Uitdagende spraak Contentmoderatie in tijden van de Grote *De-platforming*

Juncal Montero Regules*

Na de Capitoolrellen op 6 januari 2021 schorsten Twitter en Facebook het profiel van de toenmalige Amerikaanse president Donald Trump. Die Grote De-platforming (*Great De-Platforming*) ontketende felle debatten over vrije meningsuiting en censuur, en over de macht van online platformen en de beweegredenen achter hun beslissingen. Deze bijdrage onderzoekt hoe de contentmoderatie-kaders van sociale-mediaplatforms de vrijheid van meningsuiting in de online wereld reguleren. Eerst wordt het concept van contentmoderatie toegelicht, met bijzondere aandacht voor de regels en processen ervan. Vervolgens worden de voornaamste uitdagingen van contentmoderatie geïdentificeerd. Hierbij wordt in het bijzonder nagegaan of de *Facebook Oversight Board*, een quasi-judicieel orgaan opgericht door Facebook zelf, een antwoord kan bieden op deze uitdagingen.

Online platformen, online meningsuiting en contentmoderatie: what's in a name?

Dertig jaar na de oprichting van het wereldwijde web worden dagelijks miljoenen online interacties gereguleerd door een overvloed aan private en publieke regels. Tegenwoordig zijn alle vormen van internetinhoud in toenemende mate toegankelijk via private sociale-mediaplatformen.[1] Grote technologiebedrijven hebben hierdoor een enorme macht verworven. Vandaag organiseert een oligopolie van particuliere bedrijven de transnationale online-informatie met behulp van geautomatiseerde technologieën. Hierbij leggen zij zich een functionele soevereiniteit[2] én eigen regels op.

Online platformen worden gedefinieerd als online sites en dito diensten die gedeelde inhoud of sociale interacties van gebruikers hosten, organiseren en voor hen laten circuleren. Kenmerkend is dat zij het grootste deel van die inhoud niet zelf hebben geproduceerd of laten produceren. Zij bieden enkel, of in de eerste plaats, een infrastructuur aan voor de verwerking van gegevens voor een reeks verschillende doeleinden, waaronder het genereren van winst.[3] Hierbij spelen online platformen een fundamentele rol bij het vaststellen van de normen voor vrije meningsuiting die gehanteerd worden op het platform en, zodoende, geven zij tevens vorm aan de democratische cultuur op wereldschaal.[4]

Online platformen beheren de meerderheid van alle online meningsuitingen.[5] Dit nieuwe tijdperk van 'platformkapitalisme' heeft radicale veranderingen teweeggebracht in het maatschappelijke communicatieproces. Via online platformen zoals Facebook, Twitter en YouTube verwerven burgers wereldwijd toegang tot informatie en kunnen zij hun meningen, ideeën en gedachten uiten. De enige vereiste om van deze mogelijkheid gebruik te kunnen maken is toegang tot het internet, waardoor een ongekend open en inclusief expressieforum ontstaat. Online platformen spelen aldus een belangrijke rol in de digitale publieke ruimte van de 21e eeuw.[6] Zij bevorderen de uitwisseling van meningen en ideeën op wereldschaal en worden daarom soms de 'gouverneurs van online meningsuiting'[7] of 'hoeders van de publieke ruimte'[8] genoemd.

Deze verscheidenheid aan online sites en diensten

* Juncal Montero Regules is FWO-aspirant Fundamenteel Onderzoek aan de Universiteit Hasselt.

1 T. Flew, F. Martin, N. Suzor, "Internet Regulation as Media Policy: Rethinking the Question of Digital Communication Platform Governance", *Journal of Digital Media & Policy*, 2019, 33-50. Het artikel werd oorspronkelijk geschreven in het Engels. Vertaling door Caroline De Geest.
2 G. De Gregorio, "Democratising Online Content Moderation: A Constitutional Framework", *Computer Law & Security Review*, 2020.
3 T. Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*, Yale University Press, 2018.
4 M. Ammori, "The New New York Times: Free Speech Lawyering in the Age of Google and Twitter", *Harvard Law Review*, 2013, 2259-2295.
5 T T. Flew, F. Martin, N. Suzor (n 3).
6 J. Balkin, "How to Regulate (and Not Regulate) Social Media", *Knight Institute's Occasional Papers*, 2020.
7 K. Klonick, "The New Governors: The People, Rules, And Processes Governing Online Speech", *Harvard Law Review*, 2018, 1599-1670.
8 T. Gillespie (n 6).

– die eigendom zijn van én beheerd worden door particuliere ondernemingen – controleren de inhoud van boodschappen die online gedeeld worden. Dit proces van screenen, evalueren, categoriseren, goedkeuren, verwijderen of verbergen van online-inhoud noemt men *contentmoderatie*.[9] Online platformen maken gebruik van moderatiepraktijken om online inhoud te beheren en presenteren. Dit moet onder meer positief online communicatiegedrag ondersteunen en afdwingen, en agressie of asociaal gedrag tot een minimum beperken.

Volgens Gillespie is contentmoderatie een bepalende functie van online platformen.[10] Het doel, vanuit het oogpunt van de platformen, is dubbel: de handhaving van hun missie ten aanzien van hun gebruikers en het veiligstellen van hun voordelen als particuliere onderneming waarvan de activiteiten gebaseerd zijn op gegevens en reclame.

Online platformen modereren dus inhoud door middel van een reeks regels, instrumenten en processen.[11] De impliciete regels – een vorm van architecturale regulering[12] – liggen vast in ontoegankelijke codes en algoritmen, die bepalen hoe inhoud wordt georganiseerd, gepromoot en gepresenteerd aan gebruikers.[13] Algoritmen verlenen en beheren de zichtbaarheid, sturen de informatiestroom en structureren de participatie.[14] Daarnaast worden expliciete regels vastgelegd in drie schriftelijke en deels toegankelijke documenten: de servicevoorwaarden, communityrichtlijnen en interne richtsnoeren. *Servicevoorwaarden* maken deel uit van het contract tussen het platformbedrijf en de gebruiker en leggen de regels voor interactie binnen het platform vast. Communityrichtlijnen zijn de normen die platformen toepassen om inhoud te controleren. Ze leggen gebruikers uit welk gedrag en welke inhoud aanvaardbaar zijn op hun site. Maar niet alle inhoudsnormen zijn terug te vinden in de communityrichtlijnen. De meeste platformen vertrouwen op *interne richtsnoeren*, waarin de moderatienormen worden beschreven. Deze documenten vormen een gedetailleerd regelboek over wat de gepubliceerde richtsnoeren in de praktijk betekenen en zijn vaak niet toegankelijk voor de gebruikers.

De expliciete regels worden geïmplementeerd en gehandhaafd door middel van een reeks moderatie-instrumenten en -technieken. Deze bepalen of de door gebruikers gegenereerde inhoud in overeenstemming is met de regels van het platform (en mag blijven staan), dan wel in strijd is met de inhoudsnormen (en moet worden verwijderd). Die moderatie gebeurt vóór (ex-ante) of na (ex-post) de publicatie van de inhoud. De eerste methode is doorgaans geautomatiseerd en omvat bijvoorbeeld technieken die afbeeldingen herkennen en inhoud scannen.[15] De tweede methode gebruikt zowel algoritmische als menselijke middelen en kan het resultaat zijn van geautomatiseerde markering[16] of door de gemeenschap opgelegde normen. De moderatieprocessen variëren naar gelang van de specifieke kenmerken van de platforms, zoals hun functie, omvang, middelen en beleid.[17] Ze hebben tot doel de inhoud, bedoeling, gevolgen en betekenis ervan te toetsen aan de regels van de platformen.[18]

Contentmoderatie raakt aan meer dan vrijheid van meningsuiting

Door inhoud te modereren, geven platforms vorm aan de publieke meningsuiting. Zoals hierboven toegelicht, is dit deels positief, aangezien individuen hierdoor meer kansen krijgen om hun grondrechten, zoals de vrijheid van meningsuiting, uit te oefenen. Maar contentmoderatie geeft ook aanleiding tot bezorgdheid. De uitdagingen zijn terug te brengen tot drie clusters .[19]

Particuliere belangen vs. publieke waarden

9 T. Flew, F. Martin, N. Suzor (n 3) 40.
10 Zie: T. Gillespie (n 6); T. Gillespie, "Platforms Are Not Intermediaries", Georgetown *Law Technology Review*, 2018, 198-216.
11 Zie T. Gillespie (n 6); K. Klonick (n 10).
12 T. Gillespie (n 6) hfdst. 7.
13 Zoals Lessig meer dan twintig jaar geleden al aangaf. Zie L. Lessig, *Code and Other Laws of Cyberspace*, Basic Books, 1999.
14 Zie M. Bunting, "From Editorial Obligation to Procedural Accountability: Policy Approaches to Online Content in the Era of Information Intermediaries", *Journal of Cyber Policy*, 2018, 165-186.
15 B. Sander, "Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights Based Approach to Content Moderation", *Fordham International Law Journal*, 2020, 939-1006, 946.
16 S. T. Roberts, "Content Moderation", *Encyclopedia of Big Data*, University of California, 2017.
17 Caplan (n 5). Caplan onderscheidt drie grote benaderingen van platformmoderatie: ambachtelijke benaderingen, waarbij inhoud geval per geval wordt beoordeeld door teams van 5 tot 200 menselijke moderatoren (bv. Medium); community-afhankelijke benaderingen, waarbij overkoepelende beleidsbeslissingen worden gecombineerd door een klein team van bedrijfsmedewerkers met een grotere groep vrijwillige menselijke beoordelaars (bv. Reddit); en industriële benaderingen, waarbij doorgaans een beroep wordt gedaan op grootschalige bureaucratieën van tienduizenden menselijke beoordelaars (bv. Facebook).
18 S. T. Roberts, *Content Moderation in the Shadows of Social Media*, Yale University Press, 2019.
19 Deze karakterisering is niet uitputtend.

De 'platformisering' van het internet heeft ook geleid tot een 'platformisering' van spraak. Een beperkt aantal online platformen bepaalt de voorwaarden waaronder enorme hoeveelheden inhoud online stromen. Deze digitale convergentie vormt een grote bedreiging voor het pluralisme in de media.[20]

Bovendien zijn de technologiebedrijven die platformen beheren en exploiteren particuliere ondernemingen die winst najagen. Hun waarden verschillen bijgevolg van de waarden van publieke actoren en zijn niet noodzakelijkerwijze afgestemd op de bredere waarden in de samenleving. Online platformen beweren vaak dat hun diensten ten goede komen aan het grote publiek en dat zij regeringen en gemeenschappen kunnen bijstaan bij de zelforganisatie van mensen online.[21] Terecht rijzen evenwel vragen bij de oprechtheid van platformen die zich verbinden aan publieke waarden als democratie en de eerbiediging van grondrechten. De vraag naar de verenigbaarheid van particuliere belangen van grote *tech*-bedrijven met publieke waarden is vooral ethisch van aard: in hoeverre zijn wij als samenleving bereid te aanvaarden dat leidende beginselen van het publiek debat worden opgelegd door commercieel aangestuurde private actoren?

Op juridisch vlak houdt het particuliere karakter van platformen in dat zij niet gebonden zijn door mensenrechten. Internationale mensenrechten zijn immers alleen verticaal bindend voor overheidsactoren en zijn slechts in een beperkt aantal gevallen horizontaal van toepassing in de relatie tussen particuliere actoren.

Wat met algoritmische besluitvorming?

De uitdagingen van algoritmische contentmoderatie zijn gestoeld op vier pijlers:[23]

- **Een gebrek aan transparantie:** geautomatiseerde besluitvorming introduceert een aanzienlijke laag van complexiteit, aangezien algoritmen en de specifieke databases voor de besluitvorming opzettelijk vaag worden gelaten.[24]
- **Een discriminerend effect**: algoritmische besluitvormingssystemen kunnen verschillende wettelijk beschermde groepen oneerlijk behandelen. Algoritmen bestaan niet in een leegte en er wordt vaak gezegd dat zij sociaal-politieke vooroordelen reproduceren.[25]
- **Een toenemende mathematisering van het recht**: het gebruik van AI voor contentmoderatie bakent in toenemende mate de grenzen van online meningsuiting af, aangezien het bepaalt wat wel en wat niet tot de expressieve publieke sfeer kan behoren. Algoritmische berekening draagt dus bij tot de definitie van het begrip legaliteit.[26]
- **Het risico op fouten**: geautomatiseerde systemen nemen waarschijnlijk honderden, zo niet duizenden onjuiste beslissingen per dag, aangezien het voor algoritmen zeer moeilijk is om complexe, contextgebonden beslissingen te nemen, vooral met betrekking tot open concepten, zoals haatspraak.

Procedurele gebreken van contentmoderatie

De informatie die online platformen verstrekken over hun regels en procedures voor contentmoderatie is ondoorzichtig. Dit algemene gebrek aan transparantie wordt omschreven als wetteloos en bedreigend voor de rechtsstaat.[27] Ondanks de inspanningen van bedrijven om transparanter te lijken, blijkt dat de beschikbare publieke informatie slechts een fractie van de moderatiepraktijken van online platformen vertegenwoordigt. Deze hoge mate van ondoorzichtigheid bedreigt democratische waarden.

Bovendien zijn de mogelijkheden om in beroep te gaan tegen moderatiebesluiten schaars en ontbreekt het aan onafhankelijkheid en transparantie. Er is weinig of geen openbare informatie voor gebruikers over hoe zij te werk moeten gaan wanneer zij het niet eens zijn met een besluit tot schorsing.[28] De beroepsprocedures zijn vooral ondoorzichtig wat het besluitvormingsproces betreft, wat ook vragen

20 De Gregorio (n 2) 6.
21 J. van Dijck, T. Poell, M. de Waal, *The Platform Society: Public Values in a Connective World*, Oxford University Press, 2018.
22 Er zijn academici die ervoor pleiten de internationale mensenrechtenwetgeving te gebruiken als een kader voor inhoudelijke matiging. Zie D. Kaye, Speciaal rapporteur van de VN-Mensenrechtenraad voor de bevordering en bescherming van het recht op vrijheid van mening en meningsuiting en Secretariaat van de VN-Mensenrechtenraad, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, 2018; B. Sander (n 19); E. Douek, "The Limits of International Law in Content Moderation", *UCI Journal of International, Transnational, and Comparative Law*, 2021, 1556-5068.
23 Zie: H. Bloch-Wehba, "Automation in Moderation", *Cornell International Law Journal*, 2020, 41-96.
24 Zie bijvoorbeeld: M. Ananny, K. Crawford, "Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability", *New Media & Society*, 2018, 973-989.
25 R. Gorwa, R. Binns, C. Katzenbach, "Algoritmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance", *Big Data & Society*, 2020.
26 De Gregorio (n 2) 3.
27 N. Suzor, *Lawless: The Secret Rules That Govern Our Digital Lives*, Cambridge, Cambridge University Press, 2019.
28 Onlinecensorship.Org - How to Appeal' https://onlinecensorship.org/resources/how-to-appeal.

doet rijzen over de mate van onafhankelijkheid ervan. Aangezien de beroepsbeslissingen binnen het platform door menselijke moderatoren worden genomen, is hierop geen extern toezicht of controle. De beroepsmogelijkheden zijn dus beperkt.

Een experimentele oplossing: de Facebook Oversight Board

Facebook is het meest dominante online platform ter wereld, met 2,32 miljard maandelijks actieve gebruikers in meer dan 100 verschillende talen.[29] Dagelijks wordt een enorme hoeveelheid door gebruikers gegenereerde inhoud geüpload en gedeeld binnen het platform.[30] Facebook ontvangt per dag meer dan een miljoen meldingen van schendingen van zijn inhoudsnormen. Deze cijfers illustreren een van de belangrijkste uitdagingen waarmee grote platformen bij contentmoderatie worden geconfronteerd: de schaal.

Facebook beantwoordt deze schaalgrote met een industriële aanpak. Het bedrijf legt duidelijke regels vast voor de toepassing van gemeenschapsnormen, om de besluitvorming over contentmoderatie routinematig en efficiënt te maken. Hierdoor ontstaat een 'besluitvormingsfabriek'. Die maakt gebruik van zowel algoritmen als mensen om contentmoderatie beheersbaar te maken. Door algoritmen gemarkeerde inhoud wordt beoordeeld door een menselijke moderator, die bepaalt of de inhoud in strijd is met de community-normen van het platform, in welk geval het van de site wordt verwijderd. Het is mogelijk om tegen verwijderingsbeslissingen in beroep te gaan bij inhoudmoderatieteams die klachten doorsturen naar speciale beoordelaars die over het beroep beslissen.[31]

Daarnaast werd een uniek en onafhankelijk orgaan gecreëerd dat zou fungeren als beroepsinstantie voor de contentmoderatiebeslissingen van het platform.[32] De Facebook Oversight Board (hierna: de Oversight Board of de Raad) is operationeel sinds december 2020.[33] De Raad bestaat uit 20 experten, allen door Facebook benoemd en betaald, op vlak van mensenrechten, journalistiek, politiek en computerwetenschappen.

De Oversight Board heeft twee hoofdfuncties. Ten eerste onderzoekt ze concrete gevallen van potentieel aanstootgevende inhoud en beoordeelt ze of de inhoud in kwestie al dan niet op Facebook mag blijven. De Raad ontvangt verzoeken van Instagram- of Facebook-gebruikers die rechtstreeks getroffen zijn door een moderatiebesluit: zij vragen om bepaalde inhoud die van een van beide platforms is verwijderd weer op te nemen of om bepaalde toegelaten inhoud te verwijderen.[34] De bindende besluiten van de Raad moeten het bedrijf helpen bij de ontwikkeling van onafhankelijke richtsnoeren voor contentmoderatie. Ten tweede mag de Raad beleidsaanbevelingen doen over de moderatiepraktijken van Facebook. Deze adviezen zijn niet bindend, maar Facebook moet er wel publiekelijk op reageren.

De Facebook Oversight Board als oplossing voor alle contentmoderatie problemen?

Zowel de oprichting van de Raad als zijn eerste beslissingen worden omschreven als een scharniermoment in de geschiedenis van het beheer van online meningsuiting.[36] Het is inderdaad de eerste keer dat een platform een toezichtmechanisme opzet dat fungeert als beroepsinstantie voor zijn eigen moderatiebesluiten. De relevantie van deze ontwikkeling mag niet worden onderschat, noch overschat. De vraag rijst met name in hoeverre de Raad oplossingen kan bieden voor de hierboven beschreven problemen bij contentmoderatie.

De Raad versterkt het probleem van de spanning tussen particuliere belangen en publieke waarden

De oprichting van de Oversight Board is een vorm van zelfregulering. Het marginaliseert de rol van publieke actoren bij het waarborgen van de online handhaving van grondrechten. Dit wordt nog versterkt door het feit dat de besluitvorming van de Oversight Board onvermijdelijk wordt beïnvloed door de private normen die Facebook heeft gedefinieerd in zijn communityrichtlijnen en interne richtsnoeren. Aan de quasi-normatieve en uitvoerende bevoegdheden

29 Bedrijfsinfo' (Over Facebook) https://about.fb.com/company-info/ geraadpleegd op 30 april 2021.
30 Elke dag worden meer dan 2,5 miljard stukken content op het platform gedeeld.
31 K. Klonick (n 10) 1648.
32 De Oversight Board manifesteert Zuckerbergs idee dat "Facebook niet in ons eentje zoveel belangrijke beslissingen moet nemen over vrije meningsuiting en veiligheid".
33 Voor een gedetailleerde uiteenzetting over het totstandkomingsproces van de Oversight Board, zie K. Klonick, "The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression", *Yale Law Journal*, 2019, 2418-2499.
34 Deze laatste optie is ingevoerd op 13 April 2021.
35 Sinds april 2021 heeft de Oversight Board in acht zaken uitspraak gedaan. In zes daarvan heeft de Raad besloten het verwijderingsbesluit van Facebook ongedaan te maken en is de inhoud dus teruggezet op het platform, terwijl in twee andere gevallen het verwijderingsbesluit van Facebook is gehandhaafd en de inhoud dus verwijderd is gebleven. De beslissingen in deze zaken zijn hier beschikbaar: 'Decision | Oversight Board' https://www.oversightboard.com/decision/; zie ook: 'Overseeing the Facebook Oversight Board' (*Lawfare* 30 April 2021).
36 E. Douek (n 34).

die het platform zelf uitoefent via zijn regels voor contentmoderatie en handhavingsprocedures, voegt de Oversight Board een quasi-rechterlijke functie toe, aangezien het *de facto* oordeelt in grondrechtenkwesties. In die zin is de Raad een nieuwe stap in de richting van private handhaving en de institutionalisering van private rechtspraak.

Gebrek aan transparantie wordt deels ondervangen maar blijft ook deels bestaan

In zijn besluitvorming en beleidsaanbevelingen onthult de Oversight Board interne moderatieprocessen en -normen die niet publiek beschikbaar zijn. Bovendien zorgt het toezicht ervoor dat Facebook meer verantwoording aflegt over zijn beheer van online meningsuiting. Deze transparantie en verantwoordingsplicht zijn opmerkelijk, maar de relevantie ervan wordt enigszins ondermijnd door het feit dat de Oversight Board alleen beslissingen kan nemen in post-moderatiezaken. Tegen ex-ante moderatieprocessen, zoals het aanbevelen van algoritmen, is geen beroep mogelijk. Bovendien hangt de jurisdictie van de Raad af van wat Facebook in zijn werkingssfeer besluit.

Gemiste kans op betere procedurele rechten

De Raad voert een beroepsprocedure in die fungeert als instantie van tweede aanleg voor moderatiebesluiten. Die beroepsmogelijkheid komt rechtstreeks tegemoet aan de roep om betere procedurele rechten. De doeltreffendheid ervan wordt echter beperkt door de toegang tot de beroepsprocedure, aangezien Facebook zelf beslist welke zaak voor het Hof komt.

Het de-platformen van Trump: een primeur met grote waarde

De Raad nam, zeer recent, het besluit van Facebook om de profielen van Donald Trump voor onbepaalde tijd te schorsen onder de loep. Dit is de meest geruchtmakende zaak van de Raad – en waarschijnlijk een van de meest relevante gevallen van contentmoderatie waarmee sociale media platformen tot dusver te maken kregen. De perceptie heerst evenwel dat de zaak enkel voor de Raad werd gebracht door Facebook om de hete aardappel over de schorsing van Trumps profielen door te schuiven naar het quasi-judicieel orgaan. De CEO van Twitter heeft Trump wél definitief van Twitter geweerd ten gunste van de mensenrechten en ter ondersteuning van onderliggende progressieve sociale bewegingen in de Verenigde Staten.[37]

Facebook stelde de Raad twee vragen: was zijn beslissing om het profiel van Donald Trump voor onbepaalde tijd te schorsen juist, en moet die schorsing wel of niet worden gehandhaafd? Het profiel van een president schorsen, is geen rechttoe rechtaan besluit. Het doet de vraag rijzen in hoeverre Facebook bij dergelijke beslissing rekening moet houden met de bredere sociale context en de gebeurtenissen buiten zijn platform. Het besluit om Trump te bannen was meer het gevolg van de gebeurtenissen op het Capitool – en de omringende politieke context – dan van iets wat de toenmalige president direct voorafgaand aan de schorsing had gepost.

De tweede vraag was een verzoek om beleidsrichtlijnen: heeft de Raad opmerkingen of aanbevelingen over hoe Facebook schorsingen moet behandelen wanneer een gebruiker een politiek leider is? Hoe Facebook publieke figuren behandelt, is een van zijn meest controversiële beleidsterreinen. Het platform beschouwt uitspraken van politici als inherent nieuwswaardig en daarom in het algemeen belang, zelfs wanneer deze in strijd zijn met de gemeenschapsnormen van het platform. Men kan namelijk aanvoeren dat de democratie vereist dat kiezers weten wie hun kandidaten werkelijk zijn en wat zij geloven, zelfs (of misschien vooral) wanneer die overtuigingen weerzinwekkend zijn. Ook kan worden aangevoerd dat publieke figuren met aanzienlijke invloed juist aan hogere normen moeten worden gehouden dan andere gebruikers.

Op 5 mei 2021 heeft de Raad in haar beslissing het besluit van Facebook bekrachtigd om het profiel van Trump te schorsen van het sociale media platform.[38] De Raad is van oordeel dat de twee posts die de aanleiding vormden voor de schorsing van het profiel een ernstige schending inhouden van de *Facebook Standard on Dangerous Individuals and Organizations*, en dat de bescherming van het publiek (de waarde "Safety") een beperking van de vrijheid van meningsuiting (de waarde "Voice") rechtvaardigde. Belangrijk is dat de Raad evenwel van oordeel is dat de door Facebook opgelegde schorsing voor onbepaalde tijd zowel vaag als onzeker was. Een dergelijke willekeurige beslissing is immers niet opgenomen in de *Community Standards*.

Na een beoordeling van "de context, de status van de spreker, de intentie, de inhoud en de vorm, de

37 Permanente schorsing van @realDonaldTrump' https://blog.twitter.com/en_us/topics/company/2020/suspension.html geraadpleegd op 1 mei 2021.

38 "Raad van Toezicht | Onafhankelijk oordeel. Transparantie. Legitimiteit." https://www.oversightboard.com/decision/FB-691QAMHJ/ geraadpleegd op 25 mei 2021.

omvang en het bereik, en de dreiging van schade", oordeelt de Raad dat de berichten van Trump geacht kunnen worden een ernstig risico op het aanzetten tot discriminatie en geweld te creëren. Het was bijgevolg gerechtvaardigd om beperkingen op te leggen. Maar het was niet passend dat Facebook een schorsing voor onbepaalde tijd oplegde, aangezien voor een dergelijke maatregel geen duidelijke en openbaar toegankelijke procedure bestond. De Raad besluit dat Facebook zijn besluit binnen de zes maanden opnieuw moet bekijken en beslissen wat er met het profiel van Trump moet gebeuren. De hete aardappel ligt dus opnieuw in de handen van Facebook.

Met deze beslissing lijkt de Oversight Board naar een gedeelde verantwoordelijkheid met Facebook te streven. Hoewel de inhoudelijke beoordeling van de zaak aanleiding geeft tot handhaving van de schorsing, heeft de Raad zich uiteindelijk gebaseerd op de procedurele kwestie van het gebrek aan duidelijkheid en openbaarheid van de procedure, om de kwestie zodoende opnieuw door te spelen naar Facebook. Dit genuanceerde standpunt geeft echter geen duidelijk antwoord op de eerste door Facebook gestelde vraag[39] en verzuimt verder specifieke principes of richtlijnen op te stellen voor het nemen van de uiteindelijke beslissing in zowel deze zaak als toekomstige soortgelijke zaken met betrekking tot politieke leiders, wijdverspreide profielen en het aanzetten tot geweld. Het is een gemiste kans, en één die voor de Oversight Board een zeker verlies aan legitimiteit kan inhouden.

Een meer richtinggevend besluit van de Raad zou pertinent zijn geweest om twee redenen: de wereldwijde impact ervan en de legitimiteit van en vertrouwen in de Raad zelf. Het eerste heeft betrekking op politici in andere landen die (beweerdelijk) aangezet hebben tot haat of geweld. Hun sociale media profielen werden immers niet verwijderd[40] en dus ontstaat er een schijnbare inconsistentie in de wijze waarop Facebook de profielen van Trump en andere leiders behandelt. Deze inconsistentie blijft overeind. Ten tweede had deze zaak ook kunnen bijdragen aan de legitimiteit van de Raad. De doorverwijzing door Facebook naar de Raad suggereerde immers dat het bedrijf vertrouwen had in het experiment en niet bang was voor consequente beslissingen. Door de bal (deels) terug te kaatsen naar Facebook heeft de Oversight Board evenwel niet noodzakelijk de kans gegrepen om zijn legitimiteit te versterken.

Besluit

De 'platformisering' van meningsuiting biedt burgers ongekende mogelijkheden om zich uit te drukken, maar geeft ook aanleiding tot ernstige bezorgdheid over de bescherming van grondrechten in de online wereld. Wanneer online platformen inhoud modereren, doen zij dit met behulp van hun eigen regels en methoden, die zij zelf ontwikkelen en toepassen. Hierbij onttrekken zij zich aan het toezicht van publieke actoren. De moderatiepraktijken van de platformen zorgen voor een aantal uitdagingen, die vooral verband houden met de aard van de meningsuiting, de aard van de platformen zelf, algoritmische besluitvorming en procedurele gebreken.

Het nieuwste experiment op het gebied van contentmoderatie is de Facebook Oversight Board. Deze Raad, die zowel geprezen als bekritiseerd wordt, kan de bestaande problemen op zijn eentje niet oplossen. De Raad kan wel zorgen voor meer transparantie en verantwoording van 's werelds meest dominante platform. De beslissing van de Raad in de zaak van de schorsing van het profiel van Trump door Facebook is minder bepalend dan aanvankelijk verwacht kon worden. De Oversight Board is niettemin een nieuw model van particuliere rechtspraak dat de macht kan matigen van particuliere bedrijven die anders aan democratisch toezicht ontsnappen..

39 Zie het gedeelte "Beslissen over de ont-platting van Trump".
40 Denk bijvoorbeeld aan de Filipijnse president Duterte, en de Braziliaanse president Bolsonaro.