze hun verhaal kunnen doen. We moeten ook kijken welke andere kanalen we jongeren kunnen aanbieden die misschien niet die mogelijkheid hebben om met hun ouders daarover in gesprek te gaan."* Deelnemer focusgroep

Een andere respondent geeft ook aan dat ze liever haar broertje om hulp vroeg dan haar ouders, omdat de leefwereld van haar ouders niet helemaal zou aansluiten bij haar leefwereld. Haar gegevens waren in een Telegramgroep gedeeld, waarna ze geïntimideerd werd door verschillende nummers.

> *“Het zat me wel heel erg hoog en dat zag mijn broertje ook. Maar ik heb het niet tegen mijn ouders gezegd, die weten het nog steeds niet. Ik weet niet, dat voelt niet zo heel fijn ofzo als ze het weten. (...) Ja, ik weet niet zo goed of ze het zouden begrijpen. Want zij zijn natuurlijk helemáál niet opgegroeid met social media en dat soort dingen.* Burger

Erkenning

Ook speelde bij de vorige respondent haar angst voor victim blaming mee. Zij had op dat moment geen oordeel nodig, maar vooral praktische hulp en erkenning:

> *“ En ook, (als ik het aan mijn ouders vertel) voel ik me ook weer een beetje schuldig of zo, heb ik het idee. Dat het dan, ja, weer is van ‘dom dat je je nummer weggeeft’ en ik denk dan dat er misschien ook een licht oordeel is, die ik op dat moment niet nodig had. Bij mijn broertje is dat totaal niet, ja die zegt dan veel later van ‘nou ja volgende keer niet meer je nummer geven’. Maar ja, dat wist ik zelf ook wel.”* Burger

Motivatie

Het zou helpen om meer zicht te krijgen op de motivatie van plegers, want daar kan je uiteindelijk ook op interveniëren.

> *“ Ja, ik denk dat bij daders is er altijd de vraag: wat motiveert ze? En er moet iets met die motivatie gaan gebeuren. Maar ik denk ook echt dat we na moeten denken over veilig sociale media gebruiken en internetgedrag.”* Wetenschappelijk expert

5.1.2. Gemeenschapsniveau

Verantwoordelijkheid, bewustwording én preventie

Op gemeenschapsniveau is de rol van bewustwording én preventie volgens respondenten erg belangrijk. Burgers en respondenten geven aan dat op het gemeenschapsniveau, zeker als het gaat om jongeren, de primaire verantwoordelijkheid bij ouders en opvoeders ligt, met name op het gebied van preventie. Hoewel praktijkexperts een belangrijke rol kunnen spelen in signalering en begeleiding, geven zij aan slechts beperkt zicht te hebben op wat zich online in de privésfeer van jongeren afspeelt. Ouders hebben die volledige grip evenmin, maar kunnen wel het verschil maken door actief het gesprek aan te gaan met hun kinderen over hun online gedrag en -leefwereld. Net zoals we kinderen leren om te zwemmen of veilig deel te laten nemen aan het verkeer, rust er ook een maatschappelijke verantwoordelijkheid om (toekomstige) burgers basisvaardigheden bij te brengen voor veilig en bewust online gedrag. Een normstelling zou volgens respondenten gestimuleerd kunnen worden waarbij digitale weerbaarheid een vanzelfsprekend onderdeel zou worden van opvoeding, onderwijs en jeugdwerk.

Onderwijs

Scholen kunnen een aanvullende rol spelen bij preventie en voorlichting, al wordt door professionals tijdens de focusgroep benoemd en erkend dat er al veel op het bordje van leraren wordt geschoven. Digitale omgangsvormen zouden bijvoorbeeld wel geïntegreerd kunnen worden in het burgerschapsonderwijs, waarbij het niet alleen gaat om regels over gedrag op sociale media, maar vooral om het stimuleren van bewustwording. Jongeren ervaren de online en offline wereld namelijk niet meer als gescheiden domeinen, maar als één verweven leefomgeving.

Naast onderwijs aan jongeren, gaat het ook om onderwijs aan professionals.

> *“Ik denk alles met onderwijs, waar je grote groepen mensen bereikt, en ook hierover voor kunt lichten. Al die dingen die ik noem, maar ik denk ook in de opleiding van de sociale professional moet hier veel meer aandacht voor zijn. Überhaupt van alle professionals. Maar ik denk ook bijvoorbeeld de hulpverleners die wij opleiden, die moeten hier meer van horen. En die denken heel van ja, die techniek het zal allemaal wel, maar heel veel problemen met mensen die wij later willen helpen, hebben hiermee te maken. Dus je moet dan wel van de hoed en de rand weten. Dus in de opleiding van professionals moet dit veel meer aandacht krijgen. En dan denk ik, breed hè, dan denk ik aan hulpverleners, maar dan denk ik ook aan de politie.”* Wetenschappelijk expert

Gehele hulpverleningsketen

Wetenschappelijke en praktijkexperts pleiten er daarom voor dat het besef van die verwevenheid breder doordringt in de gehele hulpverleningsketen. Het vraagt van professionals dat zij digitale leefwerelden serieus nemen en niet langer als een apart of secundair fenomeen behandelen. In die context worden *digitale jeugdagenten en -jeugdwerkers* positief benoemd: zij spreken de taal van jongeren, herkennen online dynamieken en kunnen de brug slaan tussen digitale signalen en de juiste hulpverlening.

5.1.3. Wet- en regelgevingsniveau

Informatiepositie en deling

Een veelgenoemde belemmering vanuit de praktijkexperts is de onduidelijkheid wat ketenpartners, zoals scholen, hulpverlening, politie en ambtenaren, wel of niet met elkaar mogen delen. De AVG wordt in die context genoemd, waarin niet altijd duidelijk is wat de grenzen zijn van wat wel en niet mag volgens de privacywet. Hierdoor blijven signalen soms te lang binnen één organisatie hangen.

> *“Het is veel lastiger om online die informatiepositie goed te houden dan in de fysieke wereld, want in de fysieke wereld stonden ze gewoon op de hoek van de straat, en dan kon je daar op de hoek van de straat gaan kijken.”* Praktijk expert

Onwetendheid wat wel en niet mag

Onder professionals en experts is er nog veel onduidelijkheid over wat wel en niet mag volgens de regels/richtlijnen. De manier waarop iedereen daarmee omgaat is in de praktijk ook nog erg divers.

> *“… eigenlijk heeft iedereen wel een puzzelstukje. Want de een heeft weer een clipje gezien uit eigen interesse, of juist omdat een jongen dat zelf liet zien. De ander heeft juist zonet iet anders gehoord dus je ziet dat zij zeg maar de puzzel compleet maken om het soms online wel inzichtelijk te hebben. We zijn er wel echt mee bezig wat we online kunnen doen. Maar je merkt ook dat de ene ketenpartner vindt het normaal om iemand zeg maar te volgen en zegt dat ook: ‘ik wil graag je snap’. En de andere denkt: ‘ja, hoort dit wel bij mijn taak?’ Dus er is ook een heel grote onwetendheid wat nou mogelijk is en wat mag, als je het hebt over online.”* Praktijk expert

Waar de een wat terughoudender en soms ook wel te voorzichtig is (wat soms ook gepaard gaat met continu moeten verantwoorden waarom je iets wilt doen) is, is de ander juist meer op zoek naar de randen van de wet in het belang van een veilige samenleving. Er is vooral behoefte aan duidelijke wetgeving voor politie en gemeenten[11]: wat mag én kan wel?

11 Vanaf juli 2025 ligt er een wetsvoorstel bij de Tweede Kamer Wet online aangejaagde openbare-ordeverstoringen. Steeds meer verstoringen van de openbare orde beginnen online. Burgemeesters zien rellen tegenwoordig in hun gemeente vaak aankomen, omdat er online wordt opgeroepen om bijvoorbeeld de straat op te gaan. Burgemeesters kunnen nu vaak pas ingrijpen als alle ellende al heeft plaatsgevonden of plaatsvindt. De indiener wil dat burgemeesters een verwijderbevel kunnen opleggen aan plaatsers van online berichten die de openbare orde verstoren of waardoor ernstige vrees bestaat voor het ontstaan van een openbare-ordeverstoring (Overheid.nl | Consultatie Wet online aangejaagde openbare-ordeverstoring).

Bewijslast en gebrek aan casussen

Professionals lopen tegen de bewijslast van online geweld aan. Dit is in de praktijk soms erg complex vanwege verschillende redenen:

> *Het is vaak één op één. Je hebt meerdere getuigen nodig. Helaas ook een stukje capaciteitsprobleem. Prioriteit waar je mee te maken hebt. Dat is een afweging die van tevoren wordt gemaakt."* Deelnemer focusgroep

> *"Waar wij nu tegenaan lopen is veel via Snapchat. Dus alles loopt via Snapchat. En dan zit je ook met de bewijsbaarheid om dat bij justitie te krijgen. En dat is dan vaak ingewikkeld."*
>
> Deelnemer focusgroep

Ook slachtoffers geven aan dat de bewijslast een lastig obstakel is. Het versterkt ook het gevoel van onmacht maar ook hangt het samen met het gebrek aan privacy:

> *Dat er niets is wat je online kan doen als iemand bijvoorbeeld zegt dat je anaal verkracht moet worden. Dat kan je rapporteren, maar als tienduizend mensen dat zeggen, dan is het gewoon niet mogelijk om dat te doen. Dan is het niet mogelijk om een filmpje offline te krijgen waarin je zelf wordt afgebeeld waarvan je weet dat iedereen daar doodsbedreigingen onder zet, zonder de maker op de hoogte te stellen van wie je bent. Je hebt geen privacy als je een rapportering doet op social media. Dus als slachtoffer, om maar even dat rotwoord te gebruiken, ben je compleet onbeschermd. En alle moeite en alle bewijslast ligt bij jou, als slachtoffer. Om jezelf te beschermen of om aan te tonen dat iets haatdragend is."* Burger

De matige bewijslast maakt het soms ook moeilijk om zaken te kunnen opvolgen/vervolgen. Hierbij speelt het ook een rol dat de incidenten internationaal zijn:

> *En ook hier geldt weer en wat ik ook al aangaf, het komt vanuit het buitenland. Dus als je wat strafrechtelijk zou willen, dan is het eigenlijk al redelijk kansloos. Dus je moet bij dit soort dingen ook echt inzetten op preventie. Je moet inzetten op bewustwording. Wat mag nou wel, wat mag nou niet."* Deelnemer focusgroep

Überhaupt geven professionals aan dat het aantal zaken (aangiftes/meldingen) dat binnenkomt over online geweld nihil is (zie ook hierboven). Het gebeurt maar dit zien ze niet terug in de casussen, terwijl ze bijvoorbeeld wel weten dat het onder jongeren vaak voorkomt. Bijvoorbeeld bij Halt komen casussen van online geweld niet of nauwelijks binnen. Maar ook andere professionals herkennen dit. Er wordt aangegeven dat dit waarschijnlijk door de lastige bewijslast op het bordje van bijvoorbeeld Jeugdzorg komt, maar dat de druk op de jeugdzorg ook erg hoog is waardoor ze mogelijk weinig tijd hebben om dit verder op te pakken.

Ver stadium

Verschillende professionals geven ook aan dat wanneer online geweld boven de radar komt, de situaties meestal is geëscaleerd of al in een vergevorderd stadium zit. Dit maakt de aanpak dan ook complexer.

> *Als er in de online wereld jongeren af dreigen te gaan naar wel strafbare gedragingen. Dan is dat eigenlijk bijna nooit zichtbaar. Dus met andere woorden, een mooie vergelijking die wel eens werd gezegd. Als iemand op straat een prullenbakje in elkaar staat te trappen. Dan word je gecorrigeerd en dan zeg je dat mag niet. En als het helemaal is dan kan je zeggen die gaat nu naar Halt om dat goed te doen. Maar als iemand in de online wereld afdrijft... is er bijna niemand die al in een vroeg stadium kan corrigeren. Vaak zitten ze dan al heel ver."* Deelnemer focusgroep

Internationaal

Een fundamentele uitdaging in de aanpak van online geweld is dat wat online plaatsvindt, zich niet houdt aan nationale wetten of structuren. Plegers kunnen internationaal opereren, ook de platforms waarop het online geweld zich afspeelt opereren vaak vanuit een internationale context. Een Europese aanpak blijft noodzakelijk, maar respondenten benadrukken ook het belang van samenwerkingsverbanden tussen de nationale overheid en techbedrijven, om de relatie op nationaal niveau te versterken en afspraken te kunnen maken.

Strengere regelgeving social-mediagebruik bij jongeren

Daarnaast komt de wens naar voren, bij professionals zowel als burgers, voor strengere regelgeving rondom het sociale-media-gebruik bij jongeren. Zo benoemen meerdere respondenten een leeftijdsgrens, bijvoorbeeld een verbod op social media gebruik tot 14 of 16 jaar, als maatregel om jonge kinderen beter te beschermen.

5.1.4. Technisch niveau

De rol van sociale media platforms en technologische interventies

Tijdens dit onderzoek is stilgestaan bij de verschillende rollen die sociale mediaplatforms spelen in de aanpak van online geweld. Die rol is niet eendimensionaal, maar bestaat uit drie samenhangende pijlers: moderatie van content, technologische interventies, en samenwerking met overheid en maatschappelijke partijen bij preventie en handhaving. Allereerst vervullen platforms een belangrijke rol in de moderatie van schadelijke content, zoals haatzaaien of online intimidatie. Automatische detectie van dergelijke berichten wordt in de praktijk vaak gezien als een noodzakelijk instrument. Tegelijkertijd zijn er duidelijke kanttekeningen. Algoritmes zijn niet waterdicht; veel schadelijke berichten weten door de filters heen te glippen, bijvoorbeeld doordat gebruikers kleine aanpassingen doen aan aanstootgevende woorden. Ook het zogenoemde waterbedeffect blijft een hardnekkig probleem: zodra moderatie op een groot platform toeneemt, wijken gebruikers eenvoudig uit naar minder gereguleerde of alternatieve platforms (zie ook hoofdstuk 3). Daarnaast experimenteren sommige platforms met technologische interventies die gericht zijn op het voorkomen van grensoverschrijdend gedrag. Zulke normerende technologieën kunnen bijdragen aan de preventieve aanpak van online geweld, zonder direct te censureren. Een derde rol die van platforms wordt gevraagd, is actieve samenwerking met overheden, maatschappelijke organisaties en experts. Alleen door gezamenlijk beleid, kennisdeling en handhaving kan effectieve bestrijding van online geweld plaatsvinden. Juist op dat vlak bestaan zorgen: bij enkele grote sociale mediabedrijven neemt de inzet op moderatie en samenwerking af, onder meer door kostenbesparingen of politieke motieven. Denk hierbij aan X (voorheen Twitter), Meta en Telegram. Deze ontwikkelingen onderstrepen de noodzaak van een stevige, grensoverschrijdende aanpak op Europees niveau, in combinatie met een sterke, integrale samenwerking tussen platforms, overheid en maatschappelijke partijen (zie ook paragraaf 5.2).

Sommige platforms experimenteren met nieuwe technologische interventies die gericht zijn op het stimuleren van sociaal wenselijk gedrag. Zo heeft TikTok in een aantal landen de zogeheten *empathy-knop* ingevoerd. Wanneer een gebruiker een reactie plaatst die door kunstmatige intelligentie wordt herkend als mogelijk kwetsend of grof, verschijnt een pop-upmelding. Hierin wordt de gebruiker aangespoord nog eens kritisch naar de toon of inhoud van het bericht te kijken, met een oproep tot empathie en respect. Deze subtiele vorm van normering werkt zonder directe censuur en onderzoek toont aan dat het effect kan hebben: een deel van de gebruikers past het bericht aan of besluit het uiteindelijk niet te plaatsen.

Ook heeft TikTok het European Safety Forum opgericht, waarin experts, beleidsmakers, maatschappelijke organisaties en vertegenwoordigers van TikTok samenwerken aan het verbeteren van de veiligheid op het platform. Hiermee benadrukt TikTok dat het tegengaan van online geweld een gedeelde verantwoordelijkheid is, waarbij samenwerking tussen platform, samenleving en beleidsmakers essentieel is. Een vergelijkbaar voorbeeld komt van NU.nl, dat de afgelopen jaren werk heeft gemaakt van het verbeteren van de reactieomgeving onder nieuwsberichten. Hoewel het platform anonieme reacties toestaat om de drempel tot deelname laag te houden, leidde een gebrek aan duidelijke huisregels en moderatie aanvankelijk tot een verharding van het debat. Na aanscherping van de huisregels, strengere moderatie en inzet van een multidisciplinair moderatieteam is het reactieklimaat merkbaar verbeterd, zonder dat de anonimiteit volledig is afgeschaft. Dit laat zien dat ook nieuwsmedia zich bewust zijn van hun verantwoordelijkheid in het begrenzen van online geweld.

Ondanks deze positieve voorbeelden signaleren zowel praktijk- als wetenschappelijke experts een structureel gemis aan effectieve handhaving door grote sociale mediaplatforms. De vrijblijvendheid waarmee sommige platforms hun verantwoordelijkheid invullen, bemoeilijkt het creëren van een veilige online omgeving.

De rol van gemeenten

Ook gemeenten krijgen steeds vaker te maken met de complexiteit van online moderatie. Er wordt gewerkt aan protocollen voor online monitoring die gemeenten houvast moeten bieden, maar in de praktijk blijkt het onderwerp ingewikkeld en gevoelig te liggen. Bestaande instrumenten zoals de Wegwijzer Grip op Informatie, Handreiking gebruik social media en de Gesprekstarter Gebruik social media door gemeenten zijn niet bij iedereen bekend.[12] Niet alle gemeenten zijn zich voldoende bewust van de juridische grenzen. Het komt voor dat ambtenaren zonder zich kenbaar te maken deelnemen aan besloten online groepen, zoals Telegram, om informatie te verzamelen of door te spelen.[13] Dit is in strijd met de geldende regels. Er is daarom behoefte aan meer

12 Wegwijzer Grip op Informatie april 2022.pdf; 20141208-8a-handreiking-gebruik-social-media.docx; VNG rapport.

13 Zorgen om corrupte ambtenaren: 'Georganiseerde misdaad kan niet zonder ze'.

bewustwording en vooral aan duidelijke, toepasbare richtlijnen voor gemeentelijke medewerkers. De bestaande juridische kaders zijn vaak abstract en roepen vragen op over de concrete toepassing. Zo geldt bijvoorbeeld dat een ambtenaar zich altijd moet identificeren bij deelname aan besloten online groepen, maar de praktische invulling daarvan blijkt in de praktijk lastig.

Bovendien speelt het bredere vraagstuk van de verschuiving naar minder gereguleerde, alternatieve platforms *(alt-tech*). Zelfs als de moderatie op bekende platforms verbetert, blijven deze uitwijkmogelijkheden bestaan. Dit creëert een glijbaan van steeds minder toezicht naar volledige ongereguleerde online omgevingen, met alle risico's van dien, zoals de verspreiding van haatzaaien, desinformatie en deepfakes. Tegelijkertijd zijn er positieve ontwikkelingen. Europese regelgeving en de stijgende hoogte van boetes voor techbedrijven kunnen een stimulans vormen voor strengere moderatie en betere naleving. Toch bestaan er zorgen, bijvoorbeeld over het afnemende moderatiebeleid bij grote platforms als Meta en de opkomst van geavanceerde misbruikvormen zoals deepfakes. Samenvattend staan gemeenten voor de uitdaging hun rol bij online moderatie verder te professionaliseren. Dat vraagt om heldere juridische kaders, praktische richtlijnen, meer bewustwording binnen gemeenten én een gezamenlijke aanpak met andere overheden en maatschappelijke partners.

> *Eigenlijk, ja, die doen er eigenlijk vrij weinig aan. Ze moeten niks. In Europa proberen ze er misschien iets mee te doen om ze een bepaalde kant op te dwingen, maar dat komt gewoon totaal niet van de grond. Ik heb vooral een ervaring bij gemeentes. Als je dit soort verhalen hoort, wat kun je? Ik zou het niet weten. Dat maakt het voor mij heel complex hoe dit op te lossen of hoe deze mensen te helpen."* Praktijk expert

> *Omdat het gewoon voor ons ook vaak wel een black box is en het berustte heel vaak nog op toevalligheden merk ik. Wat natuurlijk super is, maar je wil natuurlijk eigenlijk je aanpak zo inrichten dat waar ik nu inzet op een gebied, dat dat juist het online ook heel bewust meeneemt. En dat is voor ons echt nog heel lastig. Ja, ik ben eerlijk, vind ik echt nog heel lastig. Dus ja, we proberen zoveel mogelijk bij andere gemeenten te vragen, maar ook bij andere ketenpartners wat zij doen. Elk haakje dat we hebben om online toch te benoemen doen we. En ook steeds bij ministeries neerleggen dat we heel erg zoekende zijn. Maar het antwoord hebben we niet. Het is heel lastig."* Praktijk expert

5.2. Integrale samenwerking

5.2.1. Belemmeringen

Versnippering, geen eigenaarschap, gebrek aan capaciteit en signalen blijven liggen

Effectieve aanpak van online geweld vraagt om een integrale samenwerking tussen ketenpartners zoals scholen, jeugdhulpverlening, politie, justitie, maatschappelijke organisaties en techbedrijven. Dit gebeurt in de praktijk al wel, maar deze samenwerking blijkt volgens de respondenten echter ook vaak versnipperd. De experts geven aan dat de veelzijdigheid en onderlinge wisselwerking van de verschillende vormen van online geweld het voor ketenpartners extra lastig maakt om op te treden, mede door de versnippering binnen de keten. Politie, hulpverlening en andere instanties beschikken elk slechts over een deel van de informatie, waardoor het totale plaatje moeilijk zichtbaar wordt. Incidenten worden vaak als 'bedreiging' of 'intimidatie' geregistreerd zonder expliciete vermelding van de online component. Hierdoor blijven de digitale wortels van het probleem onzichtbaar en blijft gerichte preventie of handhaving uit. Het leidt er ook toe dat organisaties en partijen eigenaarschap missen.

Intern zijn er soms ook wat belemmeringen zoals een gebrek aan capaciteit of prioritering. Een moeder vertelt over de situatie rondom haar dochter dat er na het incident op school (zie ook casus school) de school aanvankelijk aangaf niks te gaan doen:

> *En het was gebeurd op de vrijdagochtend, we hebben e-mails gestuurd op zaterdag naar school dat het echt niet kan en het gedrag van de kinderen. En op maandagochtend kregen wij gelijk te horen van ja, maar wij gaan hier niks mee doen. Wij gaan dit niet aanpakken. En dan denk ik ja, maar hoe kun je dit nou niet aanpakken? … Er is een politieagent uiteindelijk in de klas gekomen. Die heeft natuurlijk een gesprek gehad met de kinderen. Maar dat maakte eigenlijk geen indruk. Want de dag daarna kreeg mijn dochter een sms'je met... "Ja, waarom heb je de politie betrokken? De verkrachting was maar een grapje hoor."* Burger

Gebrek aan capaciteit en expertise en deskundigheid

Niet alle partijen beschikken over voldoende bewustzijn, kennis en expertise van digitale dynamieken. Respondenten geven aan dat ketenpartners, zoals politie en jeugdzorg, de kennis en tools missen om effectief op te treden tegen online vormen van criminaliteit, met name in snel bewegende contexten zoals de drillrapscene.

> *Mij valt op dat heel veel expertise zit vooral heel erg apart op dat online. Er wordt niet altijd de relatie met offline gelegd. En ik denk dat er nog steeds heel veel behoefte is aan technische expertise, omdat er nog wel wat juridische vragen aanzitten. 'Waar staat de server en hoe spoor je het op, hoe leg je sporen vast?' Ik denk dat er ook voor een bredere groep professionals nog heel veel behoefte is aan technische kennis…"* Wetenschappelijk expert

> *Nou ja, kijk, bewijsvoering is ontzettend belangrijk, want zonder bewijs ben je nergens, dus dat moet heel goed geborgd zijn. En dit zijn nu nog echt toch nog wel specialismen. Dit zit nog niet in het generieke. Je krijgt natuurlijk ook jongere mensen die er al meer affiniteit met en kennis van hebben, maar de vraag is… ik denk dat je specialisten keihard nodig hebt, maar dit moet ook bij de generalisten terecht gaan komen. Ik denk dat het ook logisch is om hierop door te vragen. Ik denk dat daar nog veel moet verbeteren."* Wetenschappelijk expert

> *Wat ik wel merk is inderdaad dat er heel veel tekort is in capaciteit. En dat je eigenlijk overal een soort van online aanpak moet hebben. Maar dat eigenlijk bij heel veel ambtenaren kennis ontbreekt, capaciteit ontbreekt."* Deelnemer focusgroep

Stigma en houding

Daarnaast horen we terug dat soms de houding/visie of het stigma dat er bestaat rondom online geweld nog in de weg staat.

> *En wat ook niet helpt, is dat er een soort van stigma is rondom digitale veiligheid. Er wordt heel vaak gedacht, dat is technisch. Maar heel vaak, ik merk dat heel veel collega's ook een beetje zich ervan houden, dat ze denken, 'oeh, maar dat gaat mijn pet te boven'. Dus dat helpt denk ik niet in het iedereen meenemen van, we moeten ook online zichtbaar worden. Eigenlijk alle beleidsdomeinen hebben een soort van online component, maar daar lijken we een soort van in gemeenteland, ons verre van te willen houden".* Deelnemer focusgroep

Naast het stigma dat er soms rondom online geweld en digitale veiligheid heerst, is ook de positie en houding van verschillende stakeholders ten aanzien van dit vraagstuk van invloed." *Scholen zeiden aan het begin echt twee jaar geleden tegen mij. 'Ben jij gek? We gaan ons niet bemoeien met die WhatsApp›s. Want we moeten gewoon lesgeven. We moeten gewoon dit.' Ik snap dat jullie druk zijn. Maar met deze generatie kinderen vanaf groep 6, 7, 8 speelt ook dat online stuk heel erg mee. En dan doen ze even een training media wijsheid of zo. Dan halen ze even voor 2000 euro van een subsidie iets voor de klas voor één keer. En dan denken ze oké check. Maar we doen wel wat dan online. Dan denk je, ja maar wat doe jij nou als juf of meester als er zo›n groepsapp gaande is? Zeg jij dan, als school wil ik daar in of wil ik daar niet in? Dat zijn allemaal vraagstukken. Daar zijn ook die beleidsafdelingen nog helemaal niet op ingericht. Dus ik merk gewoon dat het lijkt net alsof het online domein ons heeft overvallen vanuit alle beroepsgroepen. En dat we daar nog niet goed in onze eigen plek weten te pakken of zo. En dat is echt stoeien in de praktijk."* Deelnemer focusgroep

Belang van gedragsverandering

Het gaat om het samenbrengen van allerlei domeinen in de aanpak online geweld. In de praktijk zijn hier echter nog wel wat hobbels.

> *Dus als je professionals ook wil meenemen in een andere manier van werken, dan is het ook gedragsverandering. Dus daar zul je langer de aandacht op moeten hebben. Het hebben van handelingsperspectieven en de juiste beschrijving is nog niet het gedrag tonen in de praktijk. Om het daar ook toe te gaan passen, dat moet je nog een laagje dieper..."* Deelnemer focusgroep

5.2.2. Kansen en verbeteringen

Structurele en domeinoverstijgende samenwerking

Een structurele, domeinoverstijgende samenwerking is nodig om zowel de huidige preventie, de signalering als de opvolging van online geweld te versterken. Tegelijkertijd merkt een deelnemer op dat het een lastig vraagstuk blijft of iedereen nu inderdaad een rol hierin moet innemen, of dat het goed is om meer gericht te gaan werken.

> *Dus iedereen heeft daar eigenlijk een rol in. En aan de andere kant, voel ik soms ook wel een rem. Moeten we hier nu met z'n allen volledig opduiken? Of moeten we daar toch een soort nuance in vinden. En gewoon heel slim gaan kijken. Oké maar wie doet wat?"* Deelnemer focusgroep

Integraal onderdeel

Om tot een effectieve aanpak te komen, is het cruciaal dat online geweld niet als een op zichzelf staand probleem wordt benaderd. Het *onderscheid tussen online en offline* is voor veel jongeren en volwassenen achterhaald: beide leefwerelden zijn nauw met elkaar verweven. Wanneer beleid, richtlijnen of interventies uitsluitend zijn gericht op fysieke vormen van geweld of afzonderlijk digitale veiligheid behandelen, ontstaat het risico dat slachtoffers tussen wal en schip vallen. Ketenpartners en beleidsmakers onderstrepen daarom het belang van het *integraal opnemen van de online leefwereld in bestaande beleidsplannen/programma's, veiligheidsstrategieën, zorgstructuren en onderwijsprogramma's*. Dit vereist niet alleen inhoudelijke kennis, maar ook gedeelde verantwoordelijkheid, duidelijke taakafspraken en ruimte voor professionele ontwikkeling. Hierbij wordt ook de koppeling aan bestaande programma's als veelbelovend gezien, bijvoorbeeld die van de integrale aanpak op onlinefraude.

> *„ Daar is al vanuit het Ministerie van Justitie die integrale aanpak op online-fraude. Die bestaat al, dus eigenlijk zou je ook zoiets moeten hebben.„*
>
> Deelnemer focusgroep

Andere programma's waarin een online aanpak een link mee kan leggen zijn bijvoorbeeld de projecten Online Content Moderatie maar ook bestaande initiatieven/organisaties zoals Help Wanted of Fier.

Leren van de werkwijze bij traditionele criminaliteit

Ook geven professionals aan dat ook veel geleerd kan worden van de publiek-private samenwerking die tot stand is gekomen met betrekking tot de traditionele criminaliteit. Bijvoorbeeld de persoonsgerichte aanpak waarbij zowel wordt ingezet op preventie als repressie. Maatschappelijke verantwoordelijkheid van allerlei verschillende partijen heeft daarin ook een belangrijke rol:

> *“ Wat wij in de aanpak van de traditionele criminaliteit ook wel gedaan hebben, is dat als er partijen zijn die hun verantwoordelijkheid niet nemen, hun maatschappelijke verantwoordelijkheid als grote onderneming, we daarmee in gesprek gaan en als het moet zelfs dwingen om maatregelen te nemen. Marktplaats bijvoorbeeld, op heling -een van mijn landelijke portefeuilles – waren ze in het beginsel niet zo genegen om iets te gaan doen aan heling - het aanbieden van gestolen goederen op hun marktplaats - maar uiteindelijk door het goede gesprek aan te gaan, hebben we dat toch voor elkaar gekregen. Het maakt het wel lastiger omdat je veel eigenlijk alleen maar met buitenlandse partijen te maken hebt. Maar het aanspreken van grote bedrijven zoals deze platforms op hun maatschappelijke verantwoordelijkheid, dat is ook echt wel een hele belangrijke.”* Deelnemer focusgroep

Eigenaarschap en trekker/aanjager

Er is behoefte aan eigenaarschap en iemand die zich ontfermt als trekker/aanjager van een online geweld aanpak. Hierbij wordt het Ministerie van Justitie en Veiligheid genoemd:

> *“ De publiek-private samenwerking is daarin belangrijk, maar er moet natuurlijk altijd wel een trekker zijn of een aanjager. Ik vind wel dat daar bij het ministerie wel vooral ook een rol ligt om alle partners daarin te verzamelen en daarin de goede stappen te nemen.”* Deelnemer focusgroep

De rol van het Ministerie wordt vooral gezien op tactisch strategisch niveau.

> *“ Wel als het gaat om prioritering en aansturing, maar ook lokaal zal je toch een en ander moeten doen. Je wilt wel een overkoepelende factor hebben, maar uiteindelijk moet je het ook lokaal wel met elkaar doen.”* Deelnemer focusgroep

Doordat er aandacht is bij het Ministerie wordt het ook op tactisch operationeel niveau makkelijker om dingen uit voeren:

> *“ En dat vooral op gemeentelijk niveau. Maar ook daar is de gemeente de regiehouder op veiligheid. Dus laat de regie daar ook vooral liggen. Het is echt belangrijk, vooral als je publiek-private samenwerking doet, dat iemand de regie heeft, om het overzicht te houden, de aanjager is, de mensen bij elkaar brengt, de faciliterende rol heeft, en dat is op een strategisch niveau het ministerie, en tot aan het operationele niveau de gemeente, politie, het openbaar ministerie, die daarin participeren integraal in de driehoek, en daar dan alle partijen bij trekken”* Deelnemer focusgroep

Vroegsignalering

Meer inzetten op *vroegsignalering* en de rol die verschillende partijen daarbij kunnen spelen, zou de aanpak ook versterken. Bijvoorbeeld door het werk van digitale wijkagenten die al in verschillende politieregio's hun werk doen. Halt heeft bijvoorbeeld ook een dubbele rol, bij de repressie maar ook bij preventie.

> *Er zijn ook HALT-medewerkers die standaard aansluiten voor schoolinterventies op scholen. Ja, dan kan je het al opvangen bij de eerste signalen, althans dat hoop je. Dan is die lijn heel kort".* Deelnemer focusgroep

Allerlei partijen kunnen hierin een rol spelen zoals gemeenten, jeugdzorg, gezinswerkers, maatschappelijk werkers, leraren/scholen, sportverenigingen etc.

" In de gemeente … zijn we bezig met een pilot vroegsignalering. Ik vraag dan ook aan de gezinscoaches, hebben jullie een plan van aanpak online domein? Zo makkelijk is het. Je kan je plan van aanpak aanpassen en met een ouder ook je standaard uitvraag doen, heeft je kind een mobiel? Heeft je kind een Playstation? Wat zijn zijn social media accounts?"

Deelnemer focusgroep

Heel vaak wordt volgens respondenten een partij vergeten: namelijk de ouders. Die kunnen een belangrijke rol spelen maar in de praktijk blijkt dit vaak ook lastig.

> *Ik hoor je allemaal partijen noemen. En ik denk partijen die we heel vaak vergeten. Want wij wijzen heel vaak inderdaad naar docenten. Maar de docenten zitten bomvol met gewoon hun primaire taak lesgeven. Maar een hele belangrijke partij die we niet benoemd hebben... zijn de ouders van al die kinderen. Die net zo goed niet weten wat ze hiermee aan moeten. Die ook maar denken... 'oh, mijn kind is inderdaad blijkbaar in groep 6 of 5 al. Oh, geef me een telefoon'. Maar waar we wel bij onze eigen kinderen vragen... 'hoe was het op school en met wie speel je eigenlijk?' En je ook enigszins wel zicht hebt op wie er thuis... met wie ze omgaan, zien we dat online niet. En eigenlijk is het ook wel iets om misschien uit te leren... dat het ook gewoon net zo normaal moet worden om te vragen, wat doe je eigenlijk online? Met wie heb je daar contact? Maar ook om misschien gesprekken te voeren over, wat doe jij online en hoe ga je dat gesprek aan? Ik heb bijvoorbeeld één collega, die gaat een beetje tot het extreme door, maar die heeft wel een soort van regels opgesteld van, 'nou, wat jij offline niet durft te zeggen, doe je online ook niet. En je gaat nooit, nooit foto›s doorsturen.' En nog een soort van heel contract heeft ze eigenlijk echt letterlijk opgesteld. Maar goed, dat is een manier, maar laat ons wel zien dat heel veel ouders dit volledig aan hun kinderen overlaten. En ja, ik hoor als mensen een mooie vergelijking maken dat je dan eigenlijk kinderen geblinddoekt een soort van snelweg opstuurt. Want ze zijn eigenlijk helaas gewoon los in een jungle waar we helemaal geen zicht op hebben. Klopt, inderdaad. Ik vind ook die ouderbetrokkenheid is het allermoeilijkste."* Deelnemer focusgroep

> *Want wat ik vertelde, die groep zes, waar dus allerlei jongeren elkaar bedreigen en gekheid, tot aan messen mee naar school en nou ja, de meest rare dingen, hebben we gezegd, we gaan met een jeugdagent, een specialistisch jongerenwerker en mezelf de klas in om met die jongeren het gesprek aan te gaan. We hebben wel echt een goede wisselwerking gehad. Groep 6 hè, dus wel heel ja, hoe doe je dat, maar goed, dat ging goed. Toen hebben we via de schooldirectie gevraagd of ouders ook als ze hun kind komen ophalen, een half uurtje willen blijven plakken, zodat we ook die ouders kunnen vertellen: 'hé, dit speelt er, dit zijn de risico's dit zijn de dingen waar je rekening mee moet houden. En we willen graag dit doen.' Dus ook de thuissituatie betrekken bij het probleem. Maar dan heb je dus twee klassen van 25 ouders. Misschien de helft dan, 12, omdat veel alleenstaande moeders in dit gremium zitten. En dan krijg je drie ouders. Dan denk ik, 'ja, waar zijn dan de andere 22?' En dan is het zo moeilijk, dat je denkt, hoe kan dat nou? En toch gebeurt het. En dat heb je ook op middelbare scholen. Dat ik denk. Onraakbaar soms..."* Deelnemer focusgroep

Voorbeelden

In sommige gemeenten wordt al tussen verschillende ketenpartners samengewerkt waarin het thema online geweld al wordt meegenomen. Ook bij de politie zijn hier en daar al initiatieven gaande, bijvoorbeeld waarbij de politie samenwerkt met het onderwijs.

> *Wat wij op ons bureau altijd proberen, is dat wij een korte lijn hebben met basisscholen en middelbare scholen. We komen daar veel op school, we proberen daar op mentorlessen in ieder geval een soort gesprek mee aan te gaan. Ook in de klas. Daarbij heeft de gemeente hier in Almere ook een stukje 'oké op school'. Die geven workshops, daar kunnen scholen zich voor inschrijven. Dat stukje online problematiek wordt daarbij ook benoemd."* Deelnemer focusgroep

Digitale wijkagenten zijn ook al op sommige plekken actief (zie ook vroegsignalering). Ook werken jeugdagenten bijvoorbeeld al met Instagram.

> *We hebben een speciaal jeugdagenten Instagram. Waarin we heel veel DM's, dus direct messages ontvangen van kinderen, jongeren die dus gewoon ook vragen stellen. Ik moet eerlijk zeggen, het is niet altijd zo dat het heel persoonlijk wordt. Maar dat we ze wel doorverwijzen en met name onder andere naar Help Wanted is ook een van de plekken waar we heel veel naar doorverwijzen."* Deelnemer focusgroep

In Rotterdam worden ook sportverenigingen projecten opgezet, wat volgens respondenten ook een mooi voorbeeld is wat van wat zou kunnen werken. Dit soort initiatieven zijn kansrijk.

> *We hebben Rotterdam Protected opgezet met bijvoorbeeld Sparta erbij als een van de hoofddoelen. Feyenoord doet ook een soortgelijk project met allemaal sociale dingen. En waarom ik dit benoem is dat heel veel jongeren vaak iets hebben aan een sportvereniging. Die hebben ook best wel een mate van invloed op die gasten. Misschien soms wel meer dan scholen".* Deelnemer focusgroep

6 Op weg naar een effectieve aanpak van online geweld

6.1. Belangrijkste bevindingen

Hieronder presenteren we kort de belangrijkste bevindingen van dit onderzoek aan de hand van ‘stellingen’ waarmee we antwoord geven op de deelvragen.

Online geweld is diffuus maar heeft herkenbare patronen

Online geweld is lastig af te bakenen, vaak wordt het gebruikt als een overkoepelende term voor allerlei vormen van geweld en het heeft vele facetten, gradaties en verschijningsvormen. Verschillende vormen worden door elkaar heen gebruikt en vaak wordt online geweld gecombineerd met fysiek geweld. Tegelijkertijd is er een aantal herkenbare patronen te onderscheiden:

- Iemands identiteit en (publieke) positie in de samenleving en het maatschappelijke debat – zoals die van mensen uit de LHBTQIA+-gemeenschap, activisten of burgers die deelnemen aan protestbewegingen – beïnvloedt de manier en de mate waarin zij met online geweld te maken krijgen.
- De maatschappelijke context kan een aanzuigende werking hebben, zoals de coronacrisis of een actueel discussiepunt.
- Online geweld wordt vaak doelbewust en strategisch ingezet tegen specifieke groepen, zoals LHBTQIA+-personen, vrouwen en mensen met een migratieachtergrond, en kan zowel voortkomen uit persoonlijke relaties en conflicten als worden versterkt door de anonimiteit van de online omgeving.
- Ook groepsdruk en groepsdynamiek, bijvoorbeeld op school of in de gamesfeer, spelen een belangrijke rol bij het ontstaan en versterken van online geweld.

Online geweld is wijdverspreid

De platforms waarop online geweld zich manifesteert, opereren veelal grensoverschrijdend en de verspreiding van problematische content stopt niet bij landsgrenzen. Iets wat online plaatsvindt en ook als het verwijderd wordt van een platform, kan jarenlang iemand blijven achtervolgen (bijvoorbeeld als het al gedownload en gedeeld is voordat het eraf wordt gehaald).

Online geweld is een continuüm

Online geweld is een continuüm. Dat betekent dat het niet alleen bestaat uit extreme uitingen zoals doodsbedreigingen maar ook subtiele, alledaagse vormen kan aannemen. Deze vormen zijn met elkaar verbonden en kunnen in ernst escaleren. Meerdere vormen van online geweld lopen in elkaar over, er is sprake van een glijdende schaal. In dit onderzoek is een schaal ontwikkeld van online gedrag, variërend van acceptabel tot strafbaar, om het fenomeen online geweld beter te begrijpen. Deze ‘glijdende schaal’ benadrukt dat online onbeschoft gedrag, hoewel niet strafbaar, bijdraagt aan verharding van het publieke debat en uitsluiting van mensen in kwetsbare posities.

Online geweld heeft een hybride karakter

De werelden van online en offline geweld zijn niet meer te onderscheiden waardoor we beter over hybride vormen van geweld kunnen spreken. In verschillende casussen blijken meerdere vormen van online en offline geweld door elkaar heen plaats te vinden.

Online geweld laat zich kenmerken door een escalatieladder en als katalysator van ‘nieuw’ geweld

De escalatieladder is een model dat geweld beschrijft als een proces dat stapsgewijs toeneemt in intensiteit.[14] Er is sprake van een hybride geweldspiraal waarin verschillende vormen en fases elkaar in ‘no tempo’ kunnen opvolgen. Iets wat eerst onaardig is, kan razendsnel uitmonden in een gewelddadige geweldssituatie waarin niet alleen individuen maar ook hele groepen en de samenleving worden geraakt. Of iets wat eigenlijk begint als ‘onschuldig’, kan door de grootschaligheid van gebruikers van sociale media platforms snel omslaan in geweld. Tegelijkertijd kan iets wat online begint, uitmonden in fysiek geweld of omgekeerd. Waar bijvoorbeeld een ruzie tussen scholieren online begint (in een groepsapp bijvoorbeeld), gaat dit vervolgens offline weer verder en andersom. Online en offline volgen elkaar over en weer op en beïnvloeden elkaar wederzijds: wat er in de online wereld afspeelt, heeft invloed op het maatschappelijke debat en fysieke vormen van geweld en andersom.

Online geweld is dubbel

De dubbelheid – persoonlijk én anoniem, offline én online – maakt online geweld moeilijker te bestrijden. Bovendien worden de verschillende vormen van online geweld niet alleen ingezet als uitlaatklep, maar ook strategisch gebruikt om iemand persoonlijk te raken, individuen of groepen uit te sluiten en/of uit de samenleving en het publieke debat te weren.

14 Zie ook de escalatieladder in Odekerken (2017). Het incasseren van ongenoegen: gerechtsdeurwaarders en schuldenaren. Boom Juridisch.

Online geweld is moeilijk in beeld te brengen: er ontbreekt een eenduidig beeld

Het beeld (de cijfers) is onvolledig en niet consistent. Aangiftes ontbreken en is er geen gedragen/gedeelde monitor. Registraties zijn onduidelijk omdat het vaak ook nog samenvalt met traditionele criminaliteit, en online geweld onder die noemer geregistreerd wordt. Momenteel ontbreekt het aan een gedeeld, gedragen en voldoende uitgewerkt beeld om online geweld als complex en multidimensioneel fenomeen goed te begrijpen en effectief aan te pakken. Het is veelal een overkoepelende term die gebruikt wordt voor verschillende delicten (zie ook eerder). Hierdoor is er een incompleet en versnipperd beeld van het totaal plaatje van online geweld. Er is behoefte aan een breed gedragen set definities van online geweld, die niet alleen de strafbare grens duidelijk maken, maar ook de grijze gebieden en minder zichtbare vormen van schadelijk gedrag omvatten.

Online geweld komt veelal laat in beeld

Wanneer incidenten van online geweld bij partijen in de uitvoering in beeld komen (bijvoorbeeld bij politie, gemeenten of scholen) is de situatie veelal geëscaleerd. Juist zicht op het voorstadium waarin de eerste spanningen plaatsvinden, blijft nog achterwege. Mede ook omdat slachtoffers veelal niet erover praten of het bagatelliseren.

Online geweld is een groeiend maatschappelijk probleem dat urgentie vereist

Online geweld is een groeiend maatschappelijk probleem dat urgentie vereist. Experts, professionals maar ook burgers onderkennen dat we in een tijdperk terecht zijn gekomen waarin online geweld ver is doorgedrongen in onze leefwereld en de schaduwzijde ervan niet meer is weg te denken.

Online geweld heeft enorme impact op (het welzijn van) individuen maar ook op gehele groepen (met specifieke kenmerken). Het werkt ontwrichtend en zet aan tot polarisatie

Online geweld raakt individuen maar tegelijkertijd worden hele groepen weggezet. De impact is zowel op het individu maar ook op de samenleving als geheel voelbaar: het werkt ontwrichtend, ondermijnend en zet aan tot polarisatie. De brede maatschappelijke impact van online geweld, loopt van psychisch welzijn tot maatschappelijke polarisatie. Online geweld is allang geen abstract probleem meer. De impact van de uiteenlopende vormen is reëel, diepgaand en vaak meer ontwrichtend dan offline geweld of criminaliteit. Voor slachtoffers is de digitale component extra ingrijpend. Niet alleen vanwege de snelheid waarmee schadelijke content zich verspreidt, maar ook door de blijvende vindbaarheid, de mogelijkheid tot herhaald delen en het gevoel van voortdurende confrontatie via sociale media. Die digitale dimensie maakt online geweld tot een hardnekkig maatschappelijk probleem, met persoonlijke en collectieve gevolgen die niet langer losstaan van de fysieke wereld.

Online geweld heeft niet alleen directe gevolgen voor individuele slachtoffers, maar tast ook de fundamenten van de samenleving aan. Het ondermijnt de sociale cohesie, belemmert het vrije publieke debat en kan leiden tot uitsluiting, polarisatie en uiteindelijk tot verzwakking van de democratische rechtsstaat. Groepen en individuen, vooral zij die zich in een kwetsbare of gemarginaliseerde positie bevinden, worden door haat, intimidatie en bedreigingen ontmoedigd om zich uit te spreken of deel te nemen aan het publieke gesprek.

Online geweld kan iedereen overkomen maar raakt ook juist kwetsbare groepen

Specifieke kenmerken maken sommige individuen en groepen kwetsbaar voor online geweld. Te denken valt aan vrouwen, jongeren (met een licht verstandelijke beperking), LHBTQIA+ personen, mensen met een migratieachtergrond, activisten etc. Daarbij kan online geweld ook iedereen treffen. Bijvoorbeeld situaties waarin online geweld op school voorkomt. De achtergrond van plegers is divers: de leeftijd varieert maar de eerste signalen zijn dat het voornamelijk mannen zijn, soms onderdeel van georganiseerde (sub)culturen en beïnvloed door verschillende maatschappelijke contexten (bijvoorbeeld corona, manosphere, publiek-gevoelige thema's, gamewereld etc.).

Online geweld wordt (nog) niet gezien als 'echt' geweld

Zowel aan de kant van plegers als slachtoffers is te zien dat ze strategieën en copingmechanismen inzetten waarbij online geweldsincidenten worden gerelativeerd, gebagatelliseerd, genegeerd of gerationaliseerd. Slachtoffers proberen het te relativeren, zijn bang voor *victim blaming* of denken dat hun omgeving hen niet begrijpt. Door plegers worden incidenten bijvoorbeeld afgedaan 'als een grapje', 'we doen het allemaal tijdens het gamen' of 'iets wat niet echt is omdat het alleen in de online wereld gebeurt'.

Online geweld kenmerkt zich nog door het gebrek aan adequate ondersteuning, serieuze opvolging van meldingen/aangiften en effectieve opsporing

Er is bij online geweld vaak sprake van schaamte of andere emoties die ervoor zorgen dat er nog niet openlijk over wordt gesproken of dat er niet wordt aangeklopt voor hulp. Slachtoffers willen niet als slachtoffer worden gezien. Ook zien we dat casuïstiek nog ontbreekt, de aangifte- en meldingsbereidheid en dat het vertrouwen in oplossingen en ondersteuning laag is. Toegang en bekendheid met laagdrempelige (anonieme) hulp, een klimaat waarin er met elkaar openlijk over gesproken wordt en een norm waarin men met elkaar weet dat 'online geweld ook geweld is' mist nog.

Online geweld is nog geen gedeeld en gedragen vraagstuk

Aan de kant van de wetenschap, beleid en de uitvoering (waaronder politie en gemeenten) zijn nog verschillende drempels die een effectieve aanpak in de weg staan: zoals een gebrek aan kennis, gebrek aan eigenaarschap, signalen die blijven liggen, gebrek aan capaciteit, onwetendheid over wat wel mag en kan en het gebrek aan een domeinoverstijgende en structurele (inter) nationale samenwerking. De benodigde domeinoverstijgende samenwerking wordt bemoeilijkt doordat private organisaties – met name sociale mediabedrijven zoals Meta, X en TikTok – een cruciale rol spelen, maar vaak niet aan tafel zitten bij de aanpak.

Online geweld mist een visie en stevige expertise

Door wetenschappers, professionals, experts, ministeries, gemeenten en allerlei andere ketenpartners die een rol kunnen spelen in de aanpak van online geweld, wordt erkend dat de specifieke expertise en kennis, in bijzonder met betrekking tot online geweld tussen burgers onderling, er nog niet is. Online geweld wordt soms als een technisch fenomeen gezien, maar de impact is veel breder (zie verder in de aanbevelingen).

Online geweld is nog "nieuw"

Het complexe vraagstuk van online geweld is nog nieuw en men is zoekende naar de juiste inzichten, antwoorden én oplossingen. Er is gebrek aan onderzoek naar de wereld achter online geweld, welke interventies kansrijk én effectief zijn (wat zijn de werkzame elementen, geleerde lessen en randvoorwaarden) en (nieuwe of aangepaste) wet- en regelgeving.

Online geweld is toe aan een inhaalslag

Als we kijken naar hoe complex het vraagstuk van online geweld is, naar de impact ervan, het gebrek aan specifieke wetgeving en het ontbreken van het nemen van verantwoordelijkheid van allerlei verschillende actoren waaronder burgers zelf, professionals en experts, dan liggen er nog genoeg kansen voor verbetering om dit met elkaar te doen. We staan aan het begin van een noodzakelijke stevige maatschappelijke en juridische inhaalslag.

6.2. Aanbevelingen

Doorpakkend op de belangrijkste bevindingen omschrijven we onze overkoepelende aanbevelingen. Deze zijn niet op prioriteit gerangschikt. Om tot een effectieve aanpak van online geweld te komen is het van belang ze in samenhang in te zetten. Onder de verschillende aanbevelingen benoemen we concrete acties en stappen:

1. Duurzame domeinoverstijgende en publiek-private (inter)nationale samenwerking met een integrale en holistische benadering.
2. Centrale rol vanuit het Ministerie van Justitie & Veiligheid in samenhang met andere afdelingen en Ministeries.
3. Verantwoordelijkheid van techbedrijven.
4. Investeren in kennis, onderzoek, monitoring en (vroegtijdige) interventies over online geweld als complex en multi-dimensioneel vraagstuk.
5. Preventie en bewustwording op individueel en maatschappelijk niveau.

6.2.1. Duurzame domeinoverstijgende en publiek-private (inter)nationale samenwerking met een integrale en holistische benadering

Mede omwille van de impact van online geweld en het ontwrichtende en polariserende karakter, is het noodzakelijk dat online geweld dezelfde prioriteit en ernst krijgt als andere vormen van geweld en criminaliteit — een zogenoemde *parity of esteem*. Slachtoffers moeten kunnen rekenen op adequate ondersteuning, serieuze opvolging van meldingen/aangiften en effectieve opsporing die specifiek gericht is op online geweld tussen burgers onderling. Maar een geloofwaardige, effectieve aanpak van online geweld kan niet beperkt blijven tot één enkele sector of verantwoordelijke partij. Online geweld raakt meerdere maatschappelijke domeinen tegelijk: bijvoorbeeld veiligheid/criminaliteit, technologie, onderwijs, gezondheid, jeugd, participatie. Dat vraagt om een domeinoverstijgende aanpak waarin publieke én private partijen samen optrekken. Een duurzame aanpak van online geweld vraagt bovendien om structurele samenwerking en gedeelde verantwoordelijkheid tussen overheden, maatschappelijke organisaties, onderwijsinstellingen, media, techbedrijven en burgers.

Gezien het grensoverschrijdende karakter van online communicatie zijn internationale samenwerkingsverbanden onmisbaar om standaarden en effectieve methoden op Europees of mondiaal niveau te bevorderen. Door deze gezamenlijke inspanning en gedeelde verantwoordelijkheid ontstaat een samenhangend en effectief preventie- en interventiebeleid, dat recht doet aan de

complexiteit van online geweld in diverse digitale omgevingen. Alleen door afspraken te maken op Europees en mondiaal niveau, bijvoorbeeld via de Digital Services Act, door samenwerking met internationale techbedrijven en door aansluiting bij internationale kennisnetwerken, kan een effectieve en toekomstbestendige aanpak vorm krijgen.

Onderliggend aan al deze stappen is de overtuiging nodig dat online geweld een gedeeld maatschappelijk vraagstuk is. Geen enkele partij, sector of overheidsonderdeel kan het alleen oplossen. Publiek-private samenwerking, domeinoverstijgend beleid en internationale afstemming zijn essentieel. Alleen dan kunnen we de structurele uitsluiting, ontwrichting en schade die online geweld veroorzaakt, daadwerkelijk tegengaan en bouwen aan een samenleving waarin zowel de fysieke als de digitale publieke ruimte veilig, inclusief en toegankelijk is voor iedereen. De brede maatschappelijke impact van online geweld, van psychisch welzijn tot maatschappelijke polarisatie, vraagt om een integrale en holistische benadering.

6.2.2. Centrale rol vanuit het Ministerie van Justitie en Veiligheid in samenhang met andere afdelingen en Ministeries

Het is van belang dat er een centrale rol komt voor een Ministerie die in samenhang met andere afdelingen en Ministeries de regierol heeft op het thema online geweld tussen burgers onderling. Het Ministerie van Justitie en Veiligheid kan een actieve en coördinerende rol pakken in de (inter) nationale aanpak van online geweld. Daarbij is het noodzakelijk dat niet alleen de bekende veiligheids- en justitieketens betrokken worden, maar ook andere afdelingen binnen het Ministerie van Justitie en Veiligheid, zoals Slachtofferbeleid en Ondermijning en andere relevante Ministeries zoals Sociale Zaken en Werkgelegenheid (SZW) Volksgezondheid, Welzijn en Sport (VWS), Onderwijs, Cultuur en Wetenschap (OCW) en Binnenlandse Zaken en Koninkrijksrelaties (BZK). Hierbij is het van belang dat er niet alleen focus ligt op de 'achterkant' – de strafrechtelijke en repressieve kant, maar juist ook de 'voorkant' – preventie, vroegsignalering en organisaties die dichtbij burgers staan zoals jeugdhulp, onderwijs etc. (zie ook aanbeveling 6.2.5).

Projectgroep of Taskforce

Een concrete stap hierin is het oprichten van een Nationale Projectgroep naar het voorbeeld van bestaande samenwerkingsverbanden zoals PersVeilig, dat zich inzet voor de veiligheid van journalisten of een Taskforce voor online geweld zoals de Taskforce Onze Hulpverleners Veilig. Zo'n interdepartementale Projectgroep of Taskforce kan zorgen voor coördinatie, kennisdeling en gezamenlijke actie. Belangrijk is dat hier niet alleen het Ministerie van Justitie en Veiligheid bij betrokken is, maar ook de andere ministeries (zoals hierboven benoemd). Zoek hierbij bewust de verbinding op met andere ministeries en afdelingen die ook bezig zijn met een aanpak, bijvoorbeeld de aanpak Online discriminatie van het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (Online discriminatie: een gezaamenlijke aanpak) of waar raakvlakken liggen om samen in de aanpak Online geweld op te trekken.

Koppelen aan bestaande programma's en initiatieven

De publiek-private samenwerking waarin het Ministerie van Justitie en Veiligheid en coördinerende rol kan pakken, moet niet op zichzelf staan, maar kan verbonden worden met bestaande lokale en landelijke programma's en initiatieven. Te denken valt aan lopende projecten, programma's en initiatieven die nu veelal gericht zijn op bijvoorbeeld cybercriminaliteit, online seksuele intimidatie, geweld tegen vrouwen etc. waarin de focus van dit onderzoek – *online geweld tussen burgers onderling* - aan toegevoegd kan worden.

Onlangs werd bekend dat er een nieuw onderzoeksprogramma is gelanceerd. Lancering grootschalig onderzoeksprogramma Weerbaar tegen online criminaliteit. Het zou mooi zijn om vanuit de inzichten die zijn opgedaan in dit onderzoek, een volgende stap te zetten middels dit onderzoeksprogramma. Zie ook verder aanbeveling 4.

Meldpunt en vertrouwensorganisatie specifiek gericht op online geweld tussen burgers onderling

Daarnaast is het wenselijk om een laagdrempelig, breed toegankelijk meldpunt in te richten waar slachtoffers, maar óók plegers van online geweld terechtkunnen voor adequate ondersteuning, advies en waar nodig hulp bij het doen van aangifte, het nemen van juridische stappen of begeleiding. Dit meldpunt dient nauw verbonden te zijn met bestaande organisaties, zoals Slachtofferhulp, politie en gespecialiseerde hulpverleners, zodat slachtoffers niet van het kastje naar de muur worden gestuurd. Waar mogelijk kan er bijvoorbeeld aangesloten worden bij bestaande meldpunten/initiatieven zoals Help Wanted, SafetyNed en Fier. Er is nu wel een meldpunt Online discriminatie, waardoor de meldpunten zijn versnipperd en niet specifiek zijn gericht op de vorm van online geweld waar we in ons onderzoek op focussen.

Het zou tevens mooi zijn om een vertrouwensorganisatie ook specifiek gericht op Online geweld tussen burgers onderling in het leven te roepen. Initiatieven zoals Jouw Ingebrachte Mentor van Halt, waarbij jongeren uit hun vertrouwde omgeving betrokken worden bij het bespreekbaar maken van grensoverschrijdend gedrag, bieden aanknopingspunten om ook in de aanpak van online geweld vertrouwde netwerken in te zetten.

6.2.3. Verantwoordelijkheid van techbedrijven

Ook samenwerking met techbedrijven is onmisbaar. Zij beheren de platforms waar het grootste deel van het online geweld zich afspeelt en beschikken over cruciale data die inzicht kunnen geven in aard, omvang en verspreiding van het probleem. Tegelijkertijd hebben zij een directe verantwoordelijkheid om hun digitale omgeving zo veilig mogelijk te maken. Het is dan ook noodzakelijk dat er heldere, bindende afspraken komen over hun rol in preventie, moderatie en handhaving. Techbedrijven moeten niet alleen transparant zijn over hun beleid en procedures voor het verwijderen van schadelijke content, maar ook actief bijdragen aan de handhaving van wettelijke normen. Dat betekent dat zij meldingen van strafbaar online geweld tijdig en adequaat opvolgen, dat zij meewerken aan opsporing wanneer dat nodig is, en dat zij technische mogelijkheden inzetten om recidive en herhaald daderschap te voorkomen.

Daarnaast moeten zij, in samenwerking met publieke partijen, structureel geanonimiseerde data beschikbaar stellen voor onafhankelijk onderzoek. Alleen door die combinatie van preventie, transparantie, handhaving en kennisdeling kunnen we het online geweld daadwerkelijk terugdringen en een veiliger digitaal klimaat creëren.[15]

Ontwikkel platformoverstijgende afspraken en verantwoordelijkheden

Online geweld stopt niet bij de grenzen van één platform; gebruikers, boodschappen en dynamieken verplaatsen zich vaak tussen verschillende sociale media en online omgevingen. Dit vraagt om een gecoördineerde en platformoverstijgende aanpak waarbij verschillende platforms, beleidsmakers en maatschappelijke organisaties nauw samenwerken. Het is van groot belang om afspraken te maken over monitoring, data-uitwisseling en gezamenlijke normstelling, zodat schadelijk gedrag en de verspreiding van online geweld effectief over platformgrenzen heen kunnen worden aangepakt. In deze samenwerking spelen social mediaplatforms zoals TikTok, Meta, X en kleinere alternatieve platforms een belangrijke rol. Zij zijn verantwoordelijk voor het ontwikkelen en uitvoeren van uniforme contentregels, het delen van relevante data over schadelijke content, en het verbeteren van de transparantie van algoritmische beslissingen.

Hierbij is het essentieel dat onderzoeksinstituten en kenniscentra monitoren en evalueren welke interventies het meest succesvol zijn en leveren daarmee evidence-based adviezen (zie aanbeveling hierna).

Betrek techbedrijven actief bij overleggen

In de praktijk blijft het lastig om de techbedrijven ook aan tafel te krijgen. Zoek als overheid bewust naar manieren om techbedrijven mee te krijgen in de gezamenlijke ambitie om online geweld tussen burgers onderling aan te pakken. Niet alleen door repressief en af te dwingen, maar ook in gezamenlijkheid te kijken wat techbedrijven kunnen doen en hoe je het beste elkaars krachten kan versterken.

15 Recent zijn er rechtszaken over de werkwijze geweest. Volgens Bits of Freedom is de manier waarop de tijdlijnen zijn ingericht, in strijd met de Digital Services Act (DSA). Meta krijgt nu tot eind december tijd om de aanpassingen voor een algoritme-vrije tijdlijn door te voeren, maakt de rechtbank bekend (Meta krijgt meer tijd voor aanpassingen Facebook en Instagram; Meta krijgt meer tijd voor algoritme-vrije tijdlijn op Facebook en Instagram; ECLI:NL:GHAMS:2025:2886, Gerechtshof Amsterdam, 200.360.457/01).

6.2.4. Investeren in kennis, onderzoek, monitoring en interventies over online geweld als complex en multi-dimensioneel probleem

Het aanpakken van online geweld vereist structurele investering in onderzoek, monitoring en kennisontwikkeling. Dit vraagt om een combinatie van kwantitatieve en kwalitatieve methoden zoals:

- **Systematische monitoring:** langdurig en breed, niet beperkt tot enkele zoektermen of platforms, maar ook gericht op diverse digitale omgevingen zoals Facebook, Instagram, TikTok en Reddit.
- **Kwantitatieve data:** frequentie, verspreiding en patronen van online geweld.
- **Kwalitatieve analyses:** context, intenties en impact op slachtoffers en gemeenschappen.
- **Sociale media-analyses:** inzicht in platformverschillen, cross-platform gedrag, netwerkanalyses en de relatie met maatschappelijke en politieke gebeurtenissen.

Deze aanpak maakt het mogelijk trends vroegtijdig te signaleren en de effectiviteit van interventies te evalueren. Monitoring moet worden geïntegreerd in beleidscycli en uitgevoerd in samenwerking tussen onderzoeksinstellingen, beleidsmakers, maatschappelijke organisaties en platforms, met aandacht voor privacy en ethiek. Het is van belang dat er een goede koppeling wordt gelegd tussen de monitoring van online ontwikkelingen en de lokale praktijk. Informatie uit sociale media-analyses en signaleringssystemen moet terugvloeien naar professionals en organisaties die dicht bij de leefwereld van betrokken burgers staan, zodat zij vroegtijdig kunnen ingrijpen en escalaties kunnen voorkomen.

Daarnaast is **lerende evaluatie** essentieel: onderzoek moet niet alleen registreren, maar ook inzicht geven in de werking van interventies en beleidsmaatregelen. Dit vraagt om leer-actiecycli waarin reflectie, onderzoek en bijsturing samenkomen. Daarbij worden niet alleen experts, maar ook burgers, maatschappelijke organisaties en andere stakeholders betrokken, bijvoorbeeld via burgerpanels of dialoogtafels. Gezien de grensoverschrijdende aard van online geweld is internationale samenwerking in deze leerprocessen cruciaal. Door te investeren in brede kennisontwikkeling, systematische monitoring en lerende evaluatie ontstaat een stevig fundament voor evidence-based interventies en beleid.

Het is van groot belang om het **lerend vermogen en de expertise en deskundigheid** van zowel onderzoekers als professionals die met online geweld te maken hebben, substantieel **te versterken**. Online geweld is geen eendimensionaal probleem; het vraagt om diepgaande en specifieke kennis van de dynamiek van digitale communicatie, geweld en criminaliteit, de werking van sociale platforms, juridische kaders, de psychosociale impact op zowel slachtoffers als plegers, en de bredere maatschappelijke gevolgen, zoals polarisatie, verharding van het maatschappelijk debat en structurele uitsluiting van bepaalde stemmen. Momenteel ontbreekt een gedragen en voldoende uitgewerkt beeld om online geweld als complex en multidimensioneel fenomeen goed te begrijpen en effectief aan te pakken. Het is daarom essentieel dat een gemeenschappelijk begrippenkader en een eenduidige, gedragen definitie multidisciplinair worden ontwikkeld, met actieve betrokkenheid van beleidsmakers, wetenschappers, platforms én diverse maatschappelijke groepen. Op basis van deze gedeelde definitie kan vervolgens een consistente en breed gedragen indicatorenset worden opgesteld, die het mogelijk maakt signalen van online geweld beter te herkennen, systematische monitoring op te zetten en effectieve interventies te ontwerpen die recht doen aan de complexiteit van online interacties. Door deze gezamenlijke, multidisciplinaire basis worden verschillende perspectieven en ervaringen meegenomen, wat bijdraagt aan een genuanceerd, inclusief en toepasbaar kader voor beleid, onderzoek en praktijk.

Ontwikkel een gedeelde definitie van online geweld (categorisatie/ typologie/indicatorenset)

Een belangrijke stap in de aanpak van online geweld is het ontwikkelen van een duidelijke, breed gedragen en multidisciplinaire definitie, inclusief een bijbehorende set indicatoren en een praktische typologie of categorisering van verschillende vormen en gradaties van online geweld. In de praktijk blijkt het begrippenkader vaak diffuus, wat effectieve samenwerking, handhaving en beleidsvorming belemmert. Een gezamenlijke basis kan bijvoorbeeld tot stand komen via een Delphi-studie, waarin wetenschappers, professionals, maatschappelijke organisaties, techbedrijven en ervaringsdeskundigen gezamenlijk toewerken naar gedeelde definities, indicatoren en meetinstrumenten (Terwee et al., 2018; Nasa et al., 2021). Op basis van deze gemeenschappelijke definitie kan een breed gedragen indicatorenset worden opgesteld, waarmee signalen van online geweld beter kunnen worden herkend, systematische monitoring mogelijk wordt en effectieve interventies ontwikkeld kunnen worden. Deze definitie kan als eerste aanzet worden gezien voor een breed gedragen indicatorenset en typologie, die multidisciplinair wordt ontwikkeld met betrokkenheid van beleidsmakers, wetenschappers, platforms en maatschappelijke groepen. Zo ontstaat een gezamenlijk kader voor monitoring, analyse en interventie, dat rekening houdt met de complexiteit, samenhang en maatschappelijke gevolgen van online geweld. Zo ontstaat een gezamenlijk kader voor monitoring, analyse en interventie, dat rekening houdt met de complexiteit, samenhang en maatschappelijke gevolgen van online geweld.

Eerste aanzet tot definitie online geweld

Online geweld tussen burgers onderling verwijst naar schadelijk gedrag dat via digitale middelen wordt gepleegd, met als doel of gevolg het veroorzaken van psychische, sociale, emotionele of reputatieschade bij individuen of groepen, inclusief ‘publieke figuren’. Het kan plaatsvinden via sociale media, berichtenapps, e-mail, fora of andere online platforms en kan samengaan met offline en/of fysiek geweld. Het kan verschillende vormen combineren, wordt vaak door elkaar heen ingezet en kan escaleren in intensiteit en impact. Het doel of gevolg van online geweld kan naast individuele schade ook leiden tot uitsluiting van bepaalde personen of groepen, polarisatie, afname van vertrouwen in digitale platforms en bredere maatschappelijke ontwrichting.

Componenten van definitie:

1. **Handeling of gedrag:** Geweld verwijst meestal naar een actie of reeks acties die schade veroorzaken of beogen te veroorzaken.
2. **Schade of dreiging van schade:** Dit kan fysiek, psychisch, emotioneel of materieel zijn. Het gaat niet alleen om daadwerkelijke schade, maar ook om de dreiging ervan.
3. **Intentie:** vaak gericht op het veroorzaken van schade. In veel definities is de intentie om schade toe te brengen een belangrijk element.
4. **Slachtoffer:** een persoon, groep of ‘publieke figuur’. Er is altijd een persoon, groep of object dat het doelwit is van het geweld.
5. **Macht of controle:** geweld kan worden gebruikt om macht uit te oefenen, controle te verkrijgen of dominantie te vestigen.
6. **Context:** sociale, culturele en juridische context beïnvloedt de mate en vorm van geweld. Ook beïnvloedt het de interpretatie van het geweld. De sociale, culturele en juridische context bepaalt vaak of iets als geweld wordt gezien. Bijvoorbeeld: wat in de ene cultuur als geweld wordt beschouwd, kan in een andere als acceptabel worden gezien.
7. **Hybride karakter:** online geweld heeft een hybride karakter.
8. **Type geweld:** Er zijn veel verschillende vormen van online geweld, zoals de vormen die in dit onderzoek zijn meegenomen: Bedreiging/intimidatie, Haatzaaien en Oproepen tot kleinschalige ordeverstoringen. De verschillende vormen van online geweld komen vaak in combinatie voor en worden door elkaar heen gebruikt. Daarbij is het belangrijk te benadrukken dat online geweld ook kan samengaan met offline geweld, zoals fysieke intimidatie, bedreiging of andere strafbare handelingen.
9. **Impact op individu en samenleving:** De anonimiteit en snelheid van verspreiding vergroten de impact. Slachtoffers ervaren vaak psychische klachten, angst, schaamte en sociale isolatie. De interactie tussen verschillende vormen van online geweld en de combinatie met offline incidenten kan leiden tot een versterkte en cumulatieve impact, zowel op individuen als op de samenleving. Het kan leiden tot polarisatie, uitsluiting van personen of groepen, afname van vertrouwen in digitale platforms, verhoogd gevoel van onveiligheid, reputatieschade en bredere maatschappelijke ontwrichting.

Het zou mooi zijn om deze eerste aanzet van een definitie in samenwerking met de ministeries en alle stakeholders verder te ontwikkelen in speciale creatieve werksessies. Een toegevoegde waarde is om de definitie niet alleen schriftelijk, maar ook visueel te maken zodat het op een laagdrempelige manier in de praktijk kan worden gedeeld.

Het **systematisch en langdurig monitoren van online geweld** is essentieel om inzicht te krijgen in de aard, omvang en veranderende dynamieken van dit fenomeen. Dit betekent dat onderzoek niet beperkt mag blijven tot een beperkt aantal zoektermen, onderwerpen of platformen, maar juist uitgebreid moet worden naar diverse sociale mediaplatformen zoals Facebook, Instagram, TikTok, Reddit en andere relevante digitale omgevingen. Daarnaast is het cruciaal om zowel kwantitatieve data te verzamelen over frequentie en verspreiding, als kwalitatieve analyses uit te voeren die context, intenties en de impact op slachtoffers en gemeenschappen verklaren. Deze combinatie maakt het mogelijk om trends vroegtijdig te signaleren en de effectiviteit van interventies te evalueren. Een dergelijke structurele monitoring vereist samenwerking tussen onderzoeksinstellingen, sociale mediaplatformen, beleidsmakers en maatschappelijke organisaties, zodat data-uitwisseling, privacybescherming en ethische standaarden goed geborgd zijn. Ook is het belangrijk om deze monitoring te integreren in bestaande beleidscycli, zodat het direct bijdraagt aan het ontwikkelen en bijstellen van preventieve en beschermende maatregelen. Op die manier kunnen **gerichte, evidence-based beleidsinterventies** worden gerealiseerd die aansluiten bij actuele en toekomstige vormen van online geweld.

Monitoring, registratie en kwantitatieve analyses

Daarnaast is **structurele en systematische monitoring** noodzakelijk, waarin zowel kwantitatieve als kwalitatieve methoden gecombineerd worden. Het **opnemen van online geweld als aparte categorie in bestaande registratiesystemen, of het structureel vastleggen wanneer bij offline delicten ook sprake is van een online component**, zijn belangrijke voorwaarden om de verwevenheid van online en offline geweld beter zichtbaar te maken. Ook **grondige kwantitatieve analyses**, zoals dossieronderzoek, het scannen van politieregistraties en het systematisch doorlopen van meldingen en aangiftes, zijn hierin onmisbaar.

Naast kwantitatieve analyses en monitoring is **diepgaand kwalitatief** onderzoek nodig, bijvoorbeeld in de vorm van (reflexieve) gesprekken met betrokkenen uit verschillende overheidsinstanties, maatschappelijke organisaties, burgers en slachtoffers/plegers. Dit biedt inzicht in uiteenlopende ervaringen en belevingen rondom online geweld, variërend van openlijke haat en intimidatie tot subtielere vormen van uitsluiting, framing of manipulatie. Ook sentimentanalyses in en onderzoek naar de toonzetting van online berichten kunnen hieraan bijdragen. Tevens kan door middel van uitgebreid casusonderzoek naar incidenten dieper zicht gekregen worden op de wereld achter online geweld. Die casuïstiek kan ook weer gebruikt worden voor het voeden van de typologie/categorisatie.

Gebruik het concept van de glijdende schaal als preventie-instrument en verdiep het begrip verder

Online geweld ontwikkelt zich vaak geleidelijk, beginnend met subtiele vormen zoals cynische opmerkingen en uitsluiting, en kan escaleren naar openlijke haat, bedreiging en strafbare feiten. Het concept van een 'glijdende schaal' helpt om deze opeenvolgende fasen beter te begrijpen en te signaleren. Het is van groot belang om deze schaal verder te verkennen en te concretiseren, zodat er duidelijkheid ontstaat over de grijze gebieden en de momenten waarop preventieve interventies het meest effectief kunnen zijn. Op basis hiervan kunnen systematische signaleringssystemen en gerichte interventies worden ontwikkeld die vroegtijdig ingrijpen voordat het gedrag escaleert. Dit draagt bij aan het voorkomen van schadelijke escalaties en bevordert een veiliger online klimaat.

Eveneens is het van meerwaarde om te investeren in (vroegtijdige) **interventie-/ implementatietrajecten en/of lerende evaluaties.** De complexiteit en gelaagdheid van het fenomeen online geweld vraagt om onderzoek dat niet alleen registreert wat er gebeurt, maar ook inzicht geeft in welke interventies, beleidsmaatregelen en publiek-private samenwerkingen daadwerkelijk effectief zijn, onder welke omstandigheden, voor wie, en waarom. Daarbij is het essentieel om deze lerende benadering niet te beperken tot experts en professionals, maar ook burgers en andere direct betrokkenen actief te betrekken. De meest recente inzichten over lerend evalueren benadrukken het belang van zogenoemde leer-actiecycli, waarin onderzoek, reflectie en concrete aanpassingen steeds in samenhang en herhaling plaatsvinden. Door burgers, maatschappelijke organisaties en andere stakeholders structureel onderdeel te maken van deze cycli — bijvoorbeeld via dialoogtafels, burgerpanels of participatieve monitoring — ontstaat een voortdurend proces van gezamenlijke kennisontwikkeling, handelingsverbetering en aanpassing van beleid of interventies. Dit draagt niet alleen bij aan betere inzichten in de effectiviteit van de aanpak van online geweld, maar versterkt ook de handelingsbekwaamheid, weerbaarheid en kennis van alle betrokken partijen, inclusief burgers zelf. Lerende evaluatie wordt daarmee geen eenmalige exercitie, maar een doorlopend, cyclisch proces van gezamenlijk leren, experimenteren en bijsturen. Het is daarnaast belangrijk om deze leer-actiecycli niet alleen nationaal, maar ook internationaal vorm te geven. Gezien de grensoverschrijdende aard van online geweld is samenwerking met andere landen, internationale kennisnetwerken en techbedrijven onmisbaar om ook op internationaal niveau te komen tot effectieve en lerende aanpakken (Beers & Van Mierlo 2017; Mierlo et al 2010).

Focus in onderzoek, monitoring én aanpak/interventies op online geweld tussen burgers onderling

Belangrijk is om oog te hebben voor online geweld tussen burgers onderling en de vormen waar dit onderzoek zich op richt. Bestaand onderzoek is veelal gericht op andere vormen waaronder fraude, cybercriminaliteit, seksueel geweld en grootschalige ordeverstoringen. **De vormen van online geweld die op alledaagse basis voorkomen in de leefwereld van al onze burgers, verdienen meer én snel aandacht vanwege de diepe ingrijpbaarheid op individuele levens, groepen maar ook de samenleving als geheel.**

Tegelijkertijd is het noodzakelijk om **sociale media-analyses** nadrukkelijker en structureler onderdeel te maken van onderzoek naar online geweld. Sociale mediaplatformen vormen niet alleen de plek waar veel online geweld zich manifesteert, maar bieden ook cruciale inzichten in de schaal, aard en verspreiding ervan. In het huidige onderzoek is gekeken naar Nu.nl en X (voorheen Twitter), waarbij duidelijke verschillen zichtbaar werden tussen de twee platformen. Nu.nl en X trekken vermoedelijk verschillende gebruikersgroepen aan, met uiteenlopende demografische kenmerken, interesses en communicatiestijlen. Dit vertaalt zich in verschillen in taalgebruik en in de frequentie en aard van problematische uitingen. Op Nu.nl zagen we dat het percentage problematische taal in goedgekeurde berichten lager ligt dan op X. Dit kan wijzen op effectievere moderatie, maar ook op verschillen in platformcultuur. Doordat we op X geen informatie hebben over afgekeurde berichten, kunnen moderatie-inspanningen tussen de platformen echter niet goed vergeleken worden. Bovendien ontbreekt het nog aan inzicht in het cross-platform gedrag van gebruikers. We weten niet of en hoe mensen actief zijn op meerdere platformen en of zij daar vergelijkbaar of juist afwijkend gedrag vertonen. Het combineren van data van verschillende platformen en het volgen van gebruikers over die platformen heen, biedt waardevolle inzichten in patronen van online geweld. Zo kan onderzocht worden of dezelfde personen zich op meerdere plekken negatief uitlaten, of dat gebruikers zich uitsluitend mengen in discussies over een specifiek thema dat hen raakt of interesseert. Dergelijk onderzoek helpt ook om zicht te krijgen op de schaal van het probleem: is online geweld vooral het resultaat van een beperkte groep zeer actieve gebruikers, of gaat het om een breder gedragen fenomeen binnen grotere online gemeenschappen? Hiervoor zijn onder meer netwerkanalyses en analyses van gebruikersgedrag nodig, die laten zien hoe problematische uitingen zich verspreiden, hoe daders met elkaar interacteren en of er sprake is van georganiseerde coördinatie. Ook de maatschappelijke en politieke context waarin online uitingen plaatsvinden moet nadrukkelijk betrokken worden bij onderzoek en monitoring. Eerdere studies laten zien dat maatschappelijke spanningen, nieuwsgebeurtenissen of politieke ontwikkelingen directe impact kunnen hebben op het volume en de aard van online geweld. Het is daarom van belang om systematisch te analyseren hoe pieken in problematisch taalgebruik samenhangen met actuele gebeurtenissen. In dit verband is het ook wenselijk om de thematische reikwijdte van sociale media-analyses uit te breiden. Zo kan het in vervolgonderzoek waardevol zijn om gerelateerde termen als ‘Moslims’, ‘Islam’ of ‘Islamitisch’ expliciet mee te nemen in de analyses, om inzicht te krijgen in hoe online geweld zich verhoudt tot bredere maatschappelijke discussies, beeldvorming en polarisatie.

Besteed expliciete aandacht aan taal, framing en machtsrelaties

Woorden en termen zoals ‘wappie’, ‘fascist’ of ‘tuig/moslim terroristen’ dragen maatschappelijke betekenissen die sterk kunnen verschillen per context en doelgroep. Het is daarom belangrijk om systematisch te onderzoeken hoe taal online wordt ingezet om te polariseren, te ontmenselijken of juist te mobiliseren. Dit begrip helpt beleidsmakers, onderzoekers en moderatoren om niet alleen beter in te schatten wanneer taal overgaat in vormen van online geweld, maar ook om gerichte preventieve maatregelen te ontwikkelen. Preventie kan bijvoorbeeld bestaan uit educatieve programma’s, mediatrainingen, en het stimuleren van bewustwording binnen gemeenschappen. Zo kan zowel beleid, handhaving als preventie effectief bijdragen aan het verminderen van schadelijke taal en de gevolgen daarvan.

Betrek visuele content in signalering en analyse

Online communicatie beperkt zich niet tot tekst; visuele content speelt een grote rol bij het overbrengen van boodschappen, ook bij online geweld en haatuitingen. Memes, emoji’s, video’s en afbeeldingen kunnen subtiele, maar krachtige vormen van intimidatie of uitsluiting zijn. Effectieve signalering, onderzoek en beleidsmaatregelen moeten daarom ook deze visuele uitingen systematisch meenemen. Daarnaast is het belangrijk om de rol van AI en nieuwe technologieën in het creëren en verspreiden van visuele content te betrekken.

Investeren in brede, diepgaande kennisontwikkeling, systematische monitoring, sociale media-analyses en lerende evaluatie zal leiden tot concrete handreikingen, meetinstrumenten en tools die beter aansluiten bij de complexiteit en dynamiek van online geweld. Hiermee kunnen professionals, beleidsmakers en maatschappelijke partijen effectiever en meer onderbouwd bijdragen aan de aanpak van dit urgente maatschappelijke probleem.

6.2.5. Preventie en bewustwording op individueel en maatschappelijk niveau

Als de basis van alle plannen voor een effectieve aanpak van online geweld, staat de norm dat online geweld óók geweld is. Het verdient dezelfde serieuze benadering als andere vormen van geweld en criminaliteit, zoals verankerd in het rijksbrede beleid rondom agressie en geweld tegen hulpverleners. Een gedeeld besef van de ernst en impact van online geweld is een noodzakelijke randvoorwaarde voor een effectieve, integrale en toekomstbestendige aanpak.

Norm: Online geweld is ook geweld

Ontwikkel en verspreid een duidelijke normstelling. Concreet valt te denken aan de normstelling die ook bij andere vormen van geweld, zoals de sociale norm van geweld tegen hulpverleners en de norm voor werkgevers (Norm Stop Agressie Samen | Publicatie | Veiligepubliekedienstverlening.nl). Het is waardevol om voor online geweld tussen burgers onderling ook te ontwikkelen.

Het is daarbij eveneens van belang om meer in te zetten op preventie en educatie. Recent is er al veel aandacht hiervoor geweest: Kamerbrief over richtlijn gezond en verantwoord scherm- en sociale mediagebruik | Kamerstuk | Rijksoverheid.nl en is een richtlijn 'Gezond schermgebruik opgesteld.[16] Maar ook in andere landen is hiervoor aandacht, waarbij o.a. ook wordt gesproken over een verbod op social media voor kinderen onder de vijftien jaar.[17]

Het is noodzakelijk te investeren in de digitale geletterdheid en sociale weerbaarheid van burgers. Niet alleen jongeren, maar ook volwassenen en specifieke risicogroepen hebben behoefte aan handvatten om zich veilig en verantwoordelijk te bewegen in de online omgeving. Educatieve programma's over online omgangsvormen, het herkennen van schadelijk gedrag en het verantwoord gebruik van sociale media dragen hieraan bij. Digitale burgerschapsvorming verdient bovendien een structurele plek binnen het onderwijs, waarbij niet alleen technische vaardigheden, maar juist ook empathie, respectvolle communicatie, mediawijsheid en kritisch denken centraal staan. Tegelijkertijd is het minstens zo belangrijk te investeren in het versterken van een positieve online cultuur. Dit betekent het stimuleren van campagnes en initiatieven die respectvolle communicatie, empathie en constructieve meningsuitwisseling bevorderen, zowel op sociale platforms als in het dagelijks leven. Door in te zetten op bewustwording, het creëren van een veilige en inclusieve publieke ruimte, samenwerking en positieve online cultuurverandering kan de samenleving als geheel veerkrachtiger worden gemaakt tegen online geweld. Alleen zo wordt voorkomen dat online geweld leidt tot verdere polarisatie, maatschappelijke uitsluiting en verzwakking van de democratische rechtsorde.

Vroegsignalering en verbinden aan bestaande programma's en initiatieven

Gemeenten, scholen, jeugdwerkers, wijkteams en maatschappelijke initiatieven spelen een belangrijke rol bij het vroegtijdig in beeld krijgen van online geweld. Waar ze in sommige situaties 'te laat' zijn, kunnen zij juist een belangrijke beweging naar de voorkant maken. Zij bevinden zich dicht bij de leefwereld van jongeren en andere betrokkenen, en kunnen **vroegtijdig signalen oppikken en gericht ingrijpen**. Het zou zinvol zijn om een **handelingskader, signalerenkaart of gesprekwijzer** te ontwikkelen voor maatschappelijke organisaties die nauw in contact staan met burgers. Zo is er bijvoorbeeld ook een handelingskader voor doxing voor werkgevers ontwikkeld (Handelingskader Doxing).

16 Richtlijn gezond schermgebruik 2025 | Rapport | Rijksoverheid.nl; file.

17 Denemarken wil verbod op sociale media voor kinderen onder vijftien jaar | Tech | NU.nl

Stimuleer kennisdeling, dialoog en praktijkontwikkeling tussen professionals en lokale netwerken

Professionals zoals jongerenwerkers, onderwijsinstellingen, wijkagenten, maatschappelijk werkers en andere betrokkenen krijgen steeds vaker te maken met de impact van online conflicten die zich vertalen naar spanningen in de offline leefomgeving. Het is daarom cruciaal dat deze actoren goed toegerust zijn om signalen van online geweld te herkennen, te begrijpen en effectief aan te pakken. Dit vraagt om het opzetten van bredere kennisnetwerken en samenwerkingsverbanden waarin ervaringen, inzichten en best practices systematisch worden uitgewisseld en gedeeld. Gerichte trainingen, praktische tools en methodieken kunnen professionals ondersteunen om de complexe dynamieken van online haat en intimidatie beter te doorgronden en hier adequaat op te reageren.

Verder kunnen lokale netwerken en professionals een belangrijke rol spelen in het faciliteren van dialoog en participatie binnen gemeenschappen. Door het organiseren van gespreksgroepen, co-creatiesessies en inclusieve bijeenkomsten kan de betrokkenheid en het eigenaarschap van burgers worden versterkt. Dit draagt bij aan het creëren van veiligere online en offline omgevingen waarin verschillende stemmen gehoord worden. Tot slot zorgt een structurele investering in kennisdeling, capaciteitsopbouw en samenwerking tussen diverse disciplines en organisaties voor een robuuster en veerkrachtiger sociaal weefsel, dat beter in staat is om de gevolgen van online geweld vroegtijdig te signaleren en aan te pakken.

Richt preventie op normalisering en weerbaarheid, met oog voor diversiteit

Voorkom dat online haat, discriminatie of bedreiging als ‘gewoon’ wordt beschouwd. Preventieve acties moeten zich daarom richten op het versterken van positieve omgangsvormen en het ondersteunen van counterspeech. Belangrijk is dat deze aanpak afgestemd wordt op de verschillende gemeenschappen en doelgroepen, aangezien hun online mediagebruik en ervaringen met online geweld sterk kunnen verschillen. Op maat gemaakte interventies die rekening houden met culturele, sociale en digitale verschillen vergroten de effectiviteit van preventie en bevorderen inclusiviteit en weerbaarheid binnen diverse groepen.

Naast individuele preventie en bewustwording is het beschermen van de digitale publieke ruimte van wezenlijk belang. De stemmen die via online geweld doelgericht worden gemarginaliseerd of monddood gemaakt — zoals vrouwen, LHBTQIA+-personen, journalisten, activisten en mensen met een migratieachtergrond — verdienen bescherming en ruimte in het publieke debat. Zonder structurele inzet om die ruimte veilig te stellen en te herstellen, verschraalt het democratische gesprek en groeit maatschappelijke polarisatie. Een brede, samenhangende aanpak op samenlevingsniveau is nodig om deze maatschappelijke ontwrichting tegen te gaan.

Dat begint met het vergroten van de maatschappelijke bewustwording van de ernst en de impact van online geweld. Nog te vaak wordt online geweld afgedaan als een minder ernstige of vrijblijvende vorm van agressie, terwijl de psychologische en maatschappelijke gevolgen aanzienlijk kunnen zijn. Het zichtbaar en invoelbaar maken van deze problematiek is essentieel.

Zichtbaar en invoelbaar maken van online geweld

- **Te investeren in publiekscampagnes, bijvoorbeeld via animaties, persoonlijke verhalen van slachtoffers en (ex-)plegers, en mediaproducties zoals documentaires, podcasts of nieuwsitems.** Zulke verhalen maken de impact van online geweld tastbaar voor een breed publiek en dragen bij aan meer begrip, urgentiebesef en sociale normstelling.
- **Het ondersteunen van maatschappelijke rolmodellen, publieke figuren en influencers die zich uitspreken tegen online geweld is hierbij van grote waarde.** Hun zichtbaarheid en bereik kunnen bijdragen aan het versterken van een positief en inclusief online klimaat en bieden tegenwicht aan extremistische, polariserende en discriminerende boodschappen die het publieke debat kunnen domineren.
- **Het faciliteren van lotgenotencontact voor burgers en omstanders (bijvoorbeeld familie, vrienden etc.) die te maken hebben gehad met online geweld.** Dit kan zowel vanuit de kant van pleger (die later bijvoorbeeld tot het besef komen dat ze over de grens zijn gegaan) als vanuit de kant van slachtoffers.

Naast bewustwording is het van belang actief ruimte te creëren voor een veilig en pluriform maatschappelijk debat, zowel online als offline. Online geweld heeft een verstikkend effect op het publieke gesprek en leidt tot systematische uitsluiting van bepaalde stemmen.

‘Brave spaces’ faciliteren:

‘Brave spaces’ zijn plekken, fysiek en digitaal, waar uiteenlopende, soms schurende of controversiële opvattingen op een respectvolle en gelijkwaardige manier besproken kunnen worden (Stanlick 2015; Arao & Clemens 2023). Zulke veilige gespreksruimtes verdienen specifieke aandacht voor groepen die vaker het doelwit zijn van online geweld, zoals vrouwen, mensen met een migratieachtergrond, LHBTQIA+-personen, jongeren of mensen met een beperking. Ook op lokaal niveau kan hierin een belangrijke rol worden vervuld door gemeenten, maatschappelijke organisaties en gemeenschappen. Door bijvoorbeeld buurtbijeenkomsten, lokale meldpunten of trainingen voor jongerenwerkers te ondersteunen, kan op wijk- en buurtniveau worden gewerkt aan sociale weerbaarheid, ontmoeting en het tegengaan van polarisatie.

Sociale mediaplatforms en nieuwsmedia spelen hierin ook een sleutelrol. Zij kunnen actief bijdragen aan sociale veiligheid door transparanter te zijn over hun moderatiebeleid, de werking van hun algoritmes en de maatregelen die zij nemen tegen online haat en intimidatie. Overheden en maatschappelijke organisaties kunnen dit stimuleren door afspraken te maken over verantwoordelijk platformbeheer, bijvoorbeeld in de vorm van convenanten of gezamenlijke richtlijnen.

7 Literatuurlijst

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. Social Network Analysis and Mining, 13, 30. https://doi.org/10.1007/s13278-023-01028-5.

Arao, B., & Clemens, K. (2023). From safe spaces to brave spaces: A new way to frame dialogue around diversity and social justicae. In The art of effective facilitation (pp. 135–150). Routledge.

Baines, D., & Elliott, R. J. (2020). Defining misinformation, disinformation and malinformation: An urgent need for clarity during the COVID-19 infodemic. Discussion Papers, 20(06), 20–06.

Barbieri, F., Anke, L. E., & Camacho-Collados, J. (2021). XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond. arXiv preprint arXiv:2104.12250.

Beers, P. J., & van Mierlo, B. (2017). Reflexivity and learning in system innovation processes. Sociologia Ruralis, 57, 415–436. https://doi.org/10.1111/soru.12179.

Brubaker, R., & Cooper, F. (2000). Beyond "identity".

Chabalala, Y., Adam, E., & Ali, K. A. (2023). Exploring the effect of balanced and imbalanced multi-class distribution data and sampling techniques on fruit-tree crop classification using different machine learning classifiers. Geomatics, 3(1), 70–92.

Czymara, C. S., & Gorodzeisky, A. (2024). Hostility on Twitter in the aftermath of terror attacks. Journal of Computational Social Science, 7(2), 1305–1325.

De Vries, W., van Cranenburgh, A., Bisazza, A., Caselli, T., van Noord, G., & Nissim, M. (2019). Bertje: A Dutch BERT model. arXiv preprint arXiv:1912.09582.

Delobelle, P., Winters, T., & Berendt, B. (2020). Robbert: a Dutch RoBERTa-based language model. arXiv preprint arXiv:2001.06286.

Farkas, J., Schou, J., & Neumayer, C. (2018). Cloaked Facebook pages: Exploring fake Islamist propaganda in social media. New Media & Society, 20(5), 1850–1867.

García, S., Zhang, Z., Altalhi, A., Alshomrani, S., & Herrera, F. (2018). Dynamic ensemble selection for multi-class imbalanced datasets. Information Sciences, 445–446, 22–37. https://doi.org/10.1016/j.ins.2018.03.002.

González-Carvajal, S., & Garrido-Merchán, E. C. (2020). Comparing BERT against traditional machine learning text classification. arXiv preprint arXiv:2005.13012.

Huo, H., & Iwaihara, M. (2020, August). Utilizing BERT pretrained models with various fine-tune methods for subjectivity detection. In Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data (pp. 270–284). Cham: Springer International Publishing.

Ikonomakis, M., Kotsiantis, S., & Tampakas, V. (2005). Text classification using machine learning techniques. WSEAS Transactions on Computers, 4(8), 966–974.

Khan, A., Chaudhari, O., & Chandra, R. (2023). A review of ensemble learning and data augmentation models for class imbalanced problems: combination, implementation and evaluation. Expert Systems with Applications, 244, 122778. https://doi.org/10.48550/arXiv.2304.02858.

Kim, M., & Hwang, K. B. (2022). An empirical evaluation of sampling methods for the classification of imbalanced data. PLOS ONE, 17(7), e0271260.

Li, M., Li, W., Wang, F., Jia, X., & Rui, G. (2020). Applying BERT to analyze investor sentiment in stock market. Neural Computing and Applications, 33, 4663–4676. https://doi.org/10.1007/s00521-020-05411-7.

Liu, Y., & Lapata, M. (2019). Text summarization with pretrained encoders. arXiv preprint arXiv:1908.08345. https://doi.org/10.18653/v1/D19-1387.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.

Menshikova, A., & van Tubergen, F. (2022). What drives anti-immigrant sentiments online? A novel approach using Twitter. European Sociological Review, 38(5), 694–706.

Mierlo, B. C. van, Regeer, B., Amstel, M. van, Arkesteijn, M. C. M., Beekman, V., Bunders, J. F. G., Cock Buning, T. de, Elzen, B., Hoes, A. C., & Leeuwis, C. (2010). Reflexieve monitoring in actie. Handvatten voor de monitoring van systeeminnovatieprojecten. https://edepot.wur.nl/142966.

Nishigaki, D., Suzuki, Y., Wataya, T., Kita, K., Yamagata, K., Sato, J., Kido, S., & Tomiyama, N. (2023). BERT-based transfer learning in sentence-level anatomic classification of free-text radiology reports. Radiology: Artificial Intelligence, 5(2), e220097. https://doi.org/10.1148/ryai.220097.

Odekerken, M.W.A. (2017). Het incasseren van ongenoegen: deurwaarders en schuldenaren. Boom Juridisch. Dissertatie.

Pérez-Escolar, M., Lilleker, D., & Tapia-Frade, A. (2023). A systematic literature review of the phenomenon of disinformation and misinformation. Media and Communication, 11(2), 76–87. https://doi.org/10.17645/mac.v11i2.6453.

Rasmy, L., Xiang, Y., Xie, Z., Tao, C., & Zhi, D. (2020). Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. NPJ Digital Medicine, 4. https://doi.org/10.1038/s41746-021-00455-y.

Riyanto, S., Imas, S. S., Djatna, T., & Atikah, T. D. (2023). Comparative analysis using various performance metrics in imbalanced data for multi-class text classification. International Journal of Advanced Computer Science and Applications, 14(6).

Roggeband, C., & van der Haar, M. (2018). "Moroccan youngsters": Category politics in the Netherlands. International Migration, 56(4), 79–95.

Rossini, P. (2020). Beyond toxicity in the online public sphere: Understanding incivility in online political talk. In A research agenda for digital politics (pp. 160–170).

Sellars, A. (2016). Defining hate speech. Berkman Klein Center Research Publication, (2016-20), 16–48.

Stanlick, S. (2015). Getting "real" about transformation: The role of brave spaces in creating disorientation and transformation. Ann Arbor, MI: Michigan Publishing, University of Michigan Library.

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019, October). How to fine-tune BERT for text classification? In China National Conference on Chinese Computational Linguistics (pp. 194–206). Cham: Springer International Publishing.

Sun, Z., Song, Q., Zhu, X., Sun, H., Xu, B., & Zhou, Y. (2015). A novel ensemble method for classifying imbalanced data. Pattern Recognition, 48, 1623–1637. https://doi.org/10.1016/j.patcog.2014.11.014.

Taneja, K., & Vashishtha, J. (2022, March). Comparison of transfer learning and traditional machine learning approach for text classification. In 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 195–200). IEEE.

Van Dijk, R., Hofman, E., Ruiter, S., & Verbeek, J. (2023, September 6). 'Tijd voor geweld tegen deze lui.' De Groene Amsterdammer. https://www.groene.nl/artikel/tijd-voor-geweld-tegen-deze-lui.

Wiemers, S. A., Stasio, V. D., & Veit, S. (2024). Stereotypes about Muslims in the Netherlands: an intersectional approach. Social psychology quarterly, 87(4), 440-460.

Wu, Q., & Xu, B. (2024). A text classification model based on BERT and TextCNN. Proceedings of SPIE, 13184, 131841E. https://doi.org/10.1117/12.3032898.

Yu, S., Su, J., & Luo, D. (2019). Improving BERT-based text classification with auxiliary sentence and domain knowledge. IEEE Access, 7, 176600–176612. https://doi.org/10.1109/ACCESS.2019.2953990.

Zhao, X., An, Y., Xu, N., & Geng, X. (2024). Variational continuous label distribution learning for multi-label text classification. IEEE Transactions on Knowledge and Data Engineering, 36, 2716–2729. https://doi.org/10.1109/TKDE.2023.3323401.

Bijlage 1 Social media analyse

B.1.1 Uitgebreide beschrijving sociale media analyse

In dit onderzoek hebben we met Natural Language Processing (NLP) meer dan tienduizenden reacties onder nieuwsberichten en posts (voorheen tweets) op X geanalyseerd om inzicht te krijgen in het gebruik van potentieel discriminerend, beledigend of onbeschoft taalgebruik. Hiervoor hebben we reacties onder artikelen van nu.nl en posts op X geanalyseerd. We hebben hierbij naar gevoelige onderwerpen gekeken om te kijken hoe mensen hier online over praten. Hieronder zullen we de uitgevoerde analyses stap voor stap beschrijven. We beginnen eerst met het toelichten van NLP: wat is het precies en wat komt daarbij kijken? Vervolgens beschrijven we voor zowel de data van nu.nl als de data van X hoe we de teksten hebben geanalyseerd en wat het oplevert.

Wat is NLP en waarom gebruiken we het?

Natural Language Processing (NLP) is de discipline binnen kunstmatige intelligentie die zich richt op het automatisch lezen, interpreteren en verwerken van menselijke taal. In essentie stelt NLP ons in staat om grote hoeveelheden ongestructureerde tekst om te zetten in data die computers kunnen begrijpen en analyseren (Jurafsky & Martin, 2020). Door taalkundige en statistische methoden te combineren, kan een computer bijvoorbeeld vaststellen welke onderwerpen in een tekst voorkomen, welke sentimenten er spelen of welke schrijfstijl wordt gehanteerd.

Aan de hand van NLP kan de complexiteit binnen tekstuele documenten geanalyseerd worden. Dit stelt ons in staat om aspecten zoals grammatica, zinsstructuur, semantiek, context en toon op een systematische wijze te bekijken. Computers ‘begrijpen’ taal niet zoals mensen dat doen, maar kunnen met behulp van statistische en taalkundige technieken wel vaststellen wat iemand bedoelt, hoe iemand zich voelt, of welke onderwerpen in een tekst aan bod komen.

NLP kent een breed scala aan toepassingen. Zo wordt het bijvoorbeeld gebruikt voor het automatisch modereren van online discussies, het detecteren van haatspraak, het analyseren van klanttevredenheid in reviews of het volgen van politieke sentimenten op sociale media. In zakelijke en juridische contexten ondersteunt NLP het samenvatten van documenten, het herkennen van contractvoorwaarden of het classificeren van e-mails.

Sociale media-data bestaat uit enorme hoeveelheden tekst. Elke moment van de dag worden er duizenden posts geplaatst over allerlei onderwerpen. Vanwege de grote hoeveelheid posts is het niet haalbaar om het allemaal handmatig te analyseren. Door NLP toe te passen kunnen we analyses automatiseren en op een systematische manier bekijken hoe er haatdragende, discriminerende of onbeschofte taal wordt gebruikt in duizenden posts of reacties, of in welke context bepaalde groepen besproken worden. Dit maakt het mogelijk om maatschappelijke ontwikkelingen of online spanningen vroegtijdig te signaleren. De kracht van NLP ligt daarmee in de schaalbaarheid: het stelt ons in staat om niet alleen kwantitatieve inzichten te genereren uit tienduizenden teksten, maar ook om inhoudelijk betekenis te extraheren.

Verzamelen en voorbereiden van de data

Voordat we beschrijven hoe we de data precies hebben verzameld voor zowel nu.nl als X, beschrijven we eerst een aantal procedures die standaard zijn bij het uitvoeren van analyses met NLP. We hebben alle analyses uitgevoerd in de software R.

Voordat we de inhoud van reacties onder nieuwsberichten en posts echt konden onderzoeken, namen we allereerst een cruciale stap: het schoonmaken van alle teksten. Zonder deze voorbereiding zou elke woordtelling, sentimentanalyse of lexiconcheck worden verstoord door ruis—zoals gebruikersnamen, hyperlinks of toevallige leestekens—die niets zeggen over de betekenis van een post. Daarom hebben we alle elementen die mogelijk een risico vormen op de interpretatie van de posts, systematisch verwijderd. Elementen als @gebruikersnamen en #hashtags verdwijnen volledig uit de tekst, net als URL-links, zodat alleen de eigenlijke inhoud overblijft.

Vervolgens hebben we alle leestekens, cijfers en speciale symbolen weggehaald. Dit voorkomt dat varianten op hetzelfde woord—denk aan “hond!” versus “hond”—als aparte woorden worden gezien. Door elk teken dat niet in het alfabet voorkomt te verwijderen, zorgen we ervoor dat onze analyses zich richten op de woorden zelf en niet op toevallige interpunctie of cijfers in een post.

Daarna brachten we uniformiteit aan in de schrijfwijze door alle letters om te zetten naar kleine letters. Op die manier worden “Nederland”, “nederland” en “Nederland!” automatisch op dezelfde manier geschreven in het databestand, wat ervoor zorgt dat identieke woorden op dezelfde manier worden gelezen. Omdat we uitsluitend geïnteresseerd zijn in woorden die inhoudelijk wat betekenen, pasten we ten slotte een uitgebreide stopwoordenfilter toe. Naast algemeen geaccepteerde Nederlandse lijsten (zoals Snowball en stopwords-iso) maakten we een contextspecifieke aanvulling met werkwoorden en bijwoorden die in onze datasets te vaak en te algemeen voorkwamen om inhoudelijk bij te dragen.

In de laatste fase van dit schoonmaakproces verwijderden we overbodige witruimtes—dubbele spaties, voor- en achterliggende spaties—zodat elke reactie en elk post werd teruggebracht tot een gestructureerd, opgeschoond tekstveld. Deze stappen vormen de basis voor alle daaropvolgende analyses. Alleen nadat een post deze reeks bewerkingen had doorlopen, wisten we zeker dat elke vorm van ruis was verwijderd en dat alleen relevante, betekenisvolle woorden zouden worden geanalyseerd.

Kwantificering van schadelijke taal met gespecialiseerde lexicons

Voor zowel de data van Nu.nl en van X hebben we gebruik gemaakt van gespecialiseerde lexicons om kwantificeren hoe vaak er schadelijke taal wordt gebruikt. Bij deze analyses zoeken we in de teksten naar woorden die afkomstig zijn uit vooraf samengestelde woordenlijsten (lexicons). In ons geval gebruiken we lexicons die schadelijke taal beschrijven, zoals scheldwoorden, discriminerende uitdrukkingen of dreigende taal. Op die manier kunnen we systematisch vaststellen hoe vaak en in welke context deze termen worden gebruikt, en of ze vaker voorkomen in bepaalde thema's of discussies.

Voor ons onderzoek hebben we gebruikgemaakt van de lexicons van The Weaponized Word . The Weaponized Word is een organisatie dat zich bezighoudt met problematisch taalgebruik op het Internet. Op basis van hun onderzoek hebben ze vier lijsten voor verschillende talen ontwikkeld, waaronder in het Nederlands. Ze hebben de vier lijsten ontwikkeld op basis van de aard van het taalgebruik:

1. Discriminerende woorden: woorden die specifiek gericht zijn op de ontvanger zijn nationaliteit, etniciteit, religie, gender, seksuele oriëntatie, beperking en/of chronische ziekte of sociaaleconomische klasse.
2. Denigrerende woorden: woorden die beledigend worden gebruikt, maar niet direct betrekking hebben op de sociale identiteit van de ontvanger.
3. Bedreigende woorden: woorden die (direct of indirect) dreigingen of geweld impliceren.
4. Watch words: signalerende woorden die samenhangen met haatspraak, of beledigend of dreigend taalgebruik.

The Weaponized Word

The Weaponized Word maakt gebruik van dynamische woordenlijsten die voortdurend worden bijgewerkt: van bekende haatwoorden en bedreigingen tot phishing-sjablonen en bronnen van misinformatie. Daarnaast 'leert' de techniek welke woordcombinaties en zinspatronen typisch zijn voor negatieve of schadelijke uitingen. Zo kan het snel herkennen wanneer iemand in een bericht discrimineert, beledigt, bedreigt of verkeerde informatie verspreidt, zonder dat er naar persoonlijke gegevens wordt gekeken. Op die manier helpt The Weaponized Word platformen gezond te houden: het beperkt de verspreiding van schadelijke content en geeft moderators de kans om onmiddellijk in te grijpen en de dialoog veilig te houden.

Dankzij deze aanpak kunnen beheerders van online gemeenschappen razendsnel zien welke bijdragen extra aandacht nodig hebben. In tientallen talen signaleren we woorden en uitdrukkingen die vallen onder de bovengenoemde lijsten.

Platform Nu.nl

Artikelindeling en thematische groepering

In de eerste fase van onze analyse koppelden we alle reacties aan de titels van de onderliggende Nu.nl-artikelen. Om te voorkomen dat we elk artikel als een afzonderlijke case zouden behandelen, groeperen we de artikelen op basis van zes overkoepelende thema's: LHBTQIA+, Klimaat, Buitenland, Immigratie/Asiel, Transgender personen en Overig. Deze thematische indeling komt niet overeen met de standaard indeling van artikelen op Nu.nl. Zo beoogden we reacties te analyseren die samenhangen met het bredere maatschappelijk debat. Deze stap levert een dataset op waarin iedere reactie zowel tekstuele inhoud als een inhoudelijke categorie bevat, waardoor we de verdeling van reacties – en hun moderatiestatus – per onderwerp kunnen onderzoeken.

Verdeling naar moderatiestatus binnen thema

Vervolgens bepaalden we voor elk thema hoeveel reacties door het automatische moderatiesysteem van Nu.nl waren geaccepteerd en hoeveel als afgekeurd werden gemarkeerd. Hiervoor telden we allereerst het totale aantal reacties per thema en legden daar vervolgens de splitsing naar geaccepteerd en afgekeurd naast. Deze kwantitatieve basis geeft direct inzicht in de reactiesecties waarin vaker posts worden gepost die de huisregels overtreden: thema's waarin een relatief groot deel van de reacties wordt verwijderd, zijn mogelijk vaker gevoeliger of roepen heftigere reacties op dan onderwerpen waar er minder sprake is van moderatie.

Huisregels van Nu.nl

Hieronder volgen de huisregels van Nu.nl. Een uitgebreidere toelichting van de regels kan teruggevonden op de site van Nu.nl. Ook zijn er nog webpagina's voor elke regel.

"**Welkom bij NUjij! Op ons reactieplatform onder de artikelen kun je reageren op het nieuws. Om de discussie in goede banen te leiden, vragen we je wel om je aan onze huisregels te houden. Anders verwijderen we je reactie. Eerst in een notendop de belangrijkste huisregels op een rij:**

1. Wees respectvol naar je medemens: iedereen mag er zijn en heeft bestaansrecht. NU.nl is geen plek voor beledigingen, racisme, homofobie, grof taalgebruik, laster, intimidatie of spam.
2. Blijf ontopic. Beperk je reactie tot het nieuwsonderwerp en haal er geen andere zaken bij. Kritiek op ons moderatiebeleid is welkom op nujij@nu.nl, niet hier
3. Reageer op iemands argumenten, niet op de persoon. Val ook niet iemands manier van schrijven aan.
4. Verspreid geen onwaarheden. Klimaatontkenning valt daar ook onder.
5. Onderbouw je mening met argumenten en bronnen. Leg in minstens honderd tekens uit waarom je iets vindt.
6. Presenteer je mening niet als feit. Laat je mening blijken met woorden als 'ik vind'.
7. Speculeer niet over ongevallen, rampen en (mogelijke) misdaden waar nog weinig over bekend is."

Detectie van schadelijke taal

Na het opschonen van de data hebben we bekeken hoe vaak woorden van de lexicons terugkomen in de reacties op de posts op Nu.nl. Met de volledig gestandaardiseerde reacties bepaalden we vervolgens in hoeverre potentieel schadelijke taal voorkomt per thematische categorie en per moderatiestatus. Hiervoor maakten we gebruik van het Weaponized Word-lexicon. Ten eerste telden we het totale aantal lexicontermen in de dataset per categorie. Ten tweede bepaalden we voor elke reactie of er ten minste één lexiconwoord voorkwam, waardoor we het percentage reacties met schadelijke taal konden vaststellen. Door deze tellingen, zowel voor de geaccepteerde als de afgekeurde subset uit te voeren, ontstond een gedetailleerd beeld van welke typen grensoverschrijdend taalgebruik het vaakst voorkomen en waar de filters al dan niet effectief zijn.

Hiermee kunnen we bijvoorbeeld zien dat in het Immigratie/Asiel-thema vooral discriminerende taal opduikt, terwijl in discussies over Transgender personen de nadruk meer ligt op beledigingen. Tot slot maakten we voor de afgekeurde reacties per lexiconcategorie een top-10 woordfrequentielijst, zodat de daadwerkelijke termen die tot verwijdering leiden concreet in beeld komen. Hierbij is het belangrijk om op te merken dat sommige artikelen en titels al woorden bevatten die ook terugkomen in de lexicons. Neem bijvoorbeeld het artikel over de vermoorde imam in Afrika. Het zou goed kunnen dat reageerders het woord "vermoorden" gebruiken. Dit woord komt ook voor een het lexicon, wat betekent dat de analyses dit woord zullen terugvinden in de reacties. Dit wil echter niet zeggen dat dit woord bedreigend is gebruikt: het kan ook gebruikt zijn om te bespreken waar het artikel over gaat.

Platform X

Data-verzameling met Zeeschuimer

Voor onze analyse van gesprekken op X hebben we gebruikgemaakt van Zeeschuimer , een gespecialiseerd scraping¬programma dat ontworpen is om grote hoeveelheden sociale¬media¬data automatisch te verzamelen. Zeeschuimer werkt via de officiële API's en aanvullende webcrawling, waarbij het rekening houdt met limieten en de taal van de posts. Met deze tool hebben we alle Nederlandstalige posts opgehaald die tussen 1 juli 2024 en 31 december 2024 zijn geplaatst, en waarin de termen "Wappie" en "Tokkie", "Moslim" of "Facist" voorkwamen. In totaal resulteerde dit in 236.230 verzamelde posts.

Zeeschuimer[18]

Zeeschuimer is een browserextensie ontworpen voor onderzoekers die systematisch content op sociale mediaplatforms willen bestuderen, vooral op platforms die conventionele scrapingmethoden of API-gebaseerde dataverzameling bemoeilijken. De extensie is ontwikkeld door Stijn Peters voor de Digital Methods Initiative, wat onderdeel is van de Universiteit van Amsterdam.

Momenteel kan het data verzamelen van de volgende platformen:

- TikTok (posts en reacties)
- Instagram (alleen posts)
- X/Twitter
- LinkedIn
- 9gag
- Imgur
- Douyin
- Gab
- Truth Social
- Pinterest
- RedNote/Xiaohongshu

18 Peters, S. (z.d.). Zeeschuimer [Computer software]. GitHub. https://github.com/digitalmethodsinitiative/zeeschuimer

De X regels

Hieronder volgen de huisregels van X. Een uitgebreidere toelichting van de regels kan teruggevonden op de site van X. Ook zijn er nog webpagina's voor elke regel.

"Het doel van X is openbare gesprekken mogelijk te maken. Geweld, intimidatie en andere soortgelijke soorten gedrag ontmoedigen mensen om zichzelf te uiten en verminderen uiteindelijk de waarde van wereldwijde openbare gesprekken. Het doel van onze regels is ervoor te zorgen dat iedereen vrij en veilig aan openbare gesprekken kan deelnemen.

Veiligheid

- *Gewelddadige taal*: Je mag niet dreigen met geweld of (lichamelijk) letsel, niet daartoe aanzetten, dit verheerlijken of een wens daartoe uiten. Meer informatie.
- *Gewelddadige en haatdragende entiteiten*: Je kunt je niet aansluiten bij de activiteiten van gewelddadige en haatdragende entiteiten of die promoten. Meer informatie.
- *Seksuele uitbuiting van kinderen*: Op X hebben we geen enkele tolerantie voor de seksuele uitbuiting van kinderen. Meer informatie.
- *Misbruik/intimidatie*: Je mag geen beledigende content delen, deelnemen aan de gerichte intimidatie van iemand of andere mensen daartoe aanzetten. Meer informatie.
- *Hatelijk gedrag*: Je mag andere mensen niet aanvallen op basis van hun ras, etniciteit, land van herkomst, kaste, seksuele geaardheid, geslacht, genderidentiteit, religieuze overtuiging, leeftijd, handicap of ernstige ziekte. Meer informatie.
- *Personen die gewelddadige aanvallen uitvoeren*: We zullen alle accounts verwijderen die worden beheerd door individuen die terroristische, gewelddadig extremistische of massale gewelddadige aanvallen uitvoeren, en kunnen ook posts verwijderen die manifesten verspreiden of andere content die door aanvallers is geproduceerd. Meer informatie.
- *Zelfmoord*: Je mag geen zelfmoord of automutilatie promoten of aanmoedigen. Meer informatie.
- *Gevoelige media*: Je mag geen media posten die overmatig bloederig is of gewelddadige of volwassen content delen in live video›s of in de afbeeldingen in het profiel of in de koptekst. Media die seksueel geweld en/of aanvallen tonen, zijn ook verboden. Meer informatie.
- *Illegale of bepaalde gereguleerde producten of diensten*: Je mag onze dienst niet gebruiken voor elk onwetmatig doel of ter bevordering van illegale activiteiten. Daartoe behoren het verkopen, kopen of vergemakkelijken van transacties illegale producten of diensten en van bepaalde soorten gereguleerde producten of diensten. Meer informatie."

Verkenning met wordclouds

Nadat de originele posts waren verzameld, doorliepen alle posts een gestandaardiseerd schoonmaakproces om ruis en overbodige onderdelen te verwijderen. Om per onderwerp snel te ontdekken welke termen het meest geassocieerd worden met "Wappie + Tokkie", "Moslim" en "Facist", maakten we voor elke zoekterm een wordcloud. In zo'n afbeelding worden de vijftig meest voorkomende woorden visueel uitgelicht: grotere woorden komen vaker voor. Dit helpt om in één oogopslag te zien welke thema's, gevoelens of discussies domineren rond de verschillende zoekopdrachten. Zo toonden de wordclouds bijvoorbeeld of woorden als "gezondheid", "stem" of "opkomst" opspringen in de context van "Wappie + Tokkie", of juist termen als "geloof", "vrijheid" of "terreur" in discussies over "Moslim" en "Facist".

Detectie van schadelijke taal

Naast deze verkennende weergaven wilden we ook kwantitatief vaststellen hoe vaak bepaalde vormen van potentieel schadelijke taal voorkomen. Hiervoor gebruikten we weer het Weaponized Word-lexicon. Per post bepaalden we of er ten minste één woord uit elk van deze lijsten in voorkwam, en berekenden we welk percentage van alle verzamelde posts een dergelijke term bevatte. Hiermee blijkt niet alleen welke categorieën taal het meest domineren op X, maar ook in welke mate mensen zich over de drempel van beleefdheid en fatsoen heen bewegen. Daarnaast hebben we voor iedere lexiconcategorie bekeken wat de tien meest voorkomende woorden zijn. Dit leverde per lexicon inzicht in de frequentie waarmee specifieke termen worden gebruikt. Daarmee krijgen we een gedetailleerd beeld van de woorden die echt worden gebruikt in posts rondom bepaalde onderwerpen.

B.1.2 Resultaten posts Nu.nl en X

B.1.2.1. Analyses van de posts op Nu.nl

Reactie gerelateerd aan categorie berichten

In totaal zijn 37.061 reacties op Nu.nl geanalyseerd. Veruit de meeste reacties binnen het databestand zijn geplaatst onder een bericht dat viel in de categorie 'Immigratie/Asiel'. Daarna stonden de meeste reacties onder berichten over 'Transgender personen' en 'Buitenland'. De minste reacties zijn geplaatst onder berichten in de categorie 'Overig'. De bovenstaande figuur toont aan dat *de woorden uit de lexicons vaker voorkomen in de afgekeurde berichten*, zeker als we kijken naar woorden die discriminatie en bedreiging aanduiden.

Zie hieronder in tabel B.1 de verdeling van de reacties per categorie bericht:

Tabel B1. Aantal reacties per categorie bericht

Catergorie	N
Immigratie/Asiel	20,079
Transgender personen	4,862
Buitenland	3,279
Klimaat	3,245
LHBTIAQ+	2,901
Overig	2,695
Totaal	**37.061**

Zoals hierboven is te zien, zijn de meeste posts, reacties onder berichten die gingen over *immigratie en asiel*. Dit is deels te verklaren doordat we de reacties onder acht artikelen binnen deze categorie hadden ontvangen. De overige categorieën hadden maximaal vijf artikelen met reacties.

Reacties onder berichten over het buitenland, LHBTQIA+ en transgender personen het vaakst afgekeurd

Daarnaast is specifieker gekeken naar de reacties die zijn goedgekeurd en de reacties die zijn afgekeurd. Over het algemeen zien we dat er *meer goedgekeurde dan afgekeurde berichten* zijn: iets meer dan 77% van de berichten is goedgekeurd door de AI moderatie en het multidisciplinaire moderatie team van Nu.nl. Wel zien we dat er verschillen zijn tussen de onderwerpen. Zo werd meer dan een derde van de reacties in de categorie 'Buitenland' afgekeurd. Daarnaast werd iets minder dan een derde van de reacties uit de categorieën 'LHBTQIA+' en 'Transgender personen' afgekeurd. Daarna zien we de andere drie categorieën, die een stuk minder afgekeurde reacties bevatten: bijna een vijfde van de reacties in de categorie 'Klimaat' en 'Immigratie/Asiel' werd afgekeurd.

In de volgende tabel is het aantal goed en afgekeurde reacties per categorie berichten uiteengezet:

Tabel B2. Aantal goed en afgekeurde reacties per categorie bericht

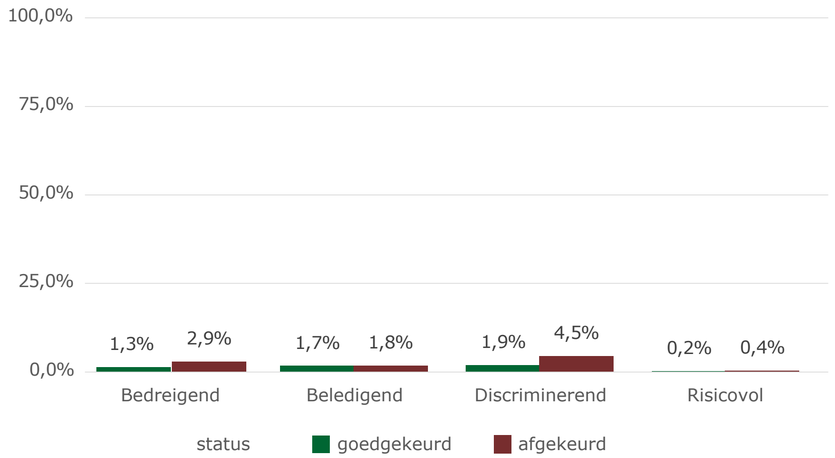
Categorie	Goed (N)	Afgekeurd (N)	Totaal	Goed (%)	Afgekeurd (%)
Buitenland	2,039	1,240	3,279	62.2	37.8
LHBTQIA+	2,018	883	2,901	69.6	30.4
Transgender personen	3,394	1,468	4,862	69.8	30.2
Klimaat	2,612	633	3,245	80.5	19.5
Immigratie/Asiel	16,460	3,619	20,079	82.0	18.2
Overig	2,249	446	2,695	83.5	16.5
Totaal	**28.772**	**8.289**		**-**	**-**

De bovenstaande tabel tonen aan dat reacties onder artikelen in de categorie *'Buitenland'* het vaakst worden afgekeurd. De artikelen binnen deze categorie gingen over homoseksualiteit in Afrika, ontwikkelingshulp, en een maatregel van de Taliban ten aanzien van vrouwen.

Discriminerende woorden komen het vaakst voor in afgekeurde berichten

Vervolgens hebben we gekeken hoe vaak de woorden uit de lexicons terugkomen in de reacties op Nu.nl. Allereerst is gekeken naar het onderscheid tussen de goedgekeurde en afgekeurde reacties. Over het algemeen zien we dat de woorden uit de lexicons niet vaak voorkomen in de reacties (zie Figuur 1). Wel zien we dat de woorden uit de lexicons vaker voorkomen in de afgekeurde reacties dan in de goedgekeurde reacties. Bij het lexicon met discriminerende woorden is het verschil tussen de goedgekeurde (1,9%) en afgekeurde (4,5%) groter dan bij de andere lexicons.

Figuur 1. Percentage berichten met lexicon-termen per status

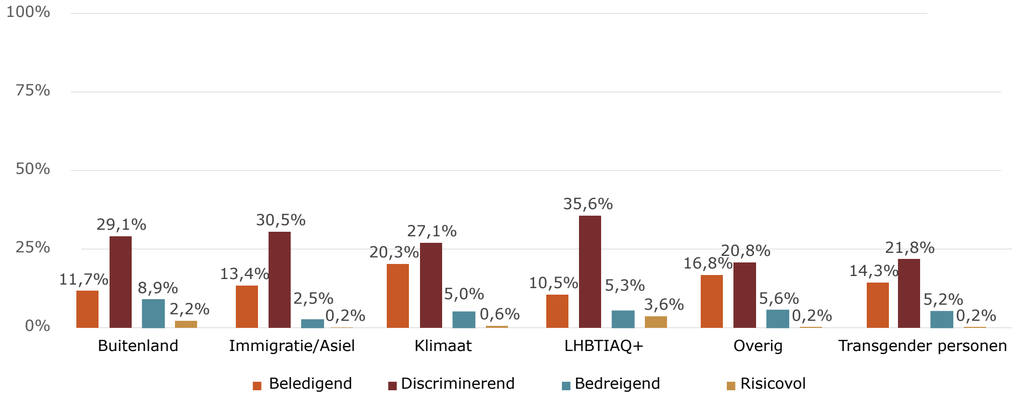


De bovenstaande figuur toont aan dat *de woorden uit de lexicons vaker voorkomen in de afgekeurde berichten*, zeker als we kijken naar woorden die discriminatie en bedreiging aanduiden.

Discriminerende woorden komen het vaakst voor

Verder hebben we gekeken hoe vaak de lexicons voorkomen per thema (Figuur 2). Ook hier zien we weer dat *discriminerende woorden* vaker voorkomen in de reacties dan woorden uit de andere lexicons. Discriminerende woorden komen het vaakst voor in reacties onder artikelen in de *LHBTQIA+ categorie* (35,6%), en het komt het minst voor bij de categorie 'Overig' (20,8%). Verder zien we in elke categorie hetzelfde patroon: *na discriminerende woorden komen beledigende woorden het vaakst voor, gevolgd door bedreigende woorden en risicowoorden*. Waar discriminerende woorden relatief het vaakst voorkomen in de categorieën 'LHBTQIA+', 'Immigratie/Asiel' en 'Buitenland', komen beledigende woorden vaker voor in de categorieën 'Klimaat', 'Overig' en 'Transgender personen'.

Figuur 2. Percentage reacties met lexicon-termen per categorie



Deze analyses tonen aan dat discriminerende woorden online het vaakst voorkomen, ongeacht het onderwerp. Denigrerende woorden komen hierna het vaakst voor.

B.1.2.2 Analyses van de posts op X

Hieronder presenteren we de analyses van de posts van X. We beginnen met de zoekterm 'Moslim', waarvoor we 121.645 posts hebben verzameld. Daarna kijken we naar de resultaten voor de zoekterm 'Fascist', waarvoor we 57.723 posts hebben verzameld. Tenslotte kijken we naar de resultaten voor de zoekterm 'Wappie' en 'Tokkie', waarvoor we 56.892 posts hebben verzameld.

A. Moslim

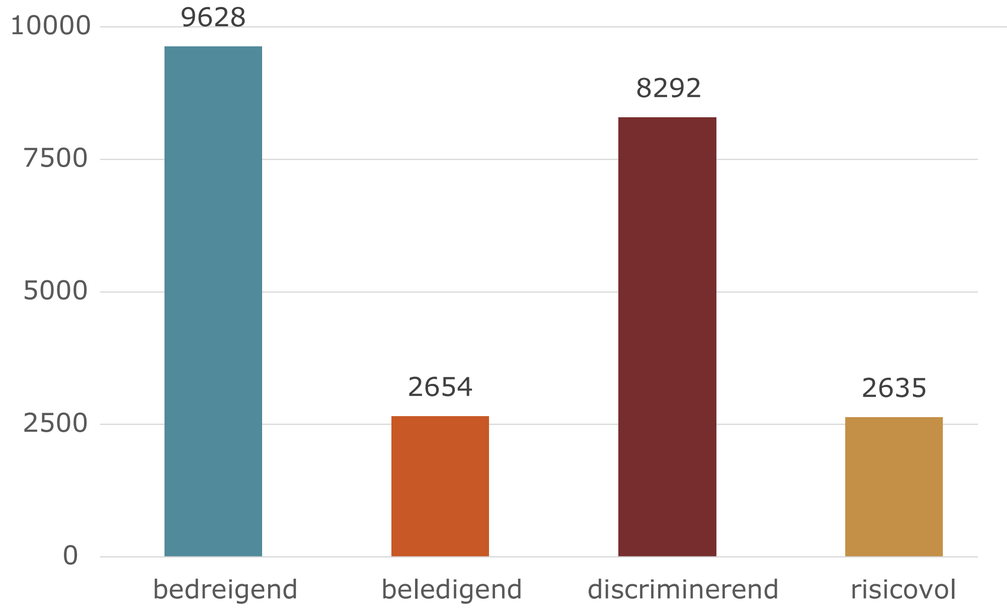
Allereerst hebben we een wordcloud gemaakt om te kijken welke woorden vaak voorkomen in posts waar het woord 'moslim' in voorkomt. Woorden die groter zijn weergeven, komen vaker voor in de posts die het woord 'Moslim' bevatten. Zoals we kunnen zien wordt er vaak gesproken over de politiek (woorden 'links', 'wilders', 'linkse', 'pvv, 'rechts') en andere religieuze gemeenschappen (woorden 'joden', 'joodse', 'christen').

Afbeelding 1 Wordcloud - Alle Tweets Moslim

Bedreigende woorden komen het vaakst voor de zoekterm 'Moslim'

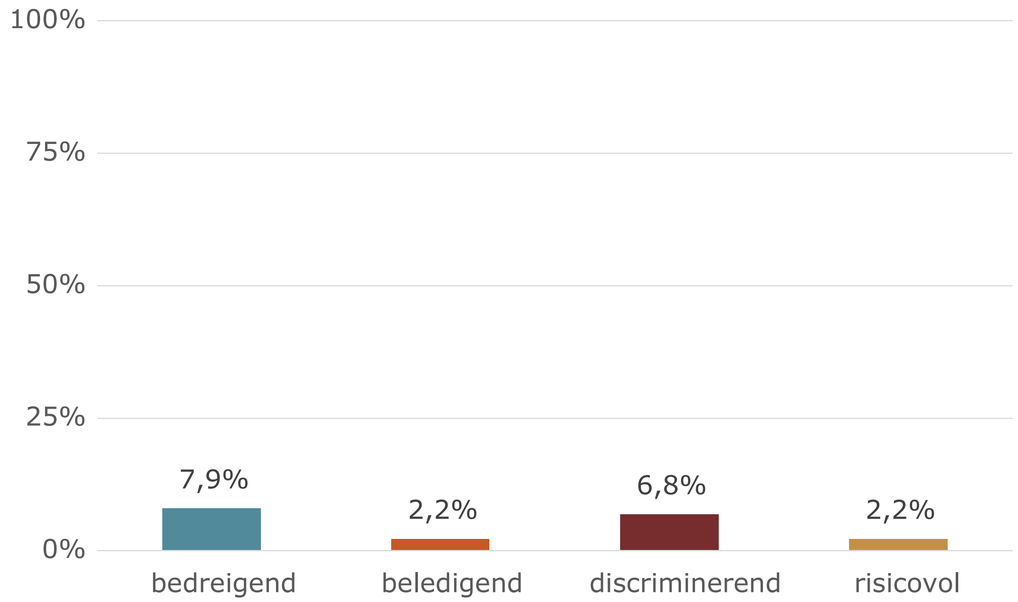
In de onderstaande figuur 3 zien we de absolute aantallen waarin termen uit elk van de vier lexiconcategorieën voorkomen in de verzamelde posts over "Moslim" (N = 121.615). Woorden uit de *categorie 'bedreigend'* worden het meest vaak gebruikt (N = 9.628), gevolgd door discriminerende woorden (N = 8.292). De categorieën 'beledigend' (N = 2.654) en 'risicovol' (N = 2.635) blijven daar duidelijk bij achter en komen in veel mindere mate voor.

Figuur 3. Aantal tweets met lexicon-termen Moslim



Figuur 4 laat hetzelfde zien maar dan in percentages: het percentage van alle posts die een of meer lexiconwoorden bevatten. Ook hier zien we dat bedreigende woorden (7,9%) het vaakst in een post opduiken. Direct daarachter komen discriminerende woorden (6,8 %) het vaakst voor, waarna zowel de beledigende als risicovolle ongeveer even vaak voorkomen (2,2 %). Hoewel deze percentages relatief laag lijken, tonen ze aan dat bijna één op de twaalf berichten rond "Moslim" gewelddadige of discriminerende taal bevat.

Figuur 4. Percentage tweets met lexicon-termen Moslim

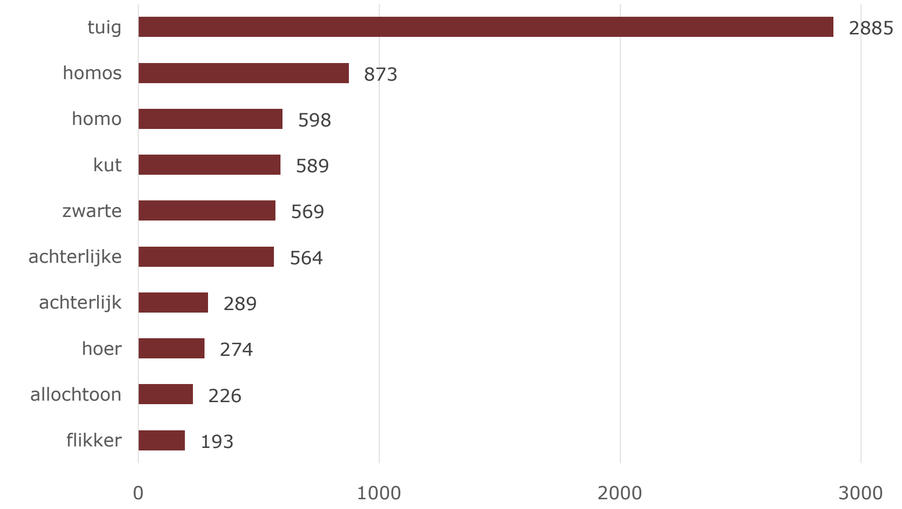


Deze bevindingen suggereren dat er vaak *dreigende taal* wordt gebruikt in posts waarin het woord 'moslim' voorkomt. Wanneer we dit breder interpreteren, kan dit erop wijzen dat dreigende taal wordt gebruikt om het gedrag of de denkwereld van moslims te beschrijven, maar het kan ook betekenen dat dreigende taal juist wordt geuit richting moslims. Hierbij lijken de tweets in te spelen op de aannames dat moslims 'tuig' zijn of vaker crimineel gedrag vertonen.

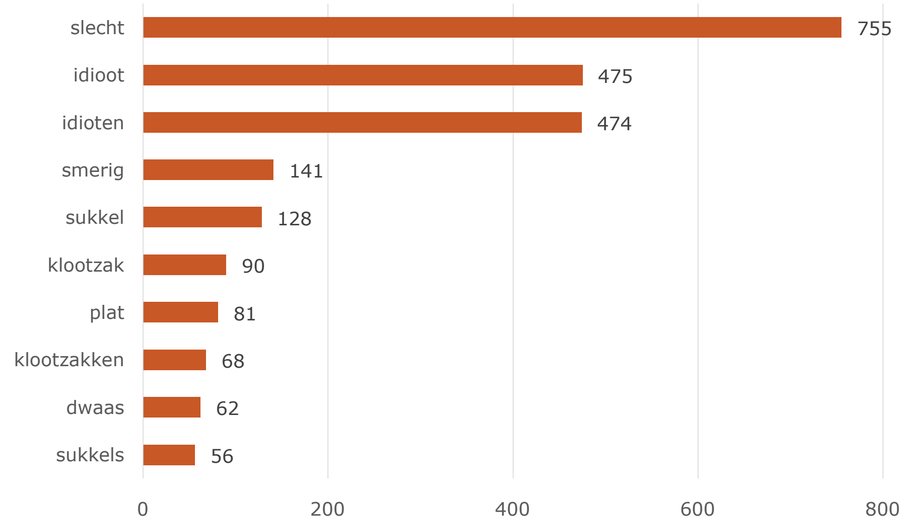
Het woord 'Moslim' veruit het vaakst gebruikt samen met het woord 'tuig'; dreigende taal gaat vaak over 'dood' en 'moord'

Vervolgens hebben we naar de meest voorkomende woorden gekeken voor elke lexicon. Als we kijken naar *de meest voorkomende discriminerende woorden*, dan komt het woord 'tuig' veruit het meest voor. Sterker nog, het woord 'tuig' kwam in 2885 posts voor, wat betekent dat het in meer dan een derde van de berichten voorkwam. De woorden 'homos' en 'homo' kwam een stuk minder vaak voor. Bij de categorie 'beledigend' zien we de aanwezigheid van veel scheldwoorden. Bij de woorden uit de categorie 'bedreigend' zien we dat het vaak gaat over de 'dood', 'doden' of 'vermoorden. Hierbij is het weer belangrijk om op te merken dat deze woorden niet per se discriminerend, beledigend of bedreigend gebruikt hoeven te zijn. Gebruikers kunnen ook een bepaald woord uit een nieuwbericht herhalen of ze doen een oproep om deze woorden niet te gebruiken.

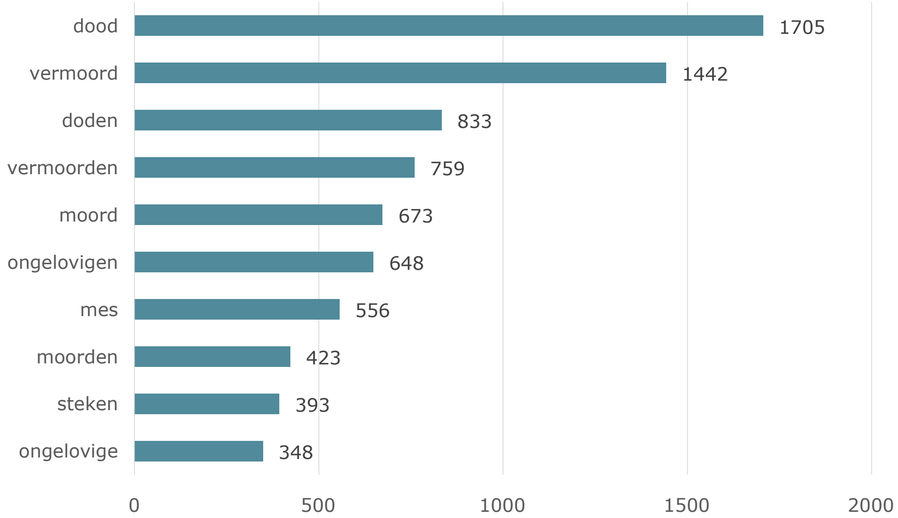
Figuur 5. **Top 10 discriminerende woorden Moslim**



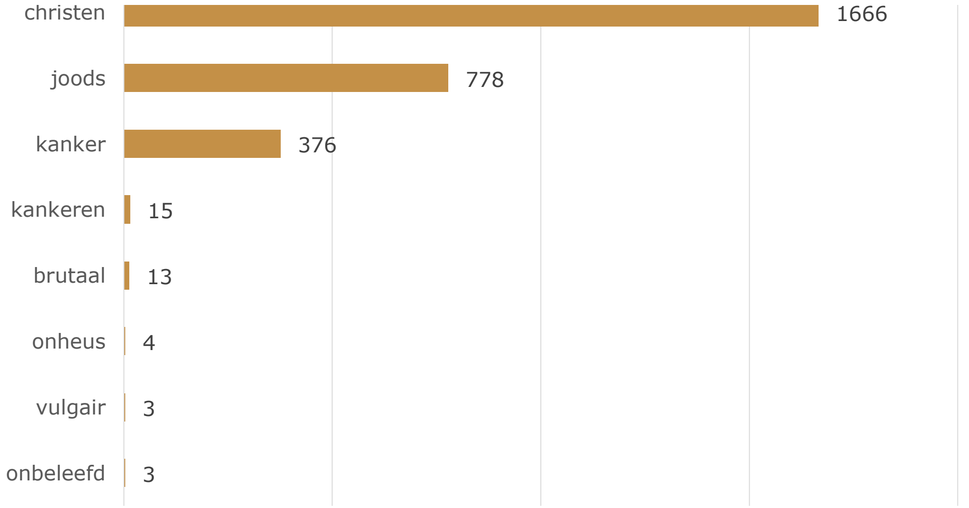
Figuur 6. **Top 10 beledigende woorden Moslim**



Figuur 7. **Top 10 bedreigende woorden Moslim**



Figuur 8. **Top 10 risicolvolle woorden Moslim**



B. Fascist

Onderstaande wordcloud toont de meest voorkomende termen in berichten waarin het woord "fascist" voorkomt. Het woord racist valt daar het meest op, wat aangeeft dat gebruikers fascisme vaak verbinden met racisme. Daarnaast zien we vaker de woorden "links" en "linkse" dan "rechts" of "rechtse", wat erop wijst dat het begrip "fascist" zowel aan linkse als aan rechtse politieke opvattingen wordt gekoppeld. Ook komen de termen "extreemrechts" en "nazi" voor.

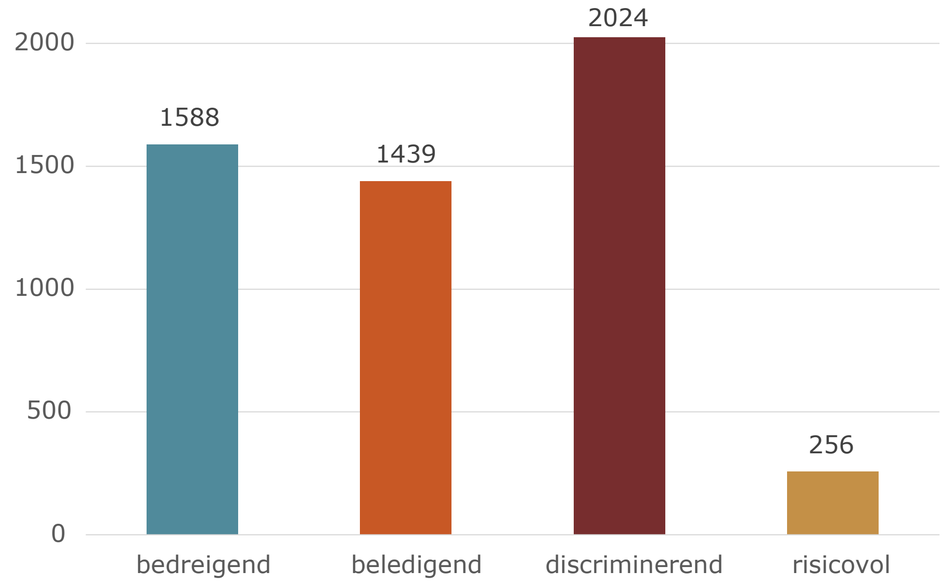
Verder verschijnen namen van politieke figuren zoals "Wilders", "Trump", "Bosma", "Poetin" en "PVV" in de wordcloud. Dit laat zien dat gebruikers het woord "fascist" op een bepaalde manier gebruiken in combinatie met deze personen en partijen. Werkwoorden als "uitgemaakt", "uitmaken", "genoemd" en "roepen" komen eveneens voor, wat kan suggereren dat iemand een ander "fascist" noemt. Tegelijkertijd duiken woorden als "haat" en "democratie" op. Tot slot zijn er ook beledigende termen zoals "vieze", "domme", "dom" en "smerige" te zien. Deze suggereert dat "fascist" soms gebruikt kan worden als belediging.

Afbeelding 2. Worldcloud - Alle Tweets Fascist

Discriminerende woorden worden het vaakst gebruikt in posts met het woord 'Fascist'

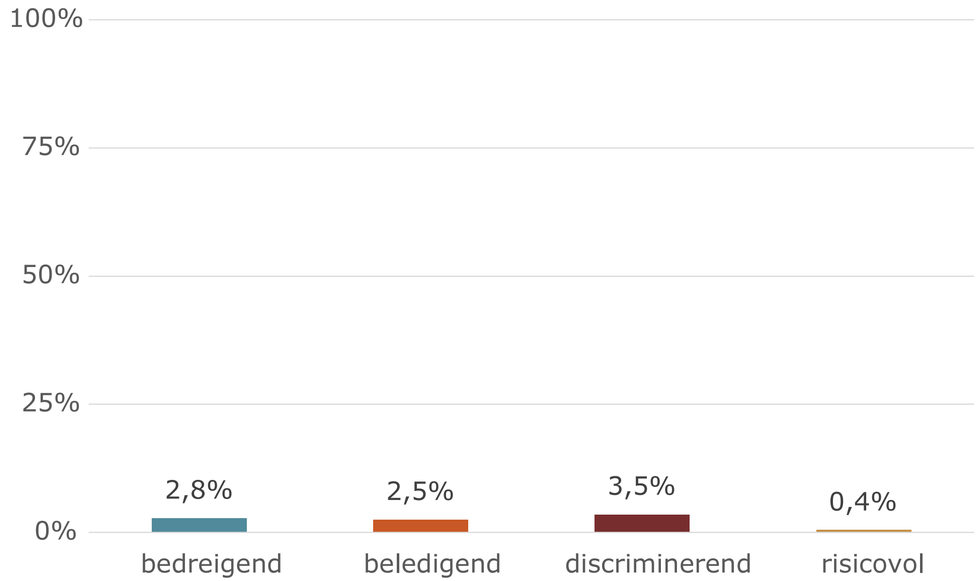
In figuur 9 zien we de absolute aantallen waarin termen uit elk van de vier lexiconcategorieën voorkomen in de verzamelde posts over "Facist" (N = 5 307). Discriminerende woorden worden hier het vaakst gebruikt (N = 2 024), gevolgd door bedreigende termen (N = 1 588). Beledigende uitdrukkingen komen op de derde plaats (N = 1 439), terwijl risicovolle woorden met N = 256 duidelijk minder frequent zijn.

Figuur 9 Aanal tweets met lexicon-termen Fascist



Figuur 10 toont hetzelfde overzicht, maar dan als percentage van alle posts waarin ten minste één lexiconwoord voorkomt. Ook hier blijkt dat discriminerende taal met 3,5 % van de berichten het meest dominant is. Bedreigende woorden volgen met 2,8 %, terwijl beledigingen met 2,5 % op de derde plek staan. Risicovolle termen duiken met slechts 0,4 % bijna niet op in de discussies rond “Facist”.

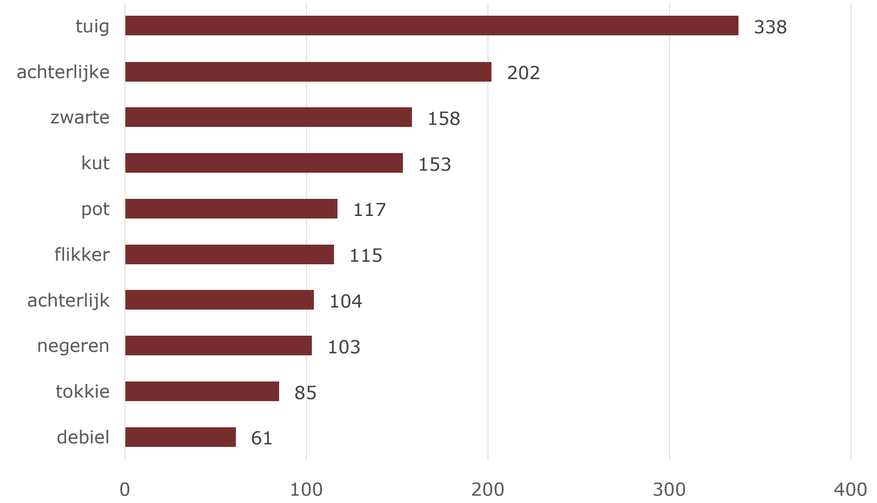
Figuur 10 Percentage tweets met lexicon-termen Fascist



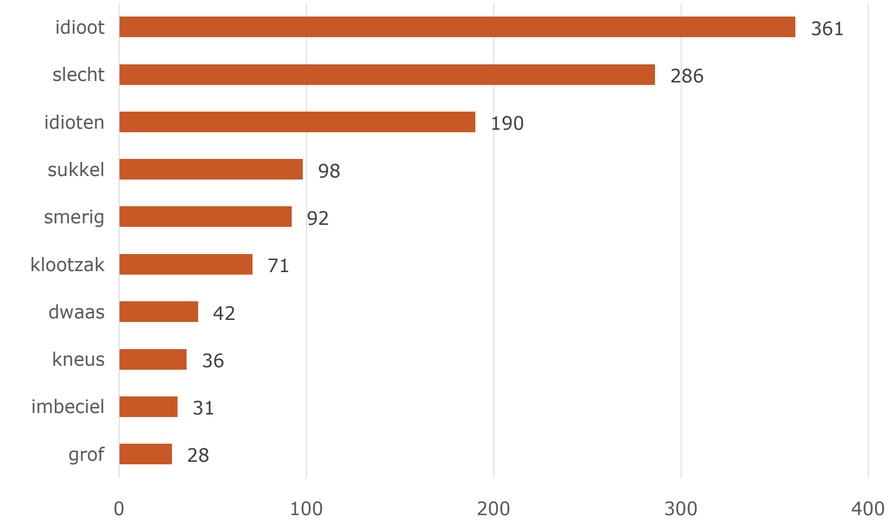
‘Dood’ het vaakst genoemd, gevolgd door ‘tuig’ en ‘crimineel’

Wederom hebben gekeken naar de tien *meest voorkomende woorden*. Ook hier zien we dat het woord ‘tuig’ ook hier het vaakst voorkomt. Dit woord komt ongeveer anderhalf keer zo vaak voor als de andere woorden binnen deze categorie. Wat betreft bedreigende woorden zien we ook hier dat de woorden ‘dood’, ‘vermoorden’ en ‘moord’ vaak voorkomen, maar hier wordt er ook vaker gerefereerd naar criminaliteit.

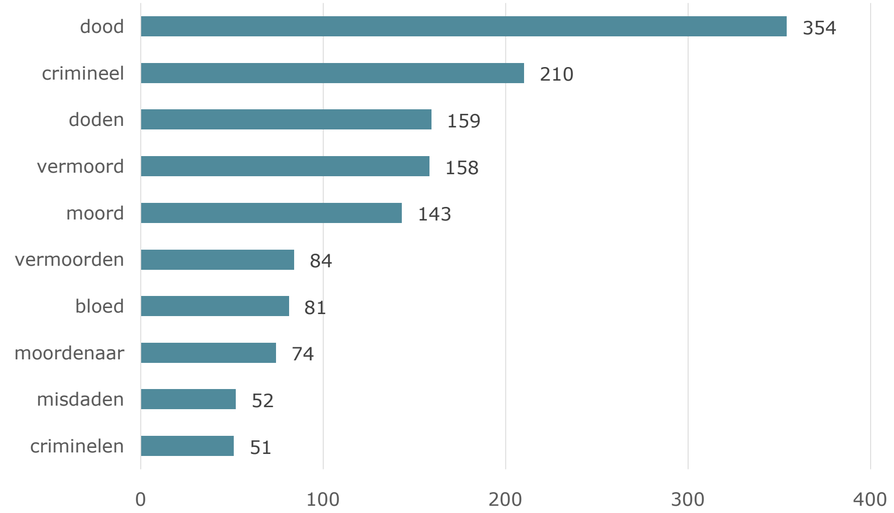
Figuur 11. Top 10 discriminerende woorden Fascist



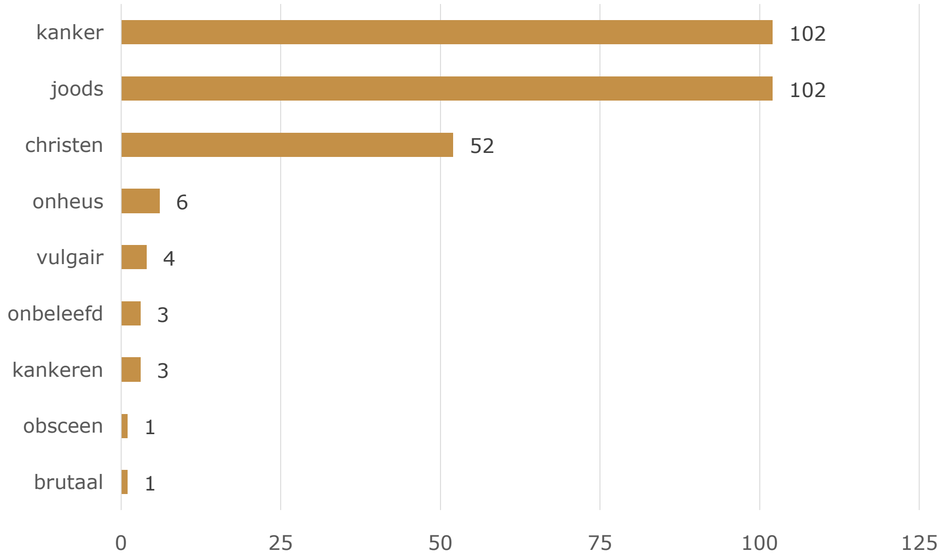
Figuur 12. Top 10 beledigend woorden Fascist



Figuur 13. Top 10 bedreigende woorden Fascist



Figuur 14. Top 10 risicolvolle woorden Fascist



C. Wappie + Tokkie

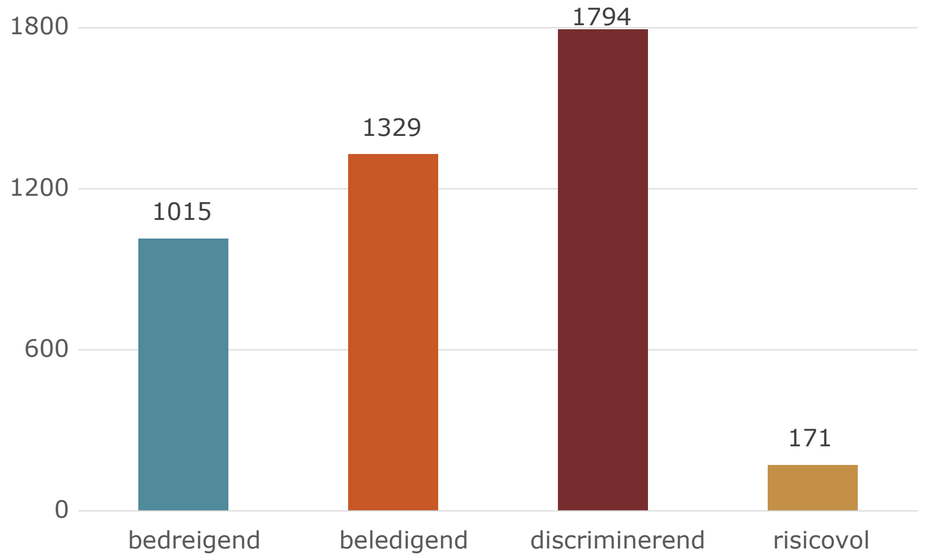
In onderstaande wordcloud staan de termen die het meest voorkomen in posts rondom “wappie” en “tokkie”. Zowel “links” als “rechts” zijn prominent aanwezig, net als “covid” en “corona”. Dit suggereert dat het in verband wordt gebracht met de politiek en de pandemie van drie jaar geleden. Verder zien we woorden als “media”, “onderzoek” en “waarheid”, gevolgd door “bewijs”, “complot” en “complotdenker”, die mogelijk verwijzen naar het gebruik van “feiten” met betrek tot bepaalde onderwerpen. Geografische aanduidingen zoals “Nederland” en “Rusland” komen eveneens voor, terwijl namen als “Trump”, “FvD” en “PVV” aangeven dat politieke actoren in deze context regelmatig worden genoemd. Tot slot zijn er beoordelende termen zoals “gek”, “dom” en “domme”, die duiden op negatief geladen taalgebruik.

Afbeelding 3. Wordcloud alle tweats ‘Wappie’en ‘Tokkie’

Discriminerende woorden het vaakst gebruikt in posts met het woord ‘Wappie’ of ‘Tokkie’

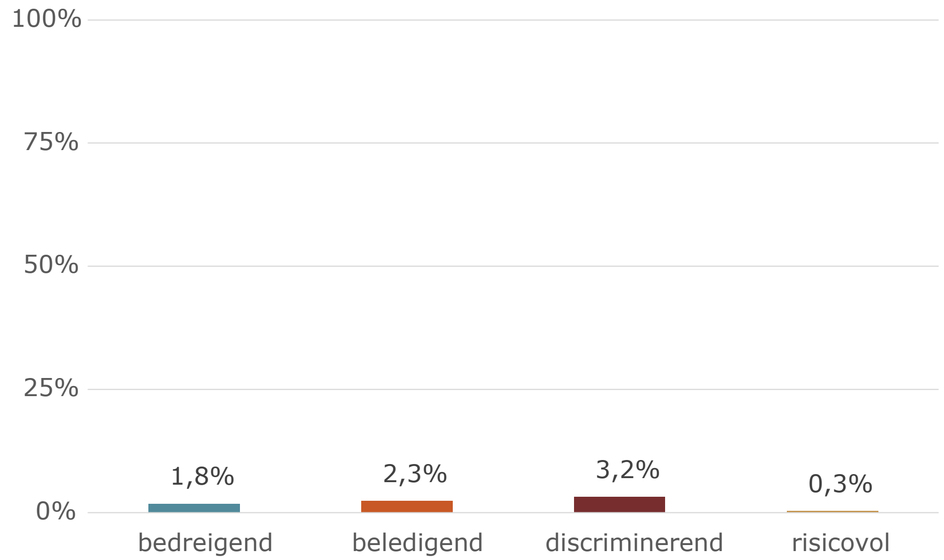
In figuur 15 zien we de absolute aantallen waarin termen uit elk van de vier lexiconcategorieën voorkomen in de verzamelde posts over “Wappie” en “Tokkie”. *Discriminerende woorden* zijn met N = 1794 het vaakst gebruikt, gevolgd door beledigende uitdrukkingen (N = 1329). Bedreigende termen komen op de derde plaats (N = 1015), terwijl risicovolle woorden met N = 171 beduidend minder voorkomen.

Figuur 15. Aantal tweets met lexicon-termen ‘Wappie’en ‘Tokkie’



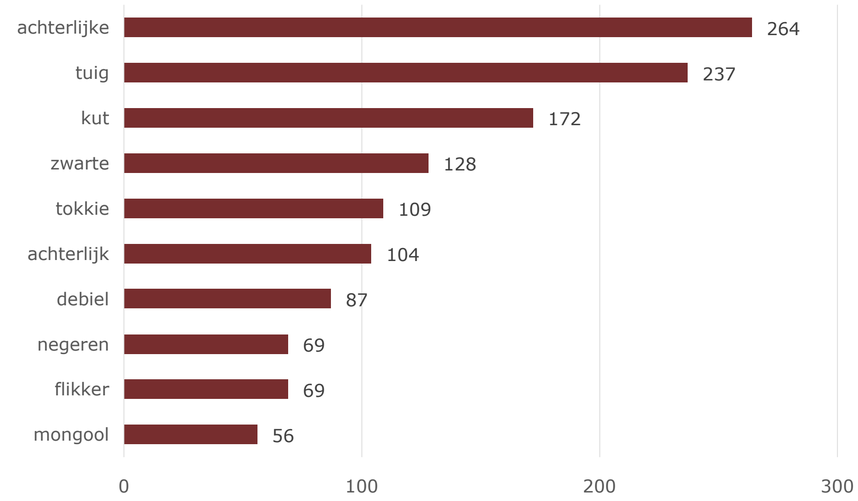
Figuur 16 presenteert dezelfde gegevens, maar nu als percentage van alle posts waarin ten minste één lexiconwoord opduikt. Ook hier is de volgorde enigszins anders dan bij de eerdere zoektermen: discriminerende taal staat bovenaan met 3,2%, gevolgd door beledigingen op 2,3%. Bedreigingen zijn in 1,8% van de berichten aanwezig, terwijl risicovol taalgebruik met slechts 0,3% vrijwel niet voorkomt.

Figuur 16. Percentage tweets met lexicon-termen ‘Wappie’en ‘Tokkie’

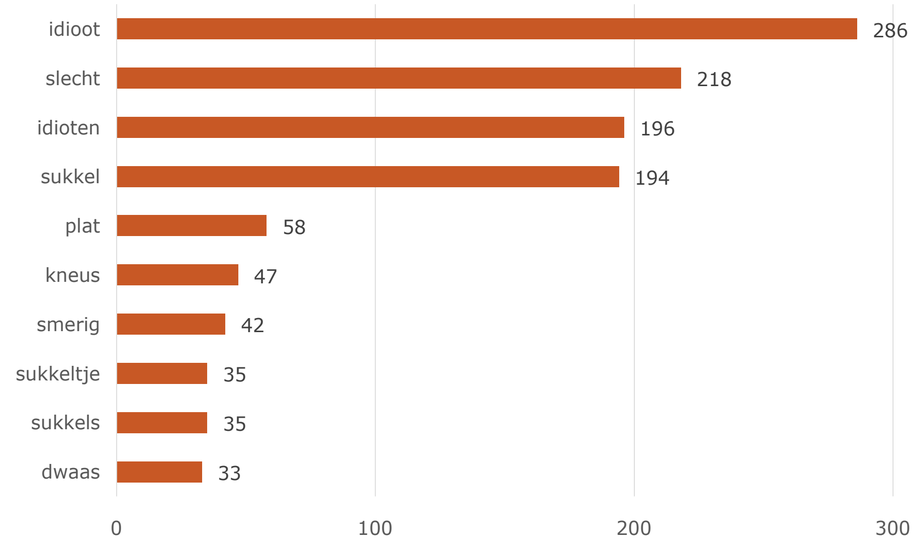


Tenslotte hebben we gekeken naar de tien *meest voorkomende woorde*n in de berichten per lexicon van The Weaponized Word. Alvorens moeten we vooropstellen dat het woord “wappie” of “tokkie” door sommigen als scheldwoord of als beledigende termen worden gezien. In tegenstelling tot de vorige onderwerpen zien we dat het woord ‘tuig’ nu niet het vaakst voorkomt; het woord ‘achterlijke’ komt vaker voor. Verder zien we de beledigende woorden die ook vaker terugkomen bij de andere onderwerpen, met hier en daar een paar verschillen. Opvallend is dat de woorden ‘idioten’, ‘idioot’ en ‘sukkel’ relatief vaak voorkomen in combinatie met de termen ‘wappie’ en ‘tokkie’. Ook de frequente aanwezigheid van de woorden ‘achterlijke’ en ‘achterlijk’ suggereert dat de tweets inspelen op het vooroordeel dat ‘wappies’ en ‘tokkies’ een lagere mate van intelligentie zouden hebben.

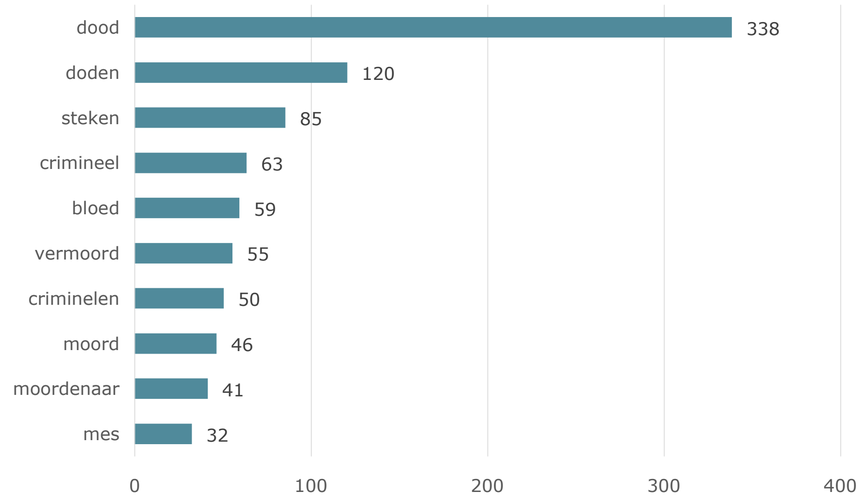
Figuur 17. Top 10 discriminerende woorden 'Wappie' of 'Tokkie'



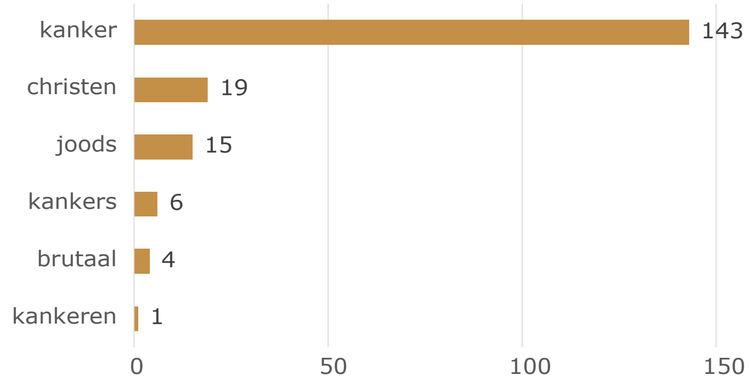
Figuur 18. Top 10 beledigende woorden 'Wappie' of 'Tokkie'



Figuur 19. Top 10 bedreigende woorden 'Wappie' of 'Tokkie'



Figuur 20. Top 6 risicovolle woorden 'Wappie' of 'Tokkie'



B.1.3 Tekstclassificatie

Aanvullend hebben we in dit onderzoek de posts op X geclassificeerd om de aard en omvang van schadelijke of problematische uitingen betrouwbaar in kaart kunnen brengen. Tekstclassificatie is een van de fundamentele toepassingen van NLP (Rayinto et al., 2023). In de kern is tekstclassificatie het proces waarbij we ongestructureerde data toewijzen aan vooraf gedefinieerde categorieën. Deze categorieën kunnen sentimentlabels zijn ("positief", "negatief"), thematische labels ("politiek", "gezondheid") of, zoals in ons onderzoek, vormen van schadelijke inhoud: discriminatie, haatspraak, belediging en dergelijke. Automatische tekstclassificatie is al decennialang een pijler van NLP, mede door de enorme groei aan digitale documenten sinds het begin van deze eeuw (Ikonomakis et al., 2005).

Oorspronkelijk draaide tekstclassificatie vooral om klassieke machine learning-modellen, maar meer recentelijk zijn er deep learning en transformer-architecturen gekomen, zoals BERT en Roberta. BERT ("Bidirectional Encoder Representations from Transformers") leert de context waarin woorden zijn gebruikt door zowel naar de woorden die voor en na een desbetreffend woord komen. Hierdoor kan het model de semantische relatie tussen woorden in kaart te brengen en complexe zinsconstructies te begrijpen. Een ander voordeel is dat BERT-modellen vooraf zijn getraind op enorme tekstverzamelingen, zoals miljoenen woorden van Wikipedia. Hierdoor hebben deze modellen al een diepgaand begrip van taal voordat ze worden ingezet voor specifieke onderzoeken. Onderzoekers kunnen deze voorgetrainde modellen vervolgens verder trainen (fine-tunen) op hun eigen datasets. Dit zorgt ervoor dat BERT zich aanpast aan de specifieke kenmerken van de nieuwe data en betere prestaties levert op de beoogde taak (Sun et al., 2019; Huo & Iwaihara, 2020; Liu et al., 2019; Liu & Lapata, 2019; Rasmy et al., 2020).

B.1.4 Data en methoden

Voor deze analyses hebben we gebruik gemaakt van de data die we eerder in dit onderzoek van X hebben verzameld via Zeeschuimer. Dit omvat alle Nederlandstalige posts die tussen 1 juli 2024 en 31 december 2024 zijn geplaatstwaarin de termen (1) "Wappie" en "Tokkie", (2) "Moslim" of (3) "Fascist" voorkwamen. In totaal resulteerde dit in 236.230 verzamelde posts.

Codeboek voor classificaties

Om sociale mediaberichten systematisch te analyseren, hebben we een indeling gemaakt in verschillende categorieën. Deze classificaties zijn gebaseerd op wetenschappelijke inzichten en juridische definities. Hieronder lichten we per categorie toe wat we eronder verstaan en op basis waarvan berichten zijn gecodeerd.

B.1.4.1. Acceptabel

Deze categorie omvat allerlei positieve en neutrale berichten maar ook berichten die negatief of scherp zijn, maar niet strafbaar of grensoverschrijdend. Denk aan ontevredenheid, boosheid of felle kritiek die niet gericht is op het aanzetten tot haat of geweld. Zulke berichten kunnen onbeleefd of polariserend zijn, maar vallen binnen het normale publieke debat. Zulke berichten vallen onder de noemer 'incivility' of 'rudeness' (i.e., onbeschoft of onfatsoenlijk) zoals beschreven door Rossini (2020), en zijn te onderscheiden van intolerante of haatdragende uitingen.

- Voorbeeld post fascist' "Als men doorvraagt nadat iemand "Racist!", "Fascist!" of "Nazi!" aan het blèren was krijgt men vrijwel nooit een bevredigend antwoord of men zwijgt. Nooit eens een uiteenzetting waarom iemand zou voldoen aan dat beladen woord. Dat komt omdat men geen argumenten heeft en dom is."

- Voorbeeld post moslim "Christendom is om te lachen en er om belachelijk te maken. Voor Islam moeten we respect hebben.. volkomen in lijn met links en progressief Europa en NL."

- Voorbeeld post wappie/tokkie: "Ik zou echt niet willen ruilen met die omhooggevallen tokkie @[naam persoon]."

B.1.4.2. Smaad en laster

Berichten waarin iemand wordt beschuldigd van iets negatiefs met de bedoeling die persoon in een kwaad daglicht te stellen, kunnen vallen onder smaad of laster. Bij smaad (artikel 261 Wetboek van Strafrecht) gaat het om het bewust aantasten van iemands eer of goede naam. Als daarbij ook bewust leugens worden verspreid, spreken we van laster (artikel 262 Sr). In beide gevallen is het niet genoeg dat iets kwetsend is — het gaat om aantasting van reputatie en de intentie om te schaden (Van Dijck et al., 2017). Zowel smaad als laster kunnen ook schriftelijk voorkomen, bijvoorbeeld via sociale media. In beide gevallen moet er sprake zijn van een aantasting van iemands reputatie en is meestal aangifte vereist voor vervolging.

- Voorbeeld post fascist: "@[naam persoon] @[naam persoon] Ze zijn fan van Alexander Dugin een Russische fascist

B.1.4.3. Discriminatie, haatspraak & haatzaaien

In deze categorie vallen berichten die groepen of individuen aanvallen op basis van kenmerken zoals afkomst, religie, gender of seksuele voorkeur. Dit kan via negatieve stereotypering, ontmenselijking, oproepen tot geweld of het gebruik van beledigende taal (zoals scheldwoorden). Discriminatie of haat kan expliciet zijn, zoals met scheldwoorden of dreigementen, maar ook subtieler via codetaal of spot. Volgens Sellars (2016) zijn elementen zoals het uiten van haat, het veroorzaken van leed, het aanzetten tot geweld en het ontbreken van enig maatschappelijk nut kenmerkend voor haatspraak. Ook content die haatideologieën promoot, zoals witte suprematie of antisemitisme, valt hieronder.

- Voorbeeld post fascist: "@[naam persoon] Jij vindt het prima dat Frankrijk nog verder wordt overspoeld met ongeciviliseerde Afrikaanse apen? Ik hoop oprecht dat een van je kinderen nog eens van hun fiets gesleurd worden. Linkse fascist"
- Voorbeeld post moslim: "@[Naam Persoon] Je bent overduidelijk een terroristen moslim eentje die we liever kwijt dan rijk zijn. Onaangepast moslim racistisch individu"
- Voorbeeld post wappie/tokkie: "Zo doe je alleen maar al weet je dat je hele tokkie familie je komt helpen als je slaags krijgt. Was vroeger al zo alleen nu zijn ze getint. Wat een armoede"

B.1.4.4. Bedreiging en intimidatie

Berichten die expliciet of impliciet dreigen met schade, geweld of wraak worden als bedreigend geclassificeerd. Het maakt hierbij niet uit of de dreiging verbaal of schriftelijk is, of via sociale media wordt verspreid. Volgens artikel 285 Sr is bedreiging strafbaar zodra er sprake is van het uiten van de intentie om een ander ernstig leed toe te brengen. Dit kan variëren van directe doodsbedreigingen tot subtielere vormen van intimidatie, zoals chantage of doxing.

- Voorbeeld post fascist: "[Naam persoon] Volgens mij ben jij de grootste gevaarlijkste extreem linkse fascist die geneutraliseerd moeten worden van X, sodemieter op naar mastodont."
- Voorbeeld post moslim: "@[Naam persoon] "als jij het gore lef zou hebben om mij aan te vallen en mijn shit te slopen en ik weet waar jij bent de komende dagen, kom ik óók het licht uit je ogen meppen. Het maakt mij niets uit of je Katholiek, Moslim, Joods of fckng Thanos opzoek naar de Infinity Stones bent."
- Voorbeeld post wappie/tokkie: "@[Naam persoon] Niemand vraagt je om lyrisch te worden over homoseksualiteit e.d., er wordt je uitgelegd dat dit normaal en natuurlijk verschijnsel is wat je te accepteren hebt, ongeacht wat je er van vindt. Nee je bent nogal wat verder naar radicaal rechts afgezakt he? Full wappie modus."

B.1.4.5. Belediging

Berichten vallen onder deze categorie wanneer ze kwetsende of haatdragende uitlatingen doen over een persoon of groep op basis van bijvoorbeeld afkomst, geloof, seksuele oriëntatie of handicap. Hiermee lijkt deze definitie echter veel op die van discriminatie en haatspraak. Dit sluit aan bij artikel 137c Sr over groepsbelediging. Belediging is subjectief en contextgevoelig, maar in deze studie richten we ons op uitingen die opzettelijk beledigend zijn, maar die niet gebaseerd zijn op de achtergrondkenmerken van een persoon of groep, zoals bijvoorbeeld afkomst, geloof, seksuele oriëntatie of handicap worden gedaan.

- Voorbeeld post fascist: "@[Naam persoon] Je bent niet goed bij je pan achterlijke fascist"
- Voorbeeld post moslim:
- Voorbeeld post wappie/tokkie: "@[Naam persoon] ben je nu nog steeds aan het drammen, tokkie? Stop met je te wentelen in je grote gelijk wanneer het overduidelijk is dat je gewoon een domme idioot uit de bodem van de maatschappij bent. Een fantast die zichzelf en z'n inzicht geweldig overschat."

B.1.4.6. Misinformatie en desinformatie

Berichten in deze categorie bevatten feitelijk onjuiste informatie. Als dat per ongeluk gebeurt, spreken we van misinformatie. Als het bewust gebeurt om anderen te misleiden of leed te veroorzaken, noemen we het desinformatie. Denk aan complottheorieën of het verspreiden van nepnieuws over verkiezingen (Baines & Elliott 2020; Perez-Escolar et al 2022; Aimeur et al 2023).

- Voorbeeld posts fascist: "@[Naam persoon] Ik weet wat je bedoelt, ik ben een(zie gift) van de één en dan van de ander; een racist en fascist. Alleen maar omdat ik geen globalisme wil en linksdenkende mensen nog steeds niet weten wat dat inhoud. De oorlogen zijn 'nodig' om bepaalde bloedlijnen te elimineren"
- Voorbeeld posts moslim: "De foto toont glashelder waarom NL totaal verloederd is tot het enige en echte, sterk Joden-hatend moslimparadijs in het universum, dankzij de 5e colonne v/d Islam."
- Voorbeeld posts wappie/tokkie: "Oorlog is gewoon beleid voor hen en daar komt bij dat zij vinden dat er teveel mensen op de planeet zijn. Jij weet vast ook wie ongeveer aan de touwtjes trekken. Misschien een goed onderwerp voor je."

B.1.4.7. Kleine ordeverstoringen en opruiing

Berichten die aanzetten tot geweld of strafbare feiten vallen onder opruiing, zoals omschreven in artikel 131 Sr. Het gaat hier om berichten die oproepen tot onrust, rellen of geweld. Het gaat dan om het aanzetten tot strafbare feiten of verstoring van de openbare orde. Ook wanneer deze oproepen via internet worden gedaan (bijv. via berichten, memes of video's), kan het strafbaar zijn.

- Voorbeeld posts moslim: "Verbrand een moslim met een koran in zijn hand"

B.1.4.8. Counterspeech

Dit zijn berichten die reageren op haatdragende of schadelijke taal met het doel deze te ondermijnen. Counterspeech kan verschillende vormen aannemen: empathie tonen, feiten aandragen, iemand terechtwijzen, of de negatieve uiting ombuigen. Het wordt in de literatuur gezien als een belangrijk niet-repressief middel om online haatspraak tegen te gaan (Dangerous Speech Project Home).

- Voorbeeld post fascist: "@[Naam persoon] Dit is al veel beter zonder dat pauperige "smerige kk fascist" in uw andere reactie. Hou dit vol! U kunt het! □□□□"
- Voorbeeld post moslim: "@[Naam persoon] Hoe weetje eigenlijk dat hij moslim is? Jouw opmerking zegt eigenlijk alles over het niveau dat je bezit, types als jij zijn een groot probleem op deze wereld. Mensen tegen elkaar opzetten zorgt voor chaos."
- Voorbeeld post wappie/tokkie: "@[Naam persoon] Waarom ben je een "wappie" als je je zorgen maakt over deze ontwikkelingen?"

B.1.4.9. Publieke Figuren

In sommige gevallen richten berichten zich op publieke personen zoals politici, artiesten of opiniemakers. Deze categorie is geen inhoudelijke classificatie op zich, maar wordt genoteerd om onderscheid te maken tussen uitingen over individuen in privé-context en personen met een publieke rol. Juridisch gezien genieten publieke figuren vaak minder bescherming tegen kritiek dan privépersonen (EHRM, Lingens v. Austria, 1986).

- Voorbeeld post fascist: "@[Naam politici] "ie 2 zo genaamde politici helpen Nederland naar de kloten. Een idee een echte fascist en de ander een echte conservatieve aartsliberaal met fascistische trekjes."

- Voorbeeld post moslim: "@[Naam politicus] "Dan moet je het laatste sprankje hoop voor vrijheid en gelijkheid bij het nekvel pakken! In Israel wonen 2 miljoen moslims met gelijke rechten en plichten, christenen, homos en iedereen leeft vredig met elkaar. Ik zou zeggen, ga lekker bekken met hamas!!!"

- Voorbeeld post wappie/tokkie: "@[Naam journalist] Iedereen met een andere mening is een wappie geworden volgens jou. Volgens mij begin jij zelf gewoon een wappie te worden!"

B.1.4.10. Uit te filteren berichten

Sommige berichten zijn niet relevant voor de analyse en worden uitgesloten. Denk hierbij aan berichten zonder inhoud (zoals alleen een emoji), Nederlandstalige berichten die betrekking hebben op de Vlaamse context, of berichten die te onduidelijk zijn om te interpreteren.

1.5 Werken naar een classificatiemodel

Om een classificatiemodel te bouwen, beginnen we met een gelabelde dataset: een steekproef van posts waar wij elke post beoordelen volgens een codeboek. Die labels vormen de basis waarvan het model leert. Daarom hebben we een random sample getrokken van 500 posts voor elke zoekterm. Deze samples gebruikten we om vijfhonderd Nederlandstalige berichten handmatig te labelen. Drie onderzoekers labelden elk bericht aan de hand van het codeboek (zie verderop in de bijlage). Vervolgens splitsten we die gelabelde data in een trainingsset (80 %) en een testset (20 %). Hiermee kunnen we kijken hoe goed het model presteert op data die het nog niet eerder heeft gezien. In de training leert het model patronen in woordgebruik, zinsbouw en context associëren met de labels. Daarna toetsen we op de testset hoe vaak het model correct voorspelt (nauwkeurigheid), hoe vaak het relevante uitingen herkent (recall) en hoe vaak de voorspellingen zuiver zijn (precisie). Die metrics geven samen een evenwichtig beeld van de kwaliteiten en valkuilen van het model (Wilkinson, Wenger & Shugarman, 2007).

Onderzoek toont telkens aan dat BERT-modellen hun traditionele tegenhangers verslaan op metrics als F1-score of accuracy, vooral bij korte, context-afhankelijke teksten of in meertalige omgevingen (Yu et al., 2019; Wu & Xu, 2024; Nishigaki et al., 2023). In dit onderzoek vergelijken we daarom drie transformer-modellen: de Nederlandse modellen RobBERT v2 (Delobelle, Winters & Berendt, 2020) en BERTje (De Vries et al., 2019), een meertalig XLM-T model (Barbieri et al. (2021). De uiteindelijke keuze voor transformer-modellen is gebaseerd op drie overwegingen. Ten eerste vraagt sociale media-taal om diepe contextuele begripsvormen die alleen transformer-modellen bieden. Ten tweede zijn dit soort in staat om al van kleine hoeveelheden data te leren, waardoor de modellen ook effectief van vijfhonderd labels leert (Taneja & Vashishtha, 2022; Nishigaki et al., 2023). Ten derde laten eerdere studies op Twitter-data met tienduizenden voorbeelden zien dat transformer-architecturen consistente winst boeken, zowel in binaire haatspraakdetectie als bij classificaties met meerdere klassen (Li et al., 2020; González-Carvajal & Garrido-Merchán, 2020).

Na fine-tunen van meerdere modellen selecteren we het model met de beste balans tussen accuracy, precisie, recall en F1 score (zie kader). Dat model passen we vervolgens toe op de volledige dataset met posts. We hebben hierbij de zowel gekeken of er een model is dat voor de 1500 handmatig gelabelde posts samen werkt, of dat sommige modellen beter werken voor een bepaalde zoekterm dan voor een andere.

Hoe beoordelen we of het model goed werkt?

Om te controleren of een model goede voorspellingen doet, kijken we naar vier veelgebruikte cijfers: accuracy, precision, recall en de F1-score. (1) Accuracy laat zien hoeveel van de voorspellingen helemaal correct zijn. Maar dat getal zegt niet alles. Stel dat één categorie veel vaker voorkomt dan de andere, dan kan het model "goed scoren" door alleen die ene categorie te kiezen. In andere woorden, het is alsof je altijd A invult op een meerkeuzetoets. Als toevallig veel antwoorden ook echt A zijn, dan scoor je best goed, maar dit komt meer door toeval dan door kennis. Daarom kijken we ook naar precision (hoe vaak het model gelijk had als het een bepaalde categorie voorspelde) en recall (hoe goed het model alle voorbeelden van die categorie weet te vinden). De F1-score is een combinatie van precision en recall en geeft een gebalanceerd beeld van hoe goed het model per categorie presteert. We kijken vooral te letten op de F1-score per categorie, omdat dat het beste laat zien waar het model goed of slecht in is. Accuracy alleen kan misleidend zijn.

1.6 Resultaten

Allereerst hebben we de handmatig gelabelde datasets opgeschoond. Dit omvatte het verwijderen van non-alfanumerieke symbolen zoals emoticons, gebruikersnamen en URLs, en een laatste controle of er echt alleen maar Nederlandse berichten in de data zitten. Niet-Nederlandse berichten kunnen er namelijk voor zorgen dat er meer ruis in de data ontstaat waardoor de andere posts niet goed geclassificeerd kunnen worden. Na het opschonen van deze datasets hadden we 469 bruikbare, gelabelde posts voor de zoekterm "moslim", 472 posts voor "fascist" en 455 posts voor "wappie" en "tokkie".

Voordat we begonnen aan het testen van de BERT modellen, hebben we gekeken naar hoe vaak de labels voor kwamen in de handmatig, gelabelde data (zie Tabel 1). Dit overzicht is mogelijk al een eerste schets van de aard en omvang van online geweld. Zo zagen we dat de meeste posts "acceptabel" waren. Dit zijn berichten die niet grensoverschrijdend zijn. Zoals uitgelegd in het codeboek, kunnen deze berichten wel ontevredenheid, boosheid of felle kritiek bevatten, maar ze zijn niet te identificeren als een mogelijk strafbaar feit. De strafbare vormen die relatief vaker voorkomen zijn belediging en discriminatie en haatspraak. Verder laat de tabel ook publieke figuren en uit te filteren berichten zien. Dit zijn berichten die buiten de scope van dit onderzoek liggen.

Tabel B.3. Frequentie van de labels

Label	Fascist	Moslim	Wappie en Tokkie	Totaal
Acceptabel	172	255	174	601
Smaad en laster	1	0	0	1
Discriminatie, haatspraak & haatzaaien	9	86	3	98
Bedreiging en intimidatie	2	1	1	4
Belediging	57	7	98	162
Misinformatie en desinformatie	5	13	5	23
Kleine ordeverstoringen en opruiing	0	1	0	1
Counterspeech	20	60	18	98
Publieke Figuren	153	13	33	199
Uit te filteren berichten	53	33	123	209

Uit dit overzicht blijkt dat we te maken hebben met een ongelijk verdeelde, gelabelde data. Dat betekent dat sommige categorieën veel minder voorbeelden bevatten dan andere. Een model dat tijdens training vooral de meerderheidscategorie ziet, loopt het risico de zeldzamere labels minder goed te herkennen. Tekstclassificatie met meerdere klassen verloopt moeizamer wanneer klassen ongelijk verdeeld zijn (Garcia et al., 2018; Zhao et al., 2024).

Omdat we onze steekproef van 500 berichten blind trokken, konden we vooraf niet garanderen dat elke categorie voldoende exemplaren zou bevatten. Methoden als *undersampling* van de meerderheid of *oversampling* van de minderheidsklassen zijn weliswaar gangbaar (Kim & Hwang, 2022; Chabalala et al., 2023), maar door het willekeurig trekken van de steekproef was dit niet mogelijk. In plaats daarvan kozen we voor een benadering die robuust is tegen ongebalanceerde klassen. Hierbij vergelijken we meerdere BERT-modellen en focussen we ons op F1-score als leidende score om het model te beoordelen.

De kern van onze aanpak is het vergelijken van verschillende modellen. Bij machine learning is het toepassen van meerdere prestatiemetrics essentieel, zeker bij ongebalanceerde data. Accuracy kan misleiden wanneer de grootste klasse de overhand heeft; een model dat altijd het meerderheidslabel voorspelt, haalt dan een hoge accuracy maar mist alle zeldzame categorieën. Precision en recall geven beter zicht op respectievelijk de zuiverheid van voorspellingen en het vermogen om alle relevante voorbeelden te vinden. Toch biedt een soort gemiddelde van precision en recall, de F1-score, de meest gebalanceerde weging van beide aspecten in ons geval (Riyanto et al., 2023).

In onze analyses trainen en evalueren we daarom drie individuele modellen (RobBERT v2, BERTje en XLM-RoBERTa) op exact dezelfde 80/20-split van de sample. Per model vergelijken we de F1-scores per zoekterm én voor wanneer we de gelabelde tweets voor de zoektermen samen nemen. Naast deze individuele modellen bouwen we een ensemble dat de voorspellingen van alle individuele modellen combineert. Hierbij leunen we op het fenomeen *majority voting*, wat inhoudt dat het label wordt gekozen die door de meerderheid van de BERT-modellen wordt gekozen. In andere woorden, als twee van de drie modellen aangeven een bepaalde post als discriminatie en haatspraak labelen, dan krijgt deze post dit label. Op deze manier kunnen we een robuustere labelling van de posts krijgen omdat de modellen elkaar aanvullen.

Onze keuze voor een ensemble vloeit voort uit een aantal overwegingen. Ten eerste verlagen ensembles de variantie van voorspellingen, wat leidt tot betrouwbaardere F1-scores. Ten tweede biedt het ensemble een vangnet voor categorieën waarin individuele modellen onvoldoende scoorden. Dit sluit aan bij de bevindingen dat gecombineerde methoden vaak beter omgaan met ongelijkheid tussen meerdere klassen (Sun et al., 2015; Khan et al., 2023)

1.7 Welk model werkt het best

In de onderstaande tabel (Tabel 2) presenteren we de prestaties van de vier modellen: Ensemble, BERTje, XLMR Stage 2 en RobBERT. We kijken hier naar hun score voor de drie zoektermen (moslim, fascist, wappie en tokkie) en op de volledige dataset (ALL). We rangschikten per model de scopes op basis van de F1score, omdat deze metric de beste balans biedt tussen precisie en recall bij een ongebalanceerde steekproef.

We hebben voor elke zoekterm drie trainingsrondes uitgevoerd op alle beschikbare gelabelde tweets binnen die zoekterm, én drie ronden op de volledige dataset (“ALL”). Voor elke zoekterm keken we welke trainingsronde de hoogste F1score gaf op de bijbehorende testset. Bij het classificeren van de ongelabelde posts laden we vervolgens voor elke zoekterm die specifieke trainingsronde met de hoogste F1-score om de nieuwe data te labelen. Hierdoor gebruiken we telkens de best presterende versie van het model per zoekterm.

Het ensemblemodel voor de specifieke zoektermen levert in alle drie de steekproeven de hoogste F1scores en is daarmee duidelijk superieur aan de individuele modellen. Het Ensemble-model behaalt met een F1 van 0,79 op de moslimsubset de hoogste score. Dit is ‘acceptabel’ in de zin dat we met bijna 80 % van de relevante berichten correcte labelling verwachten. Voor de steekproef voor de zoektermen “wappie” en “tokkie” scoort het ensemble oké met een F1 van 0,67, ondersteund door een accuracy van 0,71 en een goede precision van 0,82, wat betekent dat het merendeel van de inschattingen tijdens het trainen van het model goed lukte. De steekproef voor de term “fascist” zien we een F1 van 0,53. Dit ligt nog steeds boven de individuele BERT en RoBERTamodellen , maar wijst op meer moeilijkheden bij zeldzamere of contextueel subtiele uitingen. De lagere F1 in deze steekproeven duidt erop dat het ensemble meer moeite heeft om zeldzamere of contextueel complexe uitingen te herkennen. De andere modellen blijven allemaal ruim onder de ensemblescores en vertonen op alle scopes F1’s tussen 0,45 en 0,72.

Hoewel de ensemblebenadering consequent de beste F1scores oplevert, liggen de absolute waarden voor precisie en recall nog beneden het gewenste niveau. Een precisie onder 0,60 betekent dat meer dan de helft van de voorspelde labels onterecht is (false positives), en een recall onder 0,30 betekent dat een groot deel van de relevante berichten niet herkend wordt (false negatives). Dit betekent dat de resultaten die we in het volgende stuk zullen presenteren moeten worden gezien als een grove inschatting van de aard en omvang van online geweld, maar niet als een (volledig) betrouwbare representatie van de online leefwereld. In de reflectie gaan we dieper in op deze beperkingen en bespreken we aanbevelingen om hier in de toekomst beter mee om te gaan.

Tabel B.4. Prestatie scores van de verschillende modellen

Modeltype	Zoekterm	Accuracy	Precision	Recall	F1-score
Ensemble	Moslim	0.6559	0.6559	0.1429	0.7922
	Wappie	0.7111	0.8246	0.3671	0.6661
	Fascist	0.5806	0.6080	0.2571	0.5326
	All	0.4964	0.5317	0.2480	0.4374
BERTje	Moslim	0.5376	0.5376	0.1429	0.6993
	Wappie	0.5111	0.5598	0.2833	0.5374
	All	0.4784	0.3753	0.2994	0.4233
	Fascist	0.4839	0.3769	0.2470	0.4251
RobBERT	Moslim	0.5591	0.5591	0.1667	0.7172
	Fascist	0.4624	0.4653	0.2229	0.5455
	Wappie	0.5111	0.5851	0.1993	0.4884
	All	0.4820	0.5008	0.2600	0.4430
XLM-R	Moslim	0.5591	0.5591	0.1667	0.7172
	Fascist	0.5484	0.5619	0.1836	0.6267
	All	0.5036	0.3678	0.2480	0.5548
	Wappie	0.5444	0.5454	0.2451	0.5030

1.8 Aard en omvang van online geweld op X

Allereerst hebben we de ongelabelde datasets opgeschoond. Dit omvatte het verwijderen van non-alfanumerieke symbolen zoals emoticons, gebruikersnamen en URLs, en een laatste controle of er echt alleen maar Nederlandse berichten in de data zitten. Niet-Nederlandse berichten kunnen er namelijk voor zorgen dat er meer ruis in de data ontstaat waardoor de andere posts niet goed geclassificeerd kunnen worden. Na het opschonen van deze datasets hadden we 113.796 bruikbare, ongelabelde posts voor de zoekterm “moslim”, 53.341 posts voor “fascist” en 51.498 posts voor “wappie” en “tokkie”.

Voor het classificeren van de ongelabelde posts hebben we gebruikgemaakt van een zogenaamd ensemblemodel, dat voorspellingen combineert van drie afzonderlijke modellen: RobBERT, BERTje en XLM-RoBERTa. Eerder bleek dit ensemblemodel het beste te presteren op onze testdata. Daarom is dit model toegepast op de volledige verzameling ongeannoteerde berichten voor elk van de drie zoektermen: “moslim,” “fascist,” en “wappie/tokkie.”

In Tabel 3 en Figuur 1 beschrijven we de resultaten van de tekstclassificatie. Kijkend naar de overkoepelende resultaten van alle drie de zoektermen (“Fascist”, “Moslim” en “Wappie en tokkie”), dan zien we dat de overgrote meerderheid van de berichten is geclassificeerd als Acceptabel. Overige labels, zoals Belediging, Publieke Figuren en Uit te filteren berichten, komen relatief minder vaak voor. Sommige categorieën (bijvoorbeeld Bedreiging en intimidatie, Discriminatie en Counterspeech) zijn zelfs volledig afwezig.

Fascist

Van de geclassificeerde berichten voor “fascist” vallen er 31 218 onder Acceptabel, 291 onder Belediging, 21 812 onder Publieke Figuren en 20 onder Uit te filteren berichten. De overige categorieën (Bedreiging en intimidatie; Counterspeech; Discriminatie, haatspraak & haatzaaien; Kleine ordeverstoringen en opruiing; Misinformatie en desinformatie; Smaad en laster) bevatten geen posts.

Moslim

Voor de zoekterm “Moslim” zijn 113 796 berichten als Acceptabel geclassificeerd. Alle overige labels komen niet voor.

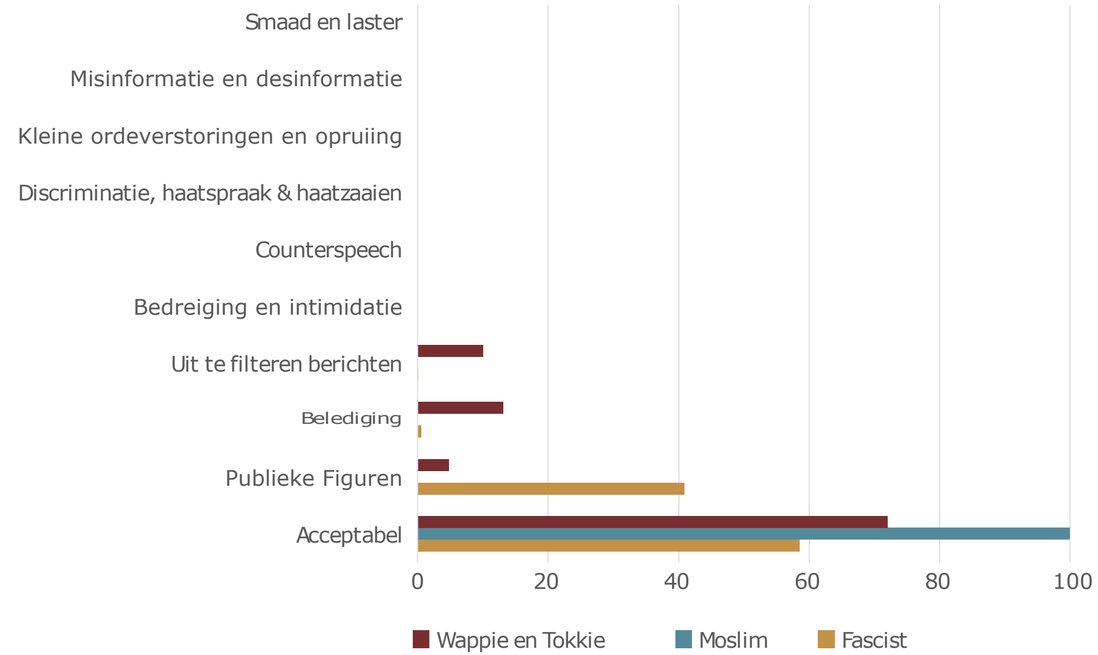
Wappie en tokkie

Voor de zoekterm “Wappie en tokkie”–subset zijn 37.067 berichten als Acceptabel gelabeld, 6.761 als Belediging, .2477 als Publieke Figuren en 5.193 als Uit te filteren berichten. Ook hier zijn er geen posts voor de overige categorieën.

Tabel B.5.

Label	Fascist	Moslim	Wappie en tokkie
Acceptabel	31218	113796	37067
Bedreiging en intimidatie	0	0	0
Belediging	291	0	6761
Counterspeech	0	0	0
Discriminatie, haatspraak & haatzaaien	0	0	0
Kleine ordeverstoringen en opruiing	0	0	0
Misinformatie en desinformatie	0	0	0
Publieke Figuren	21812	0	2477
Smaad en laster	0	0	0
Uit te filteren berichten	20	0	5193

Figuur 21. Labelverdeling per domein (in %)



Noot: Alle percentages zijn berekend op basis van het totaal aantal posts voor een zoekterm gedeeld door het totaal aantal posts gevonden binnen een categorie voor deze zoekterm.

Ongelijke verdeling in training data

Een belangrijke beperking van deze verkenning is dat we te maken hadden met een ondervertegenwoordiging van bepaalde klassen in de handmatige gelabelde steekproef die werd gebruikt voor het trainen van de modellen. Dit heeft invloed op de betrouwbaarheid van de automatische classificatie: het model lijkt beter in het herkennen van dominante categorieën, zoals acceptabele posts, terwijl zeldzamere, subtielere, of overlappende vormen van online geweld mogelijk minder goed zijn opgemerkt. Gezien deze beperking is het belangrijk om te benadrukken dat dit onderzoek slechts een eerste inschatting geeft van de omvang van online geweld.

Bovendien, doordat sommige categorieën in de gelabelde steekproef zeldzamer waren dan andere (bijv. smaad en laster), ontstond er mogelijk bij de training een bias in de richting van de meerderheid. We hebben dit deels ondervangen door meerdere BERT-modellen met elkaar te vergelijken en samen te laten werken in een ensemble model. Door te kijken naar de meest voorkomende voorspellingen van RobBERT v2, BERTje en XLM-RoBERTa voor elke post, verminderden we de variantie in de voorspellingen. Hiermee creëerden we een vangnet voor de voorspellingen waarin de individuele modellen niet goed presteerden. Hoewel deze aanpak niet geheel de ongelijke verdeling kan opheffen, levert het ensemble een robuustere labelling op en maakt het de resultaten beter bestand tegen de dominantie van de meerderheidscategorieën.

Vervolgonderzoek zou deze uitkomsten kunnen gebruiken om in de toekomst accurater te zijn op de minder voorkomende vormen van online geweld. Idealiter wordt een grotere hoeveelheid data gelabeld, zodat minderheidsklassen voldoende voorbeelden bevatten om het model tijdens training expliciet mee te nemen.

Daarnaast kunnen technieken als stratified sampling, gerichte oversampling of semi supervised learning helpen om de vertekening te verkleinen zonder blind te vertrouwen op random sampling. Door deze stappen te combineren met onze ensemble aanpak, kunnen we in vervolgonderzoek betere generalisatie en hogere betrouwbaarheid bereiken.

Een andere optie voor vervolgonderzoek omvat het uitlichten van één specifieke categorie per zoekterm en daarop te focussen. Als er veel categorieën zijn om te classificeren, dan kan dit leiden tot aanzienlijke ruis voor de taalmodellen. De kans om hiermee zeldzamere vormen van online geweld te detecteren neemt hiermee af. In het geval van de zoekterm “moslim” zou dat betekenen dat het model uitsluitend traint en evalueert op discriminatie en haatspraak. Door deze focus vermindert de overlap met andere categorieën en kan de modelperformance op juist die vormen van gewelddadige of grensoverschrijdende taal worden geoptimaliseerd.

Hierdoor ontstaat niet alleen een nauwkeurigere detectie van haatvolle uitingen, maar neemt ook de interpretatie van de resultaten toe.

Overlap en grijze gebieden

Een belangrijk inzicht is dat veel categorieën in de praktijk overlappen *(zie hier ook de eerdere rapportages:* Onderzoek naar strafbare vormen van online geweld - Verwey-Jonker Instituut*).* Belediging, discriminatie en haatspraak zijn niet altijd strikt te scheiden, en ook counterspeech kan qua toon of inhoud lijken op de uiting waartegen het zich keert. Deze overlap maakt het classificeren van berichten niet alleen technisch uitdagend, maar roept ook vragen op over hoe we online gedrag duiden en adresseren. Wat als een vorm van verdediging is bedoeld, kan in de praktijk alsnog bijdragen aan verhitte en polariserende discussies.

Dit grijze gebied en de overlap tussen categorieën viel ons ook op tijdens het handmatig labelen van de steekproef. Zo kunnen posts waarin er sprake is van discriminatie of haatspraak tegelijkertijd beledigende of intimiderende uitingen bevatten, en vormen van misinformatie kunnen bijvoorbeeld gepaard gaan met agressieve taal. In onze eerdere publicaties wezen wij al op het gebrek aan breed gedragen definities voor vormen van online geweld. We baseerden ons in dit onderzoek vooral op strafrechtelijke kaders en internationale wetenschappelijke literatuur, maar de prestatie van de tekstmodellen suggereert dat deze kaders niet altijd aansluiten op de dynamische en contextafhankelijke aard van socialmedia berichten in Nederland.

Ook dit kan een reden zijn waarom bepaalde prestatiemetrics niet goed uitvielen. Mogelijk hadden BERTmodellen moeite om deze samengestelde uitingen consequent te scheiden, omdat zij per definitie één label per post toewijzen. Ook kan het zo zijn dat het de subtiele varianten van haatspraak of intimidatie moeilijk kon herkennen omdat bepaalde vormen van online geweld erg genuanceerd zijn. Dit gegeven kan leiden tot ruis in de voorspellingen en F1scores voor de zoektermen. In vervolgonderzoek kan er in meer detail naar online uitingen van online geweld gekeken worden zodat de classificatie van de verschillende vormen nog scherper kan worden geformuleerd. Dit kan bijdragen aan een breder begrip van de vormen van online geweld, en een nauwkeurigere afbakening van deze vormen.

Een schaal van online gedrag: van acceptabel naar strafbaar

De schaal die we hebben gehanteerd – van acceptabel gedrag tot aan strafbare uitingen – blijkt waardevol om het fenomeen online geweld beter te begrijpen. Het helpt om niet alleen te kijken naar de uitersten, maar juist ook naar het brede tussengebied. Hier bevindt zich het zogenaamde ‘onbeschoft’ gedrag: uitingen die niet strafbaar zijn, maar wel bijdragen aan een verharding van het publieke debat—waardoor bepaalde meer kwetsbare groepen zich juist meer gaan terugtrekken uit de online wereld en het publieke debat dat zich online afspeelt. Deze categorie verdient meer aandacht in beleid en preventie. Door tijdiger te interveniëren in dit segment, kan mogelijk worden voorkomen dat online interacties escaleren naar vormen van strafbaar geweld en kan er voor zorgen dat iedereen, met name kwetsbare groepen zich veiliger voelen online.

De balans tussen vrijheid en bescherming

Ook deze eerste verkennende schaal onderstreept nogmaals de fundamentele uitdaging van de balans tussen vrijheid van meningsuiting en bescherming tegen online schade. Wanneer wordt scherpe kritiek of satire grensoverschrijdend? En hoe voorkomen we dat de norm vervaagt tussen polariserende maar legitieme opinies en haatdragende communicatie? In dit onderzoek hebben we geprobeerd de grenzen van online geweld scherp te definiëren aan de hand van juridische en maatschappelijke kaders, maar de interpretatie hiervan blijft deels contextueel en subjectief. Daarom is aanvullend onderzoek nodig, met name op andere sociale mediaplatforms. X wordt door slechts een selecte groep Nederlandse burgers gebruikt en heeft een eigen ecosysteem; voor een volledig beeld is het belangrijk om deze analyse aan te vullen met data van andere platforms zoals TikTok, Facebook en Instagram.

Implicaties voor beleid en praktijk

Om online geweld beter te herkennen én te voorkomen, zijn enkele structurele stappen nodig. Allereerst is er behoefte aan gedeelde en breed gedragen definities van wat onder online geweld valt. Zonder gezamenlijke taal blijft het lastig om signalen te herkennen, registreren of erop te handelen. Daarnaast laat ons onderzoek zien dat tekst slechts één deel van het verhaal vertelt: beeldmateriaal en memes spelen een minstens zo belangrijke rol in online communicatie en moeten in toekomstig onderzoek worden meegenomen. Tot slot verdient het gebruik van bepaalde woorden en termen (zoals ‘wappie’ of ‘fascist’) aandacht in termen van framing: de maatschappelijke lading ervan kan per context sterk verschillen, en dat maakt de interpretatie des te belangrijker.

Deze verkenning laat zien dat preventie niet alleen begint bij het strafbare, maar juist ook daarvóór – bij het normaliseren van online omgangsvormen en het ondersteunen van counterspeech. De glijdende schaal die we in dit onderzoek in kaart brachten, helpt om te bepalen waarlangs preventieve acties mogelijk zijn. Daarmee hopen we bij te dragen aan een meer samenhangend en proactief beleid rondom online geweld tussen burgers.

Bijlage 2 Respondenten

Wetenschappelijke- en praktijkexperts

- Politie
- Fonds Slachtofferhulp
- SafetyNed
- Avans Centre of Expertise Veiligheid en Veerkracht
- Saxion
- Nederlands Studiecentrum Criminaliteit en rechtshandhaving (NSCR)
- Universiteit Leiden
- NHL Stenden
- Haagse Hogeschool
- TikTok
- DPG Media
- Nu.nl

Deelnemers focusgroepen

- Politie (expert Osint/projectleiders)
- Jeugdagenten
- Digitale wijkagent
- Vereniging Nederlandse Gemeenten (VNG)
- Bureau Halt
- Centrum voor Criminaliteitspreventie en Veiligheid (CCV)
- Gemeenten (Adviseurs Veiligheid en Digitale Veiligheid)
- Hogeschool Saxion (Online Weerbaarheid)

COLOFON

Opdrachtgever:	Ministerie van Justitie & Veiligheid
Auteurs:	Dr. M.W.A. Odekerken Dr. E.J. de Weger M. Yenga, MSc D.J.H. Woertink, MSc
Omslag	AdobeStock of andere, nr, naam
Uitgave	Verwey-Jonker Instituut Giessenplein 59-C 3522 KE Utrecht T (030) 230 07 99 E secr@verwey-jonker.nl I www.verwey-jonker.nl

De publicatie kan gedownload worden via onze website:
http://www.verwey-jonker.nl

ISBN: 978-94-6409-397-1