

Niet verbieden, maar aanpassen

Position paper over het voorgenomen verbod op verslavende, polariserende en antidemocratische algoritmen

Het Rathenau Instituut deed op verzoek van de vaste Tweede Kamercommissie Digitale Zaken onderzoek naar de rol van aanbevelingsalgoritmen bij inmenging in verkiezingen.

Aanvullend op dat rapport reageert het Rathenau Instituut met dit position paper op het voornemen in het coalitieakkoord om “verslavende, polariserende, en antidemocratische algoritmes”, te verbieden.^{1 2} In dit paper bespreken we de mogelijke gevolgen van een verbod. Ook doen we aanbevelingen om problematische aspecten van de huidige aanbevelingsalgoritmen van socialemediaplatformen aan te pakken.

1.1 Aanleiding

Socialemediaplatformen zijn steeds sterker gaan leunen op aanbevelingsalgoritmen die zijn ontwikkeld op signalen van engagement. Platformbedrijven hebben zo een online omgeving gecreëerd waar inhoud die inspeelt op emotie, misleiding en opruiing een groot bereik kan krijgen. Kwaadwillende buitenlandse statelijke actoren kunnen inspelen op aanbevelingsalgoritmen door de zichtbaarheid van bepaalde inhoud te vergroten of te overstemmen.³

In het coalitieakkoord staat dat het huidige Kabinet zich in Europees verband wil inzetten voor het verbieden van een specifiek type algoritmen, namelijk *verslavende, polariserende en antidemocratische* algoritmen.

In dit paper richten we ons op *polariserende* en *antidemocratische* algoritmen.

In het coalitieakkoord is het verbod nog niet uitgewerkt. Het verbod zou gericht kunnen zijn op:

1. polariserende en antidemocratische *inhoud*
2. polariserend en antidemocratisch *ontwerp* van aanbevelingssystemen
3. polariserende en antidemocratische *gevolgen* voor gebruikers bij het zien van bepaalde inhoud

In dit position paper wijzen we erop dat een verbod op bepaalde inhoud die wordt verspreid door aanbevelingsalgoritmen, op gespannen voet kan staan met de vrijheid van meningsuiting. Ingrijpen in het ontwerp van aanbevelingsalgoritmen is een effectievere methode die zowel de vrijheid van meningsuiting kan blijven beschermen, als mogelijk negatieve gevolgen van deze systemen kan aanpakken.

We beschrijven hieronder eerst de verschillende onderdelen waarop een verbod gericht kan zijn. We sluiten af met aanbevelingen over hoe de Kamer de mogelijke schadelijke gevolgen van aanbevelingsalgoritmen kan minimaliseren.

¹ Dit verbod staat beschreven in het coalitieakkoord. Ook kreeg het Rathenau Instituut hierover vragen tijdens de [technische briefing over algoritmen en inmenging](#) op 5 maart 2025

² D66, VVD en CDA, [Coalitieakkoord-2026-2030](#), 59, 30 januari 2026

³ Rathenau Instituut, [Scrollend naar de stembus](#), 11 maart 2026

1.2 Polariserende en antidemocratische *inhoud* aanpakken

Partijen die overwegen antidemocratische en polariserende inhoud op sociale media te verbieden, moeten afwegen welke inhoud onder de vrijheid van meningsuiting valt en welke inhoud al verboden is. Internetgebruikers mogen zonder tussenkomst vanuit de overheid zeggen wat ze willen en inhoud zoeken en ontvangen die ze willen. Ook polariserende uitingen “*to offend, shock or disturb*” zijn volgens het Europees Hof toegestaan. Dit type inhoud verbieden is dus in strijd met het recht op de vrijheid van meningsuiting zoals die is beschermd in de grondwet en het Europees Verdrag voor de Rechten van de Mens (EVRM). Antidemocratische en polariserende inhoud waaruit smaad, laster, groepsbelediging spreekt, of die aanzetten tot haat, geweld of discriminatie is wel verboden door het nationale strafrecht.

Om het rondgaan van illegale inhoud aan te pakken heeft de overheid met de digitale dienstenverordening (DSA) een uitgebreid wettelijk instrumentarium in handen om zowel de grenzen van de vrijheid van meningsuiting online te waarborgen, als de verspreiding van illegale inhoud via platformen te beperken. De DSA bepaalt niet wat strafbare inhoud is - dat wordt bepaald door het nationale strafrecht van lidstaten. Maar het regelt wél hoe platformen moeten optreden tegen inhoud die op strafrechtelijke gronden illegaal is. Uit ons rapport komt naar voren dat er signalen zijn dat platformen tekortschieten bij het verwijderen van illegale inhoud en daarmee met het naleven van de DSA. Ook biedt de DSA mogelijkheden om platformen af te dwingen om meer informatie vrij te geven. Maar dit wettelijke Europese kader staat onder druk door de simplificeringsagenda van de Europese Commissie en oplopende geopolitieke spanningen tussen de EU en de Verenigde Staten.^{4 5}

1.3 Polariserend en antidemocratisch *ontwerp* van aanbevelingsalgoritmen aanpakken

Aanbevelingsalgoritmen zijn een reactie van platformen op de *information overload* waaraan gebruikers op sociale media worden blootgesteld. Er wordt dagelijks zoveel inhoud geüpload, dat platformen aanbevelingsalgoritmen inzetten om voor elke gebruiker een selectie te maken. Een verbod op aanbevelingsalgoritmen als zodanig is daarom ook niet per definitie wenselijk: aanbevelingsalgoritmen zijn behulpzaam om de veelheid van informatie te filteren en sorteren. Het is mogelijk effectiever om eisen te stellen aan hoe aanbevelingsalgoritmen werken. Daarmee wordt ingegrepen in *hoe* bepaald wordt *welk type inhoud* wordt aanbevolen. Oftewel: wie online een megafon krijgt en wie niet.

Platformen en hun ontwerpers hebben een voorkeur voor inhoud die gebruikers lang op het platform houdt en het platform vaak doen bezoeken. Want hoe langer gebruikers online zijn, hoe hoger de advertentie-inkomsten. Sinds 2016 stapten veel grote platformen over op aanbevelingsalgoritmen die inhoud rangschikken op basis van het verwachte engagement (*likes*, reacties, tijd besteed, enzovoort.). Inmiddels zijn dit type aanbevelingsalgoritmen de norm geworden.

Als gevolg hiervan kan de tijdlijn van gebruikers vol staan met inhoud die weliswaar rechtmatig is, maar tegelijkertijd dermate prikkelend is dat er een reactie wordt ontlokt, zoals *liken*,

⁴ De Europese Commissie heeft voorgesteld om met een zogenaamde Digitale Omnibus regels voor dataverzameling en verwerking door bedrijven te versimpelen (Digital Omnibus Proposal COM/2025/837, 2025). Hoewel bij het bedrijfsleven behoefte is aan minder regeldruk, bestaan er zorgen, waaronder bij het 127 maatschappelijke organisaties. People vs. Big Tech. (2025, 13 november). “*Stop the Digital Omnibus,*” Say 127 Civil Society Organisations. <https://peoplevsbig.tech/the-eu-must-uphold-hard-won-protections-for-digital-human-rights/>. brief van 127 ngo's. Maar ook ons eigen kabinet maakt zich zorgen: Ministerie van Buitenlandse Zaken (2025). *BNC-fiche 2: Omnibus AI en Omnibus Digitaal (bijlage bij Kamerbrief inzake informatievoorziening over nieuwe Commissievoorstellen)*. <https://www.rijksoverheid.nl/documenten/publicaties/2025/12/12/bnc-fiche-2-omnibus-ai-en-omnibus-digitaal>.

⁵ Rathenau Instituut, *Scrollend naar de stembus*, 11 maart 2026.

commentaar geven, of tijd besteden. Deze inhoud kan ongevraagd in de tijdlijn van gebruikers belanden. Reacties (*engagement*) op inhoud veroorzaken dat dit soort inhoud aan nog meer gebruikers wordt voorgeschoteld.

Uit onderzoek weten we dat inhoud die emoties opwekt, *borderline content* en inhoud die rechts-extremistische politieke uitspraken bevat, bovengemiddeld vaak wordt aanbevolen door aanbevelingsalgoritmen gebaseerd op engagement. Een effect daarvan is dat gebruikers die er niet bewust voor kiezen om dit type inhoud op te zoeken, hier wél mee worden geconfronteerd in hun tijdlijn.

Een verbod op antidemocratische en polariserende aanbevelingsalgoritmen zou gericht kunnen zijn op aanbevelingsalgoritmen die werken met engagementssignalen. Maar wie een verbod overweegt op een type aanbevelingsalgoritmen, moet tegelijk overwegen welke alternatieve aanbevelingsalgoritmen beter zijn. Alternatieve aanbevelingsalgoritmen hebben eveneens bekende en nog onbekende voor- en nadelen

In **tabel 1** (p. 4) staan suggesties voor andere vormen van platformontwerp. Sommige van deze interventies zijn mogelijk al door platformen getest. Andere zijn alleen nog getoetst in wetenschappelijke experimenten of geopperd door experts. Wat de gevolgen zijn van de alternatieve ontwerpen in de praktijk moet worden onderzocht. Maar met de huidige beperkte informatiepositie van onafhankelijke partijen ten opzichte van socialemediaplatformen, is dit vrijwel onmogelijk.

Momenteel ligt de keuze over hoe bepaald wordt welk type inhoud wordt aanbevolen, bij een handjevol grote socialemediabedrijven die een dominante marktpositie hebben. Op dit moment geven deze bedrijven weinig vrijheid aan hun gebruikers om zelf te bepalen op welke manier ze aanbevelingsalgoritmen willen gebruiken. Wanneer deze machtsconcentratie zou worden doorbroken, kunnen alternatieve aanbevelingsalgoritmen ontstaan. Het aanbevelingsalgoritme gebaseerd op engagement zou dan slechts een van de aangeboden opties zijn.

Aanpassing	Wat is het?	Voorbeelden
Rangschikking gebaseerd op het overbruggen van groepen mensen (Lasser & Poehhacker, 2025)	Platformen ontwerpen hun aanbevelingsalgoritmen met andere doelen, zoals het vergroten van vertrouwen tussen groepen mensen, in plaats van het doel van zoveel mogelijk engagement.	Recent wetenschappelijk onderzoek naar bridging based ranking.
Circuit breakers (Snap, 2025) (Pathak & Spezzano, 2024)	Extra controle voordat informatie viral kan gaan.	Snap geeft aan dat inhoud eerst door menselijke moderatoren wordt bekeken voordat het aangeraden wordt aan een groot publiek.
Kwaliteitscores (Cunningham, 2023)	Content met een lagere mate van kwaliteit, zoals met generatieve AI-gegenereerde beelden, krijgt minder gewicht door het aanbevelingsalgoritme.	
Diversificatie-algoritmen (Bengani, 2023)	Diversificatie-algoritmen kunnen bij het filteren en rangschikken van inhoud zorgen voor een diverse set aan inhoud.	Platformen geven aan dit te doen, maar hoe zwaar diversiteit weegt in de aanbevelingen is onbekend. Diversificatie van inhoud kan op veel verschillende manieren: op onderwerp, geografische afkomst, bron, type media en populariteit.
Gebruikersfeedback op inhoud zwaarder laten wegen (Cunningham et al., 2025; Stray et al., 2024)	Door gebruikers explicieter te bevragen, bijvoorbeeld door gebruikerssurveys, krijgen platformen meer inzicht in de voorkeuren van gebruikers. Dat terwijl aanbevelingsalgoritmen nu vaak uitgaan van afgeleide voorkeuren, zoals likes en kijktijd.	Wanneer mensen veel vergelijkbare inhoud te zien krijgen, zeker in verkiezingstijd, zouden platformen gebruikers actief kunnen bevragen of deze inhoud waardevol voor hen is en wat de inhoud doet met hun attitudes over verkiezingen.
Downranking van inhoud (Piccardi et al., 2025)	In plaats van inhoud te verwijderen, kunnen platformen borderline-inhoud of antidemocratische inhoud minder gewicht geven in hun aanbevelingsalgoritmen.	Een recent experiment van onderzoekers laat zien dat het lager rangschikken van antidemocratische en polariserende berichten op X een positief effect had op hoe gebruikers denken over mensen met andere politieke overtuigingen.
Alle inhoud (ook comments) handmatig goedkeuren (Ribeiro et al., 2023)	Door comments en inhoud handmatig goed te keuren, lijken comments die niet aan de regels te voldoen, te verminderen. Dat kan de kwaliteit van het gesprek op social media vergroten.	In Facebookgroepen bestaat de mogelijkheid voor moderatoren om alle comments handmatig goed te keuren.
Community notes (Chuai et al., 2024)	Door de contextuele informatie die gebruikers geven bij een post te gebruiken als signaal voor het aanbevelingsalgoritme, zou de verspreiding van misleidende inhoud kunnen afnemen. Meer onderzoek hiernaar is nodig.	Onderzoek laat gemengde resultaten zien van het effect van community notes op de verspreiding van informatie op Twitter (tegenwoordig X).
Bereik aanpassen op basis van inschatting betrouwbaarheid account (Fernández et al., 2021; Stray et al., 2024)	In de periode voorafgaand en tijdens verkiezingen zouden platformen de gewichten van hun aanbevelingsalgoritme kunnen aanpassen, op basis van intern onderzoek naar accounts. Accounts met een lage betrouwbaarheidsscore zouden minder aanbevolen kunnen worden.	Nieuwe accounts en accounts die in de periode vlak voor een verkiezing actief worden, zouden minder aanbevolen kunnen worden in verkiezingstijd, of pas na een menselijke check op het type content dat wordt gedeeld.
Gebruikers informeren na interactie met inhoud (Truong et al., 2024)	Platformen kunnen gebruikers die veel engagement aangaan met accounts die lage kwaliteit of verwijderde inhoud delen waarschuwen en helpen misleidende inhoud te herkennen.	

Tabel 1: suggesties uit onderzoek voor mogelijke aanpassingen aan aanbevelingsalgoritmen

1.4 Polariserende en antidemocratische *gevolgen* aanpakken

Wanneer het kabinet het *gevolg* van antidemocratische en polariserende aanbevelingsalgoritmen op individuele gebruikers wil verbieden, moet duidelijk zijn welke gevolgen er allemaal kunnen zijn.

Wetenschappers hebben moeite met het leggen van causale verbanden tussen de werking van aanbevelingsalgoritmen en veranderingen in denkbelden van gebruikers. Dat betekent dat causale uitspraken zoals: “verandering X van aanbevelingsalgoritme Y leidt tot hogere polarisatie onder doelgroep Z” heel moeilijk te bewijzen zijn, al neemt het bewijs langzaam maar zeker toe.⁶ Dat komt door de beperkte informatiepositie van onafhankelijke onderzoekers: alleen platformen hebben de data om deze causale verbanden te kunnen leggen. Wetenschappers wijzen op de noodzaak voor *randomized controlled experiments* en *holdout experiments* om meer te weten te komen over de gevolgen van aanbevelingsalgoritmen. Dat betekent dat platformen de uitrol van nieuwe mogelijkheden, of aanpassingen aan hun aanbevelingsalgoritmen, binnen een selecte groep testen om verschillen te kunnen meten met mensen die niet aan het experiment meedoen.

Het is niet zo dat er helemaal geen informatie is over de gevolgen van aanbevelingsalgoritmen gebaseerd op engagement. Zo lieten onderzoekers in 2025 zien dat het aanbevelingsalgoritme van Twitter (nu X) tweets aanraade die ervoor zorgden dat gebruikers zich slechter gingen voelen over mensen met een andere politieke voorkeur. Dit kan geïnterpreteerd worden als een polariserend gevolg van het aanbevelingsalgoritme (Milli et al., 2025).⁷ Dit ging om Amerikaanse gebruikers van het platform. Of dit ook geldt voor Nederlanders dient onderzocht te worden binnen de Nederlandse context.

Om een goed beeld te krijgen van de gevolgen van aanbevelingsalgoritmen zouden socialemediaplatformen meer inzicht moeten geven in de volgende zaken:

- Inzicht in *realtime* steekproeven, bijvoorbeeld van openbare inhoud die het meest wordt verspreid en inhoud die het meeste engagement publiceert
- Inzicht in het aantal mensen en de typen groepen waarbij reeds verwijderde inhoud terecht is gekomen
- Inzicht in technische specificaties van de architectuur van aanbevelingsalgoritmen
- Resultaten van interne A/B-testen van platformen
- Inzicht in platforminspanningen en de effecten daarvan in verkiezingstijd
- Standaardisatie van toegang en rapportages van platformen

Met deze inzichten zouden publieke organisaties en onafhankelijke onderzoekers meer weten over welke veranderingen in het ontwerp van aanbevelingsalgoritmen bij kunnen dragen aan bijvoorbeeld veranderende politieke attitudes of polarisatie.

Het socialmedialandschap is beperkt tot een handjevol platformen. Dit zorgt ervoor dat mensen online steeds minder keuzemogelijkheden hebben in de wijze waarop ze zich kunnen uiten, connecties willen aangaan of informatie willen achterhalen.⁸ De keuzes van enkele bedrijven die

⁶ Tot deze conclusie komt het Commissariaat van de Media in een recent rapport over socialemediafeeds in *Naar democratisch gezonde feeds*, 18 mei 2026.

⁷ Milli e.a., *Engagement, User Satisfaction, and the Amplification of Divisive Content on Social Media*, 27 februari 2025.

⁸ Rathenau Instituut, *Achter de macht van big tech*, 9 juli 2025.

online omgevingen beheren zijn daardoor invloedrijk.⁹ Als die beperkte groep ook nog eens vergelijkbare, op engagement gebaseerde aanbevelingssystemen hanteren, bestaan er geen alternatieven. Een pluriform medialandschap is daarom nodig om tegenwicht te bieden en alternatieve aanbevelingsalgoritmen te ontwikkelen.

Handelingsopties

1. Het is allereerst van belang dat het **bestaande wetgevingskader beschermd blijft én wordt aangegrepen**. Met bestaande wetgeving kan verantwoording van platformen worden afgedwongen. Denk daarbij aan de DSA, waarmee platformen ter verantwoording geroepen kunnen worden om de verspreiding van illegale inhoud te stoppen. Ook kunnen platformen worden gedwongen toegang te geven aan toezichthouders ter bevordering van transparantie over maatregelen om de integriteit van verkiezingen te beschermen.
2. Om de informatiepositie van toezichthouders en onafhankelijke onderzoekers te verbeteren kan de Kamer **aandringen op informatie van de platformen om van de interventies in tabel 1 aan te geven of deze opties zijn overwogen en getest** en wat eventueel hun afwegingen zijn geweest om de aanpassingen niet door te voeren.
3. Om de macht van platformen over aanbevelingsalgoritmen te verkleinen en de autonomie van gebruikers te vergroten, kan de Kamer de minister vragen zich in **Europees verband hard te maken voor interoperabiliteitsverplichtingen** voor sociale media. De Digital Fairness Act die momenteel in ontwikkeling is, biedt daar aanknopingspunten voor. Interoperabiliteit kan een markt creëren voor alternatieve aanbieders van aanbevelingsalgoritmen op bestaande platformen (zogenaamde *middlewares*).
4. Om ontwikkeling van alternatieve aanbevelingsalgoritmen te stimuleren kan de Kamer de minister vragen om platformen **te verplichten tot het aanbieden van andere typen aanbevelingsalgoritmen**. Tabel 1 biedt hiervoor inspiratie. De DFA biedt aanknopingspunten.
5. Om de afhankelijkheid van grote socialemediabedrijven af te bouwen is het noodzakelijk om alternatieve online omgevingen op te bouwen. Dit kan door een **pluriform, decentraal en interoperabel medialandschap van online omgevingen te stimuleren**. Bij decentrale omgevingen komt de controle niet meer bij één partij te liggen, maar is de controle verspreid over verschillende onafhankelijke servers. Dat is belangrijk voor de autonomie en controle van gebruikers. Met interoperabiliteit ontstaat ruimte voor nieuwe platformen en wordt het veel gemakkelijker voor gebruikers om te kiezen voor platformen die aan hun wensen voldoen, zonder volgers en contacten te verliezen. Daarnaast kunnen overheden het goede voorbeeld geven door **zelf gebruik te maken van decentrale online omgevingen** en zo de drempel voor overstap voor anderen verlagen. Ook kan de overheid publieke en private organisaties aanmoedigen om deze standaarden te gebruiken en onderzoek naar nieuwe decentrale technologieën ondersteunen.¹⁰

⁹ Rathenau Instituut, [Inclusief online](#). Naar een ontwerp van apps en omgevingen waar mensen vrij en veilig kunnen zijn, 28 augustus 2025.

¹⁰ Rathenau Instituut, [Inclusief online](#). Naar een ontwerp van apps en omgevingen waar mensen vrij en veilig kunnen zijn, 2025.

Referenties

Commissariaat voor de Media [Naar democratisch gezonde feeds](#). CvdM, 18 mei 2026.

Rathenau Instituut. [Scrollend naar de stembus](#). De rol van aanbevelingsalgoritmen bij inmenging in verkiezingen. Den Haag. Auteurs: Ex, L., Van Eeden, Q., Nieuwenhuizen, W., & Van Huijstee, M., m.m.v. Gerritsen, J., 11 maart 2026.

D66, VVD en CDA. [Coalitieakkoord-2026-2030](#), 30 januari 2026.

Rathenau Instituut. [Inclusief online](#). Naar een ontwerp van apps en omgevingen waar mensen vrij en veilig kunnen zijn. Den Haag. Auteurs: Nieuwenhuizen, W., Nieuwenhuis, T., Eeden Q. van, Huijstee, M. van), 28 augustus 2025

Rathenau Instituut. [Achter de macht van big tech](#). Van digitale afhankelijkheid naar digitale autonomie. Den Haag. Auteurs: Eeden, Q. van., B. Karstens, S. Roolvink en Huijstee, M. van), 9 juli 2025.

Milli, Smitha, Micah Carroll, Yike Wang, Sashrika Pandey, Sebastian Zhao, en Anca D Dragan. [Engagement, User Satisfaction, and the Amplification of Divisive Content on Social Media](#). Onder redactie van Jay Van Bavel. *PNAS Nexus* 4, nr. 3 pgaf062. <https://doi.org/10.1093/pnasnexus/pgaf062>. 27 februari 2025
