Rathenau Instituut

Digitale dreigingen voor de democratie

Over nieuwe technologie en desinformatie

Rapport

Auteurs
Pieter van Boheemen, Geert Munnichs en Elma Dujso

Foto omslag
Shutterstock

Bij voorkeur citeren als:
Rathenau Instituut (2020). *Digitale dreigingen voor de democratie – Over nieuwe technologie en desinformatie*. Den Haag (auteurs: Boheemen, P. van, G. Munnichs & E. Dujso)

Voorwoord

Tot voor kort kon Nederland zich koesteren met de gedachte dat desinformatie de afgelopen jaren geen grote impact op de samenleving heeft gehad. De hausse aan misleidende berichten die verspreid zijn rond de uitbraak van het coronavirus laat evenwel zien dat ook de Nederlandse samenleving hiervoor niet ongevoelig is. Tegelijkertijd is het nog te vroeg om een definitief oordeel te vellen over de betekenis hiervan voor de weerbaarheid van de Nederlandse samenleving tegen desinformatie.

De snelle technologische ontwikkelingen op het gebied van IT zouden het beeld echter op afzienbare termijn kunnen doen kantelen. Dit rapport, dat we schreven op verzoek van het ministerie van Binnenlandse Zaken en Koninkrijksrelaties en dat past bij het thema Digitale Samenleving uit ons Werkprogramma, geeft een breed overzicht van de technologische ontwikkelingen die de komende jaren een rol kunnen gaan spelen bij de productie en verspreiding van desinformatie. Dat overzicht stelt bepaald niet gerust. De mogelijkheden die technologieën als tekstsynthese, *voice cloning*, *deepfakes*, micro-targeting en chatbots aan producenten en verspreiders van desinformatie bieden om internetgebruikers te misleiden, zijn groot en veelvormig.

Het ziet er dan ook naar uit dat er meer maatregelen zullen moeten worden genomen om de dreigingen voor het publieke debat en het democratisch proces die uitgaan van door technologie gedreven desinformatie, het hoofd te bieden. Die maatregelen kunnen bijvoorbeeld bestaan uit een betere detectie van *deepfakes*, monitoring door platformbedrijven van eventueel misbruik door adverteerders van de mogelijkheden van micro-targeting en een betere facilitering van factcheckers.

Voor veel van de in dit rapport geopperde, nieuwe maatregelen ligt de verantwoordelijkheid primair bij de platformbedrijven. Een belangrijke vraag voor de komende jaren zal dan ook zijn welke maatschappelijke verantwoordelijkheid zij bereid zijn op zich te nemen bij de bestrijding van desinformatie. Schieten zij daarin tekort, dan kan de overheid vanwege het publieke belang dat gemoeid is met het tegengaan van de schadelijke effecten van desinformatie, ertoe besluiten platformbedrijven strenger te gaan reguleren. Voor beide opties biedt dit rapport handreikingen.

Dr. ir. Melanie Peters
Directeur Rathenau Instituut

Samenvatting

Inleiding
Het ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK) heeft het Rathenau Instituut verzocht onderzoek te doen naar de impact van technologische ontwikkelingen op de productie en verspreiding van desinformatie en naar maatregelen die kunnen worden getroffen om de mogelijke negatieve effecten daarvan te beperken. Het gaat hierbij vooral om desinformatie die gericht is op verstoring van het publieke debat en het democratisch proces. Het onderzoek vloeit voort uit de actielijnen die minister Ollongren in haar brief van 18 oktober 2019 aan de Tweede Kamer aankondigde, als onderdeel van de beleidsinzet van de regering ter bescherming van de Nederlandse samenleving tegen desinformatie.

In het onderzoek staan de volgende vragen centraal:

- Wat is de impact van technologische ontwikkelingen op de productie en verspreiding van desinformatie?
- Welke maatregelen worden reeds genomen om de bedreigingen die daarvan uitgaan voor het publieke debat en het democratisch proces, te beperken?
- Welke nieuwe maatregelen kunnen worden ontwikkeld om die bedreigingen het hoofd te bieden, met inachtneming van de vrijheid van meningsuiting en de persvrijheid?
- Welke actoren hebben daarbij een rol?

Aanpak
Voor het onderzoek is literatuurstudie verricht, zijn interviews gehouden met deskundigen en zijn twee casestudies uitgevoerd. De casestudies betreffen belangrijke technologische ontwikkelingen op het gebied van de productie en verspreiding van desinformatie: *deepfakes* en *psychographing*. De tussenresultaten van het onderzoek zijn besproken in een online bijeenkomst met deskundigen. Dit rapport beschrijft de resultaten van het onderzoek.

De onderzochte technologische ontwikkelingen hebben alle een digitale component. De ontwikkelingen betreffen zowel huidige als in de komende jaren te verwachten ontwikkelingen. Daarbij is een tijdshorizon van vijf jaar gehanteerd. Geen van de in dit onderzoek beschreven technologieën kan als ‘geheel nieuw’ worden beschouwd. Wel brengen we in beeld hoe technologische innovaties die reeds in ontwikkeling zijn of zich beginnen af te tekenen, verder vorm kunnen krijgen, en welke impact deze innovaties kunnen hebben op de productie en verspreiding van desinformatie.

Desinformatie
In dit onderzoek sluiten we aan bij de omschrijving van desinformatie die de minister van Binnenlandse Zaken en Koninkrijksrelaties in haar eerder genoemde brief geeft: ‘het doelbewust, veelal heimelijk, verspreiden van misleidende informatie, met het doel om schade toe te brengen aan het publieke debat, democratische processen, de open economie of nationale veiligheid’. Daarbij plaatsen we de kanttekening dat dit onderzoek zich primair richt op desinformatie die leidt tot schade aan of verstoring van het publieke debat en het democratisch proces. Daarbij kan bijvoorbeeld worden gedacht aan het aanwakkeren van maatschappelijke tegenstellingen of het voeden van wantrouwen in politieke instituties.

Uit eerder onderzoek blijkt dat in Nederland vooralsnog geen grote impact van desinformatie op de samenleving zichtbaar is. De meeste voorbeelden van desinformatie in dit onderzoek zijn dan ook afkomstig uit andere landen. Die voorbeelden bieden ook zicht op wat Nederland de komende jaren mogelijk te wachten staat op het gebied van desinformatie.

Het onderzoek bestaat uit drie delen, met elk een eigen karakter: een quickscan met een overzicht van technologische ontwikkelingen; twee verdiepende casestudies; en een vooruitblik met mogelijk te nemen, nieuwe maatregelen.

Deel I: Quickscan
De quickscan geeft een overzicht van technologische ontwikkelingen die de komende jaren een rol kunnen gaan spelen bij de productie en verspreiding van desinformatie. Ook wordt een beknopt overzicht gegeven van bestaande maatregelen die de negatieve effecten van desinformatie moeten tegengaan. In de quickscan maken we onderscheid tussen algemene technologieën, productietechnologieën en verspreidingstechnologieën.

Algemene technologieën
- Databasetechnologie: op grote schaal verzamelen en analyseren van (persoons)gegevens;
- Kunstmatige intelligentie: zelflerende algoritmes en systemen.

Technologieën waarmee desinformatie kan worden geproduceerd
- Tekstsynthese: algoritmes die leesbare en logische (nieuws)berichten genereren;
- *Voice cloning*: manipulatie van spraakberichten met behulp van kunstmatige intelligentie;
- Beeldsynthese en *deepfakes*: genereren en aanpassen van video’s met behulp van kunstmatige intelligentie;

- *Augmented* en virtual reality en avatars: presenteren van informatie in een virtuele omgeving;
- Memes: afbeeldingen ontworpen om op grote schaal te worden gedeeld op sociale media.

Technologieën waarmee desinformatie kan worden verspreid
- Socialemediaplatforms: onlineplatforms zoals Facebook, Twitter en TikTok, waarbij aanbevelingsalgoritmes berichten selecteren;
- Micro-targeting: specifieke doelgroepen met een op hen afgestemde boodschap bereiken (met behulp van campagnesoftware, *dynamic prospecting*, *programatic advertising*, *psychographing* en *influencer marketing*);
- Chatapps: berichten (versleuteld) uitwisselen, een-op-een of in kleine groepen; Bots: (deels) automatisch aangestuurde accounts op sociale media;
- Zoekmachines: platforms die het internet doorzoekbaar maken;
- Spraakassistenten: spraakgestuurde apparaten waarmee onder andere zoekmachines worden geraadpleegd;
- *Distributed autonomous applications*: online platforms zonder centrale aansturing;
- Games: online spellen;
- Crossmediale storytelling: bereiken van een specifieke persoon of doelgroep via diverse kanalen en apparaten.

Deel II: Casestudies
Voortbouwend op de quickscan zijn twee casestudies uitgewerkt waarmee een meer samenhangend beeld wordt geschetst van hoe technologische ontwikkelingen op het gebied van desinformatie de komende jaren zouden kunnen uitpakken en welke impact ze op het publieke debat en het democratisch proces kunnen hebben. De casestudies gaan over *deepfakes* en *psychographing*.

Deepfakes
Kunstmatige intelligentie kan worden ingezet voor de bewerking van audiovisueel materiaal. Daarmee kunnen gemanipuleerde video's – *deepfakes* – worden gemaakt die voor mensen lastig van echt zijn te onderscheiden. Zo kunnen met *face swaps* gezichten worden verwisseld of kan met *digital puppetry* een kunstmatig hoofd of lichaam worden gegenereerd. Met *deepfakes* kan bijvoorbeeld de indruk worden gewekt dat een bepaalde persoon een bepaalde uitspraak heeft gedaan, wat het publieke debat kan schaden.

Het valt te verwachten dat *deepfakes* door verdere technologische innovatie steeds moeilijker te onderscheiden zullen zijn van authentieke, niet-gemanipuleerde beelden. Ook zullen steeds geavanceerdere *deepfake*-technieken in eenvoudig te gebruiken apps en gadgets op de markt worden gebracht. Het gebruik van

deepfakes zal dan ook steeds normaler worden. Gezien het groeiende belang van beeld op internet, kan dat de geloofwaardigheid aantasten van beeldmateriaal dat afkomstig is van gevestigde nieuwsmedia.

In reactie op de toenemende mogelijkheden van platformbedrijven om *deepfakes* te detecteren, kunnen producenten en verspreiders van *deepfakes* uitwijken naar niet-gemodereerde, besloten kanalen.

Psychographing
Psychographing is een geavanceerde vorm van micro-targeting. Het is een advertentietechnologie die kan worden ingezet om berichten geautomatiseerd af te stemmen op de persoonlijkheidskenmerken van een doelgroep. De gedachte erachter is dat mensen kunnen worden beïnvloed door hen informatie voor te schotelen die is afgestemd op hun psychologische kenmerken. Internetgebruikers zouden hiermee op grote schaal kunnen worden misleid of gemanipuleerd.

De casestudie schetst een scenario waarin een groepering zich ten doel stelt om het publieke debat met behulp van *psychographing* te beïnvloeden. Door in te spelen op maatschappelijk gevoelige kwesties beoogt de groepering maatschappelijke tegenstellingen aan te wakkeren en het vertrouwen van burgers in gevestigde instituties te ondermijnen. Om maximaal onrust te veroorzaken, kunnen de berichten worden verspreid via niet-publieke kanalen, zoals privégroepen op Facebook of Telegram. Daar is de kans immers klein dat de berichten worden tegengesproken, wat het effect van de desinformatiecampagne vergroot.

Deel III: Vooruitblik
In de vooruitblik beschrijven we welke nieuwe maatregelen kunnen worden genomen om de belangrijkste door technologie gedreven dreigingen voor het publieke debat en het democratisch proces tegen te gaan.

Maatregelen tegen deepfakes
Investeren in detectie van deepfakes
Platformbedrijven kunnen investeren in een actief detectiebeleid gericht op de bestrijding van *deepfakes*. Dat is nodig om mee te kunnen komen in de wedloop die mogelijk ontstaat tussen producenten en verspreiders van steeds geavanceerdere *deepfakes* aan de ene kant en de detectie daarvan door platformbedrijven aan de andere kant.

Instellen van een meldpunt tegen kwaadaardige beeldmanipulatie
Platformbedrijven als SnapChat, Instagram en TikTok, waar *deepfakes* steeds normaler worden, kunnen een meldpunt instellen waar gebruikers (vermoedelijk) kwaadaardige beeldmanipulatie kunnen rapporteren.

Waarmerken van beeldmateriaal en overige berichten
Het digitaal waarmerken van beeldmateriaal en overige berichten stelt internetgebruikers in staat na te gaan of materiaal afkomstig is van een in hun ogen betrouwbare informatiebron. Dat vereist een betrouwbaar systeem om digitale waarmerken te registreren. De overheid en de grote technologiebedrijven kunnen hierin vooropgaan.

Inperken mogelijkheden van micro-targeting
Monitoren van gebruik advertentietechnologie
Platformbedrijven kunnen monitoringsmogelijkheden inbouwen in hun diensten om misbruik van door hen geleverde advertentietechnologie tegen te gaan.

Technische mogelijkheden advertentietechnologie inperken
Platformbedrijven kunnen adverteerders restricties opleggen bij hun doelgroepselectie, en monitoren op een verantwoord gebruik van door hen aangeboden advertentietechnologie.

Internetgebruikers transparantie bieden
Platformbedrijven kunnen internetgebruikers meer inzicht geven in het gebruik dat adverteerders maken van advertentieprofielen.

Maatregelen tegen schadelijke effecten aanbevelingsalgoritmes
Inbouwen reflectiemoment in platformdiensten
Aanbevelingsalgoritmes van platformbedrijven versterken veelal de sociale en politieke voorkeuren van gebruikers en – in het verlengde daarvan – maatschappelijke tegenstellingen. Om de schadelijke effecten daarvan tegen te gaan, kunnen platformbedrijven een reflectiemoment inbouwen in het gebruik van hun diensten. Gebruikers worden hierdoor gestimuleerd om (des)informatie minder impulsief te delen.

Transparantie bieden over aanbevelingsalgoritmes
Om de schadelijke effecten van aanbevelingsalgoritmes tegen te gaan, kunnen platformbedrijven transparantie bieden over de werking van de algoritmes. Om te beginnen door wetenschappelijke onderzoekers daar toegang tot te geven.

Waarschuwingssysteem voor besloten en versleutelde kanalen
Om verspreiding van desinformatie op besloten en versleutelde kanalen tegen te gaan, kan een onafhankelijk waarschuwingssysteem worden ingesteld dat desinformatiecampagnes rondom gevoelige maatschappelijke kwesties signaleert, en daarvoor waarschuwt. De overheid en de platformbedrijven kunnen dit waarschuwingssysteem faciliteren.

Verdienmodel platformbedrijven kritisch tegen het licht houden
Maatregelen als het beperken van het gebruik van advertentietechnologie of het bieden van transparantie over de werking van aanbevelingsalgoritmes, kunnen op gespannen voet staan met het verdienmodel van platformbedrijven. Die kunnen dan ook weinig genegen zijn om deze maatregelen te treffen. In dat geval kan de overheid overgaan tot verdergaande maatregelen, zoals het afdwingen van meer transparantie over het gebruik van aanbevelingsalgoritmes, of het kritisch tegen het licht houden van het verdienmodel van de platformbedrijven.

Investeren in factchecken blijft van belang
Omdat factchecken van belang is om houvast te bieden aan internetgebruikers die op zoek zijn naar betrouwbare informatie, kunnen de overheid en platformbedrijven (blijven) investeren in faciliteiten voor factcheckers.

Investeren in mediawijsheid blijft van belang
Technologische maatregelen en strengere regulering van platformbedrijven kunnen de productie en verspreiding van desinformatie terugdringen. Maar er zullen altijd vrijplaatsen blijven bestaan op het internet, waardoor internetgebruikers geconfronteerd zullen blijven worden met desinformatie. De overheid moet dan ook blijven investeren in mediawijsheid.

Slotsom: platformbedrijven primair verantwoordelijk, maar overheid kan ingrijpen
Voor veel van de genoemde maatregelen ligt de verantwoordelijkheid voor de bestrijding van desinformatie primair bij de platformbedrijven. Maar gezien het publieke belang dat gemoeid is bij het tegengaan van de schadelijke effecten van desinformatie, kan de overheid ertoe besluiten op te treden als platformbedrijven hun verantwoordelijkheid onvoldoende nemen. De overheid kan bijvoorbeeld aandringen op een actief detectiebeleid gericht op het tegengaan van *deepfakes*, of op monitoring van een onverantwoord gebruik door adverteerders van de mogelijkheden van micro-targeting.

En als aandringen niet helpt, zouden maatregelen kunnen worden afgedwongen. Die maatregelen kunnen ook ten koste gaan van het verdienmodel van platformbedrijven. Of de overheid daartoe moet overgaan, zal mede afhangen van

de ernst van de dreigingen voor het publieke debat en het democratisch proces die uitgaan van bijvoorbeeld de polariserende werking van aanbevelingsalgoritmes of door platformbedrijven gefaciliteerde desinformatiecampagnes van adverteerders. Om voldoende gewicht in de schaal te leggen, ligt het voor de hand om dwingende maatregelen binnen EU-verband te nemen.

Inhoud

Inleiding

1.1 Aanleiding

Politieke beïnvloedingscampagnes door Russische trollen, de bemoeienis van het politieke consultancybedrijf Cambridge Analytica bij het Brexit-referendum of verwarrende berichten over het ontstaan en de bestrijding van het coronavirus, leiden tot groeiende zorgen over de politieke en maatschappelijke effecten van desinformatie. Nu bestaan politieke propaganda en misleiding al langer. Zo was het in de zeventiende eeuw in de Nederlandse politiek niet ongebruikelijk om vanuit politiek gewin pamfletten te verspreiden met suggestieve beschuldigingen aan het adres van politieke tegenstanders.[1] Maar met de opkomst van de informatiemaatschappij heeft de manier waarop onware of misleidende informatie wordt ontwikkeld en verspreid nieuwe en veel grootschaligere vormen aangenomen, waar we bovendien dagelijks mee te maken hebben. Dat leidt ook tot nieuwe vragen.[2]

Dit onderzoek gaat dieper in op de aard en verspreiding van desinformatie in het huidige tijdsgewricht, hoe nieuwe technologieën daarop van invloed zijn, en welke maatregelen kunnen worden genomen om de negatieve effecten ervan tegen te gaan. Het Rathenau Instituut heeft dit onderzoek gedaan op verzoek van het ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK). Het onderzoek vloeit voort uit de actielijnen die minister Ollongren in haar brief van 18 oktober 2019 aan de Tweede Kamer aankondigde, als onderdeel van de beleidsinzet van de regering ter bescherming van de Nederlandse samenleving tegen desinformatie.[3] De regering toont zich daarmee bewust van het grote belang dat technologische ontwikkelingen hebben voor de wijze waarop allerlei vormen van desinformatie kunnen worden geproduceerd en verspreid, inclusief de vraag of de Nederlandse samenleving daarop voldoende is voorbereid.

---

[1] Haverkate, J.M.M. (2019). Spindoctors van de Gouden Eeuw: De eerste pamfletoorlog van Overijssel (1654-1675). https://research.vu.nl/en/publications/spindoctors-van-de-gouden-eeuw-de-eerste-pamfletoorlog-van-overij.

[2] Balaban, D. (2018). News Sharing on Social Media Platforms. Theoretical Approaches. *Communication. Strategic Perspectives*. www.academia.edu/38666829/News_Sharing_on_Social_Media_Platforms._Theoretical_Approaches

[3] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2019). *Kamerbrief over beleidsinzet bescherming democratie tegen desinformatie*. www.rijksoverheid.nl/documenten/kamerstukken/2019/10/18/kamerbrief-over-beleidsinzet-bescherming-democratie-tegen-desinformatie

1.2 Doel- en vraagstelling

Het doel van dit onderzoek is na te gaan welke impact technologische ontwikkelingen kunnen hebben op de productie en verspreiding van desinformatie en welke maatregelen kunnen worden getroffen om de mogelijke negatieve effecten daarvan te beperken. Het gaat hierbij vooral om desinformatie die gericht is op verstoring van het publieke debat en het democratisch proces.

In het onderzoek staan de volgende vragen centraal:

- Wat is de impact van technologische ontwikkelingen op de productie en verspreiding van desinformatie?
- Welke maatregelen worden reeds genomen om de bedreigingen die daarvan uitgaan voor het publieke debat en het democratisch proces, te beperken?
- Welke nieuwe maatregelen kunnen worden ontwikkeld om die bedreigingen het hoofd te bieden, met inachtneming van de vrijheid van meningsuiting en de persvrijheid?
- Welke actoren hebben daarbij een rol?

1.3 Aanpak

Voor het onderzoek is literatuurstudie verricht, zijn interviews gehouden met deskundigen en zijn op basis van de resultaten daarvan twee casestudies uitgevoerd waarin dieper op de materie is ingegaan. De opbrengst van de literatuurstudie, de interviews en de casestudies is vervolgens besproken in een online expertmeeting met deskundigen. De resultaten van deze activiteiten zijn verwerkt in dit eindrapport. De onderstaande tabel biedt een overzicht van de ondernomen onderzoeksactiviteiten.

Onderzoeksactiviteit	Functie
Literatuurscan	Overzicht van relevante technologische ontwikkelingen, hun betekenis voor desinformatie en mogelijke maatregelen om bedreigingen daarvan voor het publieke debat en het democratisch proces tegen te gaan.
Interviews	Aanvulling literatuurscan met nieuwe inzichten over technologische ontwikkelingen, hun betekenis voor desinformatie en mogelijke maatregelen.
Quickscan	Synthese van de opbrengst van de literatuurscan en de interviews. De quickscan geldt als tussenrapportage van het onderzoek.

Casestudies	Verdieping van het inzicht in enkele belangrijke technologische ontwikkelingen en daaraan gerelateerde maatregelen.
Expertmeeting	Verdere verkenning van mogelijke maatregelen, aan de hand van de quickscan en de casestudies.
Eindrapport	Synthese van de opbrengst van de diverse onderzoeksactiviteiten.

Tabel 1 Overzicht onderzoeksactiviteiten

Hieronder volgt een verdere toelichting op de aanpak, waaronder de afbakening van het onderzoek, de literatuurscan en de interviews.

1.3.1 Afbakening

De twee centrale begrippen van dit onderzoek, ‘technologische ontwikkelingen’ en ‘desinformatie’, zijn beide brede onderwerpen. Om tot een werkbare onderzoeksopzet te komen, zijn we tot de volgende afbakening gekomen.

Onder technologische ontwikkelingen die van invloed zijn op de productie en verspreiding van desinformatie, worden zowel huidige als in de komende jaren te verwachten ontwikkelingen begrepen. Vanwege de snelle technologische ontwikkelingen op dit gebied is hierbij een tijdshorizon van circa vijf jaar gehanteerd. Voor het ontwikkelen van een handelingsperspectief op een zich zo snel ontwikkelend gebied – met uitspraken over mogelijk te nemen maatregelen –, heeft het in onze ogen weinig zin om verder dan vijf jaar vooruit te kijken.

De technologische ontwikkelingen die in het kader van desinformatie worden onderzocht hebben alle een digitale component. Die component kan betrekking hebben op de productie van desinformatie, de verspreiding ervan, of allebei. Het gaat hierbij bijvoorbeeld om de mogelijkheden die kunstmatige intelligentie en online platformbedrijven bieden voor de productie en verspreiding van desinformatie.

Geen van de in dit onderzoek beschreven technologieën kan als ‘geheel nieuw’ – in de zin van: nu nog onbekend – worden aangemerkt. Een beschrijving daarvan zou sciencefiction worden. Wel brengen we in beeld hoe technologische innovaties die reeds in ontwikkeling zijn of zich beginnen af te tekenen, verder vorm kunnen krijgen, en welke impact deze innovaties kunnen hebben op de productie en verspreiding van desinformatie. Die innovaties schuilen bijvoorbeeld in de ontwikkeling van kwalitatief steeds betere toepassingen – zoals bij *deepfakes* het

geval is – of in de mogelijkheden die *psychographing* of *influencer marketing* bieden om geavanceerdere vormen van micro-targeting mogelijk te maken.

In hoofdstuk 2 wordt ingegaan op de omschrijving van het door ons gebruikte begrip van desinformatie.

1.3.2 Literatuurscan

Op basis van een literatuurscan is een overzicht gemaakt van relevante technologische ontwikkelingen en hun betekenis voor desinformatie, en een eerste inventarisatie van mogelijke beleidsopties. Hierbij hebben we gebruik gemaakt van primaire en secundaire wetenschappelijke bronnen. Relevante referenties in deze bronnen zijn geraadpleegd via online databases.

In de wetenschappelijke databases Scopus, ISI Web of Science, Google Scholar, IEEE Explore en SSRN is gezocht naar recent gepubliceerde wetenschappelijke literatuur. Daarvoor zijn de volgende zoektermen gebruikt: ‘disinformation’ (402 resultaten), ‘strategic communications’ (53 resultaten), ‘micro-targeting’ (28 resultaten), ‘deepfake’ (16 resultaten), ‘post-truth’ (711 resultaten) en ‘online-harm’ (4 resultaten). Vervolgens zijn die artikelen geselecteerd die op het eerste oog het meeste inzicht boden in het fenomeen desinformatie, daarmee samenhangende technologische ontwikkelingen en daarop betrekking hebbende maatregelen.

Verder zijn relevante publicaties geraadpleegd van nationale organisaties (AIVD, NCTV, ministerie van BZK, ROB, CPB), Europese instellingen (EC, STOA), mediaonderzoeksbureaus (Reuters, PEW Research), online platformbedrijven (Facebook, Twitter, Google) en relevante maatschappelijke organisaties (Bits of Freedom, AlgorithmWatch, The Intercept, Electronic Frontier Foundation).

Ten slotte is ook gekeken naar nieuwsbronnen die door professionals in het veld worden gevolgd (iBestuur, Emerce.nl, Security.nl, Reddit en MrDeepfakes forum).

1.3.3 Interviews

De opbrengst van de literatuurscan is aangevuld met de resultaten van interviews met deskundigen op het gebied van desinformatie en aanpalende expertisegebieden. In de interviews is vooral ingegaan op de in de ogen van de geïnterviewden meest relevante ontwikkelingen op het gebied van desinformatie en de belangrijkste maatregelen die daartegen kunnen worden genomen. De interviews zijn afgenomen op basis van een vooraf toegestuurde vragenlijst (zie

Bijlage 1) en verliepen semigestructureerd. In Bijlage 2 zijn de namen van de geïnterviewde personen vermeld.

1.3.4 Quickscan, casestudies en expertmeeting

De resultaten van de literatuurscan en de interviews zijn verwerkt in de quickscan en in twee casestudies. De casestudies betreffen belangrijke en impactvolle technologische ontwikkelingen op het gebied van de productie en verspreiding van desinformatie: *deepfakes* en *psychographing.*

De quickscan en de casestudies zijn besproken tijdens een online expertmeeting. In deze bijeenkomst lag de nadruk op mogelijke maatregelen voor het tegengaan van bedreigingen voor het publieke debat en het democratisch proces, die uitgaan van door technologie gedreven desinformatie. In Bijlage 3 zijn de namen van de deelnemers aan de expertmeeting vermeld.

De selectie van de deelnemers aan de interviews en de expertmeeting en de selectie van de casestudies zijn besproken met het ministerie van Binnenlandse Zaken en Koninkrijksrelaties. Conceptversies van de quickscan en het eindrapport zijn besproken met een interdepartementale klankbordgroep die door het ministerie voor dit doel was ingesteld. Als onafhankelijk onderzoeksinstituut heeft het Rathenau Instituut de resultaten hiervan naar eigen inzicht gebruikt. De verantwoordelijkheid voor de inhoud van dit rapport ligt dan ook volledig bij het Rathenau Instituut.

1.4 Leeswijzer

Hoofdstuk 2 beschrijft wat in dit onderzoek onder desinformatie wordt verstaan en in welke mate desinformatie in Nederland voorkomt.

De overige hoofdstukken beschrijven de opbrengst van het onderzoek. Ze zijn onderverdeeld in drie delen, die ongelijksoortig van aard zijn.

Deel I bevat de resultaten van de quickscan. Ze geeft een breed overzicht van technologische ontwikkelingen die de komende jaren een rol kunnen gaan spelen bij de productie en verspreiding van desinformatie (hoofdstuk 3 tot en met 5). Daarnaast biedt de quickscan een beknopt overzicht van bestaande maatregelen die worden genomen om de dreigingen die van desinformatie uitgaan voor het publieke debat en het democratisch proces, het hoofd te bieden (hoofdstuk 6).

Deel II bevat twee casestudies waarmee een meer samenhangend beeld wordt geschetst van hoe technologische ontwikkelingen op het gebied van desinformatie de komende jaren zouden kunnen uitpakken en welke impact ze op het publieke debat en het democratisch proces kunnen hebben. De casestudies gaan over *deepfakes* en *psychographing* (hoofdstuk 7).

Deel III bevat een vooruitblik, waarin wordt beschreven welke nieuwe maatregelen kunnen worden genomen om schade aan het publieke debat en het democratisch proces als gevolg van technologie gedreven desinformatie tegen te gaan, en welke actoren daarvoor verantwoordelijk zijn (hoofdstuk 8). Daarnaast bevat het een afsluitend hoofdstuk, waarin de belangrijkste bevindingen van het onderzoek worden opgesomd (hoofdstuk 9).

2 Desinformatie

Om de impact van technologische ontwikkelingen op de productie en verspreiding van desinformatie te kunnen beschrijven, is het nodig duidelijk te maken wat we in dit onderzoek onder desinformatie verstaan. Daarover bestaan immers diverse opvattingen. Verder beschrijven we wat bekend is over de mate waarin desinformatie in Nederland speelt, en hoe zich dat verhoudt tot andere landen.

2.1 Definitie desinformatie

Van de definitie van desinformatie bestaan diverse beschrijvingen. Veelal wordt met de term desinformatie gedoeld op het verspreiden van 'onware', 'inaccurate' of 'misleidende' informatie, maar die termen zijn vaak lastig te hanteren. Zo is het in de praktijk vaak lastig om een onderscheid te maken tussen het verschaffen van informatie van lage kwaliteit en het verspreiden van leugens.[4] [5] Het is duidelijk dat de context waarbinnen (des)informatie wordt verspreid en het doel waarmee dat gebeurt, mede bepalen of iets als desinformatie kan worden gekenmerkt.

In de loop der jaren hebben zich in de wetenschappelijke literatuur verschuivingen voorgedaan in de omschrijving van desinformatie. Zo omschrijven Schultz en Godson desinformatie als 'false, incomplete or misleading information that is passed, fed, or confirmed to a targeted individual, group, or country'.[6] Zij richten zich daarbij op de inhoud van de informatie en de beoogde doelgroep.

Meer recente definities van desinformatie richten zich ook op de intentie van de degene die desinformatie produceert of verspreidt. Iets kan pas als desinformatie worden aangemerkt als hierbij een kwaadwillende partij betrokken is, die doelbewust handelt.[7] Zo definieert het CPB desinformatie als 'het bewust creëren en verspreiden van onware, inaccurate of misleidende informatie'.[8] Het lastige aan

[4] Wardle, C., & Derakhshan, H. (2017). *INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making*. https://rm.coe.int/information-disorder-report-november-2017/1680764666
[5] RAND (2019). *What's Being Done to Fight Disinformation Online*. www.rand.org/research/projects/truth-decay/fighting-disinformation.html
[6] Shultz, R. H., & Godson, R. (1984). *Dezinformatsia: Active Measures in Soviet Strategy* (1st edition). University of Nebraska Press.
[7] Gelfert, A. (2018). Fake News: A Definition. *Informal Logic*, *38*(1), 84–117. https://doi.org/10.22329/il.v38i1.5068 en Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2), 122–139. https://doi.org/10.1177/0267323118760317
[8] CPB (2019). *Risicorapportage cyberveiligheid economie 2019*. CPB. www.cpb.nl/sites/default/files/omnidownload/cpb-notitie-risicorapportage-cyberveiligheid-2019.pdf

deze omschrijving is dat het in de praktijk niet altijd duidelijk is welke intentie toe te schrijven is aan degene die desinformatie produceert of verspreidt, en of deze daarmee doelbewust handelt.

Humprecht stelt dan ook voor om een meer specifieke doelstelling aan de definitie van desinformatie toe te voegen: de degene die desinformatie produceert of verspreidt zou daarmee uit moeten zijn op het berokkenen van schade of het behalen van winst of sociale invloed.[9]

In haar brief aan de Tweede Kamer van 18 oktober 2019 lijkt minister Ollongren aan te sluiten bij de door Humprecht voorgestelde omschrijving, waarbij ze de intentie van degene die desinformatie produceert of verspreidt beperkt tot het berokkenen van schade. Zij omschrijft desinformatie als 'het doelbewust, veelal heimelijk, verspreiden van misleidende informatie, met het doel om schade toe te brengen aan het publieke debat, democratische processen, de open economie of nationale veiligheid'.[10] Deze definitie ligt sterk in lijn met die van de Europese Commissie en het Britse Lagerhuis.[11] [12]

In haar meer recente brief aan de Tweede Kamer van 13 mei 2020 verduidelijkt de minister wat ze verstaat onder doelbewuste verspreiding. Ze stelt namelijk dat misleidende informatie ook kan worden verspreid door mensen zonder dat zij bewust schade willen berokkenen.[13] Dit niet-doelbewust verspreiden van misleidende informatie kan worden betiteld als misinformatie.[14]

Het begrip desinformatie wordt aan de andere kant begrensd door bij wet verboden uitingen als laster, haatzaaien of oproepen tot geweld. Dergelijke uitingen kunnen, anders dan misinformatie, juridisch worden vervolgd.

Dit betekent ook dat productie en verspreiding van desinformatie als zodanig niet verboden is. De overheid kan dan ook niet zonder aanvullende redenen overgaan

---

[9] Humprecht, E. (2018). Where 'fake news' flourishes : a comparison across four Western democracies. *Information, Communication and Society*, *21*, 1–16. https://doi.org/10.1080/1369118X.2018.1474241
[10] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2019). *Kamerbrief over beleidsinzet bescherming democratie tegen desinformatie*. www.rijksoverheid.nl/documenten/kamerstukken/2019/10/18/kamerbrief-over-beleidsinzet-bescherming-democratie-tegen-desinformatie
[11] EC DG CONNECT HLEG on Fake News (2018). *A multi-dimensional approach to disinformation*
[12] DCMSC of the House of Commons (2019). *Disinformation and 'fake news' Report* www.parliament.uk/business/committees/committees-a-z/commons-select/digital-culture-media-and-sport-committee/news/fake-news-report-published-17-19/
[13] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2020). *Kamerbrief ontwikkelingen beleidsinzet bescherming democratie tegen desinformatie.* www.rijksoverheid.nl/documenten/kamerstukken/2020/05/13/kamerbrief-ontwikkelingen-beleidsinzet-bescherming-democratie-tegen-desinformatie
[14] Humprecht, E., Esser, F., & Van Aelst, P. (2020). Resilience to Online Disinformation: A Framework for Cross-National Comparative Research. The International Journal of Press/Politics, 1940161219900126. https://doi.org/10.1177/1940161219900126

tot het verwijderen van misleidende informatie. Dat zou in strijd zijn met de vrijheid van meningsuiting.[15] [16] Platformbedrijven kunnen daarentegen wel overgaan tot het verwijderen van misleidende informatie, als verspreiding van desinformatie in strijd is met door hen gestelde gebruikersvoorwaarden.

Schade aan publiek debat en democratisch proces

In dit onderzoek sluiten we aan bij de door minister Ollongren geformuleerde omschrijving van desinformatie. Daarbij plaatsen we wel de kanttekening dat dit onderzoek zich primair richt op de productie en verspreiding van desinformatie die leidt tot schade aan of verstoring van het publieke debat en het democratisch proces. Voor deze schade of verstoring kan bijvoorbeeld worden gedacht aan het aanwakkeren van maatschappelijke polarisering, het voeden van wantrouwen in politieke instituties of het heimelijk beïnvloeden van de politieke meningsvorming van burgers. Zo stelt de Staatscommissie parlementair stelstel (commissie-Remkes) dat heimelijke beïnvloeding van politieke meningsvorming in strijd is met de uitgangspunten van vrije en gelijke verkiezingen.[17]

We willen hieraan toevoegen dat we het publieke debat beschouwen als een kernelement van de democratische rechtsorde. Dit debat dient ertoe om politieke voorkeuren en opvattingen publiekelijk te kunnen uiten en weerspreken. Het publieke debat moet het mogelijk maken dat kiezers hun politieke voorkeuren en opvattingen kunnen vormen en desgewenst kunnen bijstellen of herbezien, mede in het licht van wat door anderen in het debat naar voren wordt gebracht.[18] Inbreuken op deze democratische kernfunctie van het publiek debat beschouwen we als een verstoring daarvan.

Tegen deze achtergrond kan ook duidelijk worden gemaakt waarom bijvoorbeeld heimelijke beïnvloeding van de politieke meningsvorming van burgers met behulp van micro-targeting – iets waarvan Cambridge Analytica tijdens de Brexit-campagne in het Verenigd Koninkrijk werd beschuldigd – problematisch is. Doordat hierbij gebruik werd gemaakt van op specifieke doelgroepen toegesneden politieke boodschappen die voor anderen niet inzichtelijk waren, konden deze politieke boodschappen niet door andere kiezers of politieke groeperingen worden

---

[15] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2019). *Actielijnen tegengaan desinformatie.* www.rijksoverheid.nl/documenten/kamerstukken/2019/10/18/actielijnen-tegengaan-desinformatie

[16] IViR (2019). *Het juridisch kader voor de verspreiding van desinformatie via internetdiensten en de regulering van politieke advertenties*. Universiteit van Amsterdam. www.ivir.nl/publicaties/download/Rapport_desinformatie_december2019.pdf

[17] Staatscommissie parlementair stelsel (2018). *Lage drempels, hoge dijken: Democratie en rechtsstaat in balans.* www.staatscommissieparlementairstelsel.nl/documenten/rapporten/samenvattingen/12/13/eindrapport

[18] Munnichs, G.M. (2000). *Publiek ongenoegen en politieke geloofwaardigheid: democratische legitimiteit in een ontzuilde samenleving.* https://www.rug.nl/research/portal/nl/publications/publiek-ongenoegen-en-politieke-geloofwaardigheid(fffc2cec-2962-4114-b655-9118961af83c).html

weersproken. Hierdoor was het ook niet mogelijk om na te gaan in hoeverre de diverse politieke boodschappen met elkaar verenigbaar waren.

2.2 Betrokken groeperingen

Bij de productie en verspreiding van desinformatie kunnen uiteenlopende actoren betrokken zijn. Vaak worden hierbij de volgende groeperingen onderscheiden:[19]

- Statelijke actoren en daaraan gelieerde groepen;
- Extremistische groeperingen;
- Economisch gedreven actoren, zoals de Macedonische jongeren die actief waren tijdens de Amerikaanse presidentsverkiezingen in 2016;[20]
- Professionele marketingorganisaties, zoals het politieke consultancybedrijf Cambridge Analytica;
- Socialemediaplatforms.

De beweegredenen van de betrokken groeperingen om desinformatie te verspreiden, kunnen sterk verschillen. Statelijke actoren en daaraan gelieerde groepen zijn met hun desinformatiecampagnes vaak uit op het stichten van maatschappelijke onrust, door verwarring te zaaien, maatschappelijke verdeeldheid te creëren of de berichtgeving van gevestigde instituties in twijfel te trekken – vaak zonder duidelijke politieke agenda. Desinformatie die door bijvoorbeeld rechts-extremistische groeperingen wordt verspreid, heeft vaak wel een duidelijke politieke agenda. Het tegenovergestelde is echter ook mogelijk. Desinformatie kan namelijk ook worden verspreid vanuit louter economische motieven, zoals bij de Macedonische jongeren het geval was. Door berichten te verspreiden die gericht zijn op het trekken van zoveel mogelijk aandacht op sociale media, kunnen afzenders hun brood verdienen – waarbij de inhoud van de berichtgeving verder bijzaak is.

De betrokken groeperingen gaan hierbij vaak opportunistisch te werk. Ze buiten kwetsbaarheden in de samenleving uit, haken aan bij zich voordoende gelegenheden zoals oplaaiende discussies in de media, en maken gebruik van die technische middelen die het meeste effect sorteren.

Het is lang niet altijd mogelijk om te achterhalen wie verantwoordelijk is voor de productie of verspreiding van desinformatie, en om welke reden. De afzender kan

[19] Benkler, Y., Faris, R., & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford University Press. www.oxfordscholarship.com/view/10.1093/oso/9780190923624.001.0001/oso-9780190923624

[20] Subramanian, S. (2017). *Meet the Macedonian Teens Who Mastered Fake News and Corrupted the US Election.* Wired www.wired.com/2017/02/veles-macedonia-fake-news/

er namelijk belang bij hebben om niet te worden herkend, en bijvoorbeeld een valse identiteit aannemen. Zo onthulden Facebook en Twitter recentelijk dat Russische actoren betrokken zijn bij organisaties in Ghana en Nigeria die zich voordoen als Amerikaans, en zich mengen in debatten over politiek gevoelige onderwerpen.[21]

2.3 Kernelementen

In de hiernavolgende hoofdstukken maken we gebruik van diverse kernelementen met behulp waarvan we de impact van technologische ontwikkelingen op de productie en verspreiding van desinformatie beschrijven. Zie daarvoor onderstaande tabel.

Element	Toelichting
Afzender	Een of meerdere actoren die verantwoordelijk zijn voor de productie en/of verspreiding van desinformatie.
Intentie	De beweegreden van de afzender om desinformatie te produceren of verspreiden.
Inhoud	De boodschap die wordt overgebracht, die de ontvanger op andere gedachten moet brengen of tot bepaald gedrag moet aanzetten.
Vorm	De vorm waarin desinformatie wordt geuit, bijvoorbeeld in audio of video.
Medium	De manier waarop de desinformatie wordt overgebracht, en de ontvanger bereikt, bijvoorbeeld via een platform.
Ontvanger	De persoon of groep die de desinformatie ontvangt.
Effect	De (beoogde) verandering in gedachten of gedrag van de ontvanger.

Tabel 2 Kernelementen

2.4 Desinformatie in Nederland

Uit de beschikbare literatuur komt naar voren dat er in Nederland vooralsnog weinig desinformatie voorkomt. Uit eerder onderzoek van het Rathenau Instituut (2018), bleek dat in Nederland vooralsnog geen grote impact van desinformatie op de

[21] Culliford, E. (2020). Facebook, Twitter remove Russia-linked accounts in Ghana targeting U.S. *Reuters*. www.reuters.com/article/us-facebook-content-idUSKBN20Z3LW

samenleving zichtbaar was.[22] In Nederland verspreide desinformatie bleek vooral afkomstig van economisch gedreven actoren, die vaak met behulp van 'pulpnieuws' of 'click-bait' (klik-aas) mensen naar advertentiesites proberen te lokken. Slechts een beperkt deel daarvan heeft een politiek karakter of betreft maatschappelijk gevoelige onderwerpen. Onderzoek van het Oxford Internet Institute bevestigt dit beeld.[23]

De AIVD constateert in haar jaarverslag van 2019 wel pogingen tot verspreiding van desinformatie door Russische groeperingen, maar de impact daarvan lijkt vooralsnog beperkt.[24] Dit laatste wil zeggen dat ze weinig online interactie (likes en shares) teweeg wisten te brengen en dat verspreide verhaallijnen nauwelijks werden overgenomen. En uit recent onderzoek van de Universiteit van Amsterdam komt naar voren dat desinformatie tijdens de provinciale en Europese verkiezingen van 2019 geen rol van betekenis heeft gespeeld.[25]

Veel literatuur over desinformatie heeft betrekking op de situatie in andere landen, zoals de Verenigde Staten en het Verenigd Koninkrijk. Maar omdat Engelstalige berichten ook het Nederlandse publiek kunnen bereiken, kunnen Engelstalige desinformatiecampagnes wel indirect hun weerslag hebben op het publieke debat in Nederland.

Nederlandse burgers maken zich in ieder geval wel zorgen over desinformatie. Volgens een onderzoek van de Volkskant ziet 82% van de Nederlanders desinformatie als een bedreiging voor het functioneren van de democratie en de rechtsstaat.[26] [27] Volgens het CPB zijn de zorgen over de gevolgen van desinformatie voor de publieke opinie de afgelopen jaren toegenomen. Het CPB wijst daarbij op buitenlandse voorbeelden van desinformatie en op Russische trollen die op Twitter actief waren na het neerschieten van vlucht MH17.[28]

---

[22] Van Keulen, I., Korthagen, I., Diederen, P., & Van Boheemen, P. (2018). *Digitalisering van het nieuws*. Rathenau Instituut.
[23] Blood, D. (2017). *Is social media empowering Dutch populism?* Oxford Internet Institute. www.ft.com/content/b1830ac2-07f4-11e7-97d1-5e720a26771b
[24] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2020). *AIVD-jaarverslag 2019*. www.aivd.nl/documenten/jaarverslagen/2020/04/29/jaarverslag-2019
[25] Rogers, R., & Niederer, S. (2019). *Politiek en Sociale Media Manipulatie*. Universiteit van Amsterdam. www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2019/10/18/rapport-politiek-en-sociale-media-manipulatie/rapport-politiek-en-sociale-media-manipulatie.pdf
[26] Kranenberg, A. (2017). *Nederlanders bezorgd over 'nepnieuws' - een op drie weet vaak niet meer wat waar is en wat onwaar.* Volkskrant www.volkskrant.nl/nieuws-achtergrond/nederlanders-bezorgd-over-nepnieuws-een-op-drie-weet-vaak-niet-meer-wat-waar-is-en-wat-onwaar~b6914596/
[27] Kranenberg, A. (2017). *Wie weet nog wat er waar is?* Volkskrant www.volkskrant.nl/kijkverder/2017/desinformatie/
[28] CPB (2019). *Risicorapportage cyberveiligheid economie 2019*. CPB. www.cpb.nl/sites/default/files/omnidownload/cpb-notitie-risicorapportage-cyberveiligheid-2019.pdf

Desinformatie tijdens de coronapandemie
Een mogelijke uitzondering op het hierboven geschetste beeld is van recente datum, en betreft het voor Nederlandse begrippen grote aantal misleidende berichten en complottheorieën rond de uitbraak van het coronavirus. Dat doet zich overigens wereldwijd voor. Al bij het uitroepen van de pandemie bracht de Wereldgezondheidsorganisatie (WHO) ook de dreigingen die uitgaan van desinformatie onder de aandacht. Tedros Adhanom Ghebreyesus, directeur-generaal van de WHO, sprak van een *infodemie*: 'We're not just fighting an epidemic; we're fighting an infodemic'.[29]

In Nederland heeft de coronacrisis een breed scala aan desinformatie en valse geruchten losgemaakt, van falsificaties van berichtgeving van de NOS en de Rijksoverheid tot wilde complottheorieën.[30] Zo doen misleidende berichten over coronamedicijnen en waarschuwingen tegen bepaalde middelen de ronde op sociale media en in chats.[31]

Maar het is nog te vroeg om een goed oordeel te kunnen vellen over de mogelijke impact van de hausse aan coronagerelateerde desinformatie voor de manier waarop in Nederland het publieke en politieke debat wordt gevoerd. Zo is het lang niet altijd duidelijk of er sprake is van kwade opzet, en of er niet veeleer sprake is van misinformatie dan van desinformatie.

2.5 Internationale ontwikkelingen

Omdat de meeste voorbeelden van desinformatie afkomstig zijn uit andere landen, kijken we in dit onderzoek ook naar wat er in het buitenland gebeurt. Daarmee krijgen we ook meer zicht op wat Nederland de komende jaren mogelijk te wachten staat op het gebied van desinformatie.

Hieronder beschrijven we enkele trends die in het buitenland zijn waar te nemen. Hierbij moet worden bedacht dat de aard en impact van desinformatie van land tot land kunnen verschillen. Zo vonden onderzoekers hele andere desinformatiethema's in het Verenigd Koninkrijk dan in Duitsland.[32] Dat betekent ook dat de schade die desinformatie in andere landen teweegbrengt zich niet

[29] Adhanom, T. (2020). *Munich Security Conference*. www.who.int/dg/speeches/detail/munich-security-conference
[30] Vermanen, J., & Van Bree, T. (2020). *Flinke stijging van onbetrouwbaar nieuws over coronavirus op Twitter.* Pointer https://pointer.kro-ncrv.nl/node/280
[31] Kist, R., & Nieber, L. (2020). *Misinformatie over coronavirus gaat ook viraal.* NRC. www.nrc.nl/nieuws/2020/03/09/misinformatie-over-coronavirus-gaat-ook-viraal-a3993140
[32] Humprecht, E. (2018). Where 'fake news' flourishes : a comparison across four Western democracies. *Information, Communication and Society*, *21*, 1–16. https://doi.org/10.1080/1369118X.2018.1474241

noodzakelijkerwijs op dezelfde manier of met hetzelfde effect in Nederland voor zal doen.

Desinformatie is een groeiend internationaal fenomeen
Onderzoek van de Universiteit van Oxford laat zien dat desinformatie door de jaren heen in steeds meer landen voorkomt. In 2018 werden in 48 landen aanwijzingen gevonden voor georganiseerde desinformatiecampagnes, een forse stijging ten opzichte van de 28 landen in het jaar daarvoor.[33]

Desinformatie vindt plaats op grote schaal
Voorbeelden uit het buitenland maken duidelijk dat desinformatiecampagnes op grote schaal plaatsvinden.[34] Het bekendste voorbeeld hiervan zijn de desinformatiecampagnes die zijn uitgevoerd in 2016 tijdens de presidentsverkiezingen in de Verenigde Staten. Onderzoek door de Amerikaanse overheid naar Russische inmenging in de verkiezingen laat zien dat duizenden door Russische groeperingen aangestuurde accounts meer dan een miljoen tweets, honderdduizenden Facebookberichten en duizend YouTube-video's produceerden en verspreidden.[35] De tweets werden 288 miljoen keer bekeken, de Facebookberichten 126 miljoen keer. [36] Vanwege deze gigantische aantallen wordt de Russische Internet Research Agency (IRA), die hiervoor verantwoordelijk wordt gehouden, ook wel een trollenfabriek genoemd.[37]

De activiteiten van de IRA beperken zich niet tot Facebook, Twitter en YouTube, maar worden ook op Google+, Vine, Meetup, Pinterest, Tumblr, Gab, Medium en Reddit waargenomen. Tussen 2014 en 2017 wist IRA op Instagram 187 miljoen interacties te ontlokken en op Facebook meer dan 76 miljoen.[38] Het bovenstaande maakt ook duidelijk dat afzenders van desinformatie zich niet alleen richten op de grote platforms, maar ook gebruikmaken van de mogelijkheden die kleinere platforms bieden.

---

[33] Bradshaw, S., & Howard, P. (2018). *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation*. University of Oxford. http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf
[34] Nemr, C., & Gangware, W. (2019). *WEAPONS OF MASS DISTRACTION: Foreign State-Sponsored Disinformation in the Digital Age*. Park Advisors. www.state.gov/wp-content/uploads/2019/05/Weapons-of-Mass-Distraction-Foreign-State-Sponsored-Disinformation-in-the-Digital-Age.pdf
[35] US Office of the Director of National Intelligence (2017). *'Background to "Assessing Russian Activities and Intentions in Recent US Elections," The Analytic Process and Cyber Incident Attribution'*. www.dni.gov/files/documents/ICA_2017_01.pdf.
[36] US House of Representatives (2018). *Exposing Russia's Effort to Sow Discord Online: The Internet Research Agency and Advertisements* https://intelligence.house.gov/social-media-content
[37] Linvill, D. L., & Warren, P. L. (2020). Troll Factories: Manufacturing Specialized Disinformation on Twitter. *Political Communication*, *0*(0), 1–21. https://doi.org/10.1080/10584609.2020.1718257
[38] DiResta et al. (2018). *'The Tactics and Tropes of the Internet Research Agency,' en Philip N. Howard et al. (2018). 'The IRA, Social Media and Political Polarization in the United States, 2012-2018'*. https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/12/IRA-Report-2018.pdf

Zoals hierboven al is genoemd, kan desinformatie ook voortvloeien uit economisch gewin. Menig producent of verspreider van desinformatie heeft een economisch motief. Het verdienmodel is hierbij gebaseerd op het vertonen van advertenties in of bij desinformatie. Volgens de website DesinformationIndex vertegenwoordigen de producenten en verspreiders van desinformatie een totale wereldwijde marktwaarde van 235 miljoen dollar.[39] Dit aanzienlijke bedrag vormt ook een van de verklaringen voor het vermogen van afzenders van desinformatie om innovatieve technologieën in te zetten.

Professionele dienstverlening

Een andere noemenswaardige buitenlandse ontwikkeling is de toenemende professionalisering van de productie en verspreiding van desinformatie. Desinformatiecampagnes kunnen tegenwoordig worden aangeschaft op websites (*underground forums*) waarop professionele partijen hun diensten aanbieden. Recorded Future bracht deze markt in kaart. Het stuitte op uitgebreide prijslijsten, waarop bijvoorbeeld het plaatsen van een authentiek-ogend artikel in de Financial Times, met een inhoudelijke strekking naar keuze, voor ongeveer 50.000 dollar wordt aangeboden.[40]

Afzenders van desinformatie hoeven dan ook niet zelf over de benodigde kennis te beschikken om van bepaalde technologische mogelijkheden gebruik te maken. Professionele dienstverleners vergemakkelijken het gebruik van nieuwe technologieën.

[39] Disinformation Index (z.d.) https://disinformationindex.org/research/
[40] Insikt Group (2019). *The Price of Influence: Disinformation in the Private Sector.* www.recordedfuture.com/disinformation-service-campaigns/

Deel I Quickscan

3 Algemene technologieën

De quickscan geeft een breed overzicht van technologische ontwikkelingen die de komende jaren een rol kunnen gaan spelen bij de productie en verspreiding van desinformatie. Daarnaast biedt het een beknopt overzicht van bestaande maatregelen die worden genomen om de dreigingen die van desinformatie uitgaan voor het publieke debat en het democratisch proces, het hoofd te bieden.

De term quickscan geeft al aan dat het een verkenning betreft, die in relatief korte tijd is uitgevoerd. Een gevolg hiervan is dat maar beperkt op de diverse ontwikkelingen en hun onderlinge samenhang is ingegaan. De casestudies in Deel II bieden een meer uitvoerige beschrijving van enkele belangrijke technologische ontwikkelingen en hun mogelijke impact op de productie en verspreiding van desinformatie.

De quickscan bestaat uit vier hoofdstukken. Dit hoofdstuk (hoofdstuk 3) beschrijft twee algemene technologieën die vaak aan de basis liggen van de erna besproken technologieën, die worden gebruikt voor de productie (hoofdstuk 4) en de verspreiding (hoofdstuk 5) van desinformatie. Hoofdstuk 6 beschrijft de bestaande maatregelen die worden genomen om de schadelijke effecten van desinformatie tegen te gaan. Bijlage 4 geeft een overzicht van de in de quickscan besproken technologieën.

Zoals reeds in het inleidend hoofdstuk is vermeld, moet worden bedacht dat veel van de hier besproken technologieën nog sterk in ontwikkeling zijn. Dat wil ook zeggen dat in de komende jaren hun betekenis voor de productie en verspreiding van desinformatie anders kan uitpakken dan in dit onderzoek is voorzien.

De twee in dit hoofdstuk besproken, algemene technologieën zijn:

- databasetechnologie
- kunstmatige intelligentie.

3.1 Databasetechnologie

Technologieën die worden gebruikt voor de productie en verspreiding van desinformatie maken steeds vaker gebruik van databasetechnologie. Databasetechnologie maakt het mogelijk om grote hoeveelheden informatie te verzamelen en te analyseren. De gegevens in databases – kortweg data genoemd

– vormen als het ware de grondstof waaruit desinformatie wordt gecreëerd. Het kunnen beschikken over grote hoeveelheden data en de technologie om daaruit informatie te winnen, wint dan ook aan belang.

Data worden op steeds grotere schaal verzameld, een ontwikkeling die de komende jaren alleen maar verder zal doorzetten. Zo volgen online-adverteerders met behulp van cookietechnologie en trackingcodes voortdurend het online gedrag – en daarmee: de voorkeuren – van internetgebruikers, van de websites die ze bezoeken tot de tijd die ze besteden om door een webpagina te scrollen. Google scant mails en privé-chats om informatie te vergaren waarmee internetgebruikers gepersonaliseerde advertenties kunnen worden aangeboden.

In de nabije toekomst wordt het steeds beter mogelijk om data uit verschillende databases en bronnen aan elkaar te koppelen. Producenten en verspreiders van desinformatie kunnen daarmee steeds beter inschatten welke boodschap en welke vormgeving van die boodschap het beste aansluit bij de opvattingen en behoeften van de ontvanger. Dergelijke vormen van micro-targeting (zie hoofdstuk 5) kunnen dan ook worden ingezet om op specifieke persoonskenmerken afgestemde desinformatie te verspreiden.

De data die voor de productie van desinformatie relevant zijn, hoeven niet noodzakelijkerwijs op een legale manier te zijn verkregen. Data kunnen afkomstig zijn uit gehackte databases, per ongeluk gelekte databases of publiek toegankelijke (open) data. Zo zijn Amerikaanse wetenschappers erin geslaagd om op basis van kenmerken uit de openbare databases van Google Streetview in te schatten op welke partij er in een bepaalde wijk gestemd wordt. Deze informatie kan worden gebruikt voor de productie en verspreiding van op politieke campagnes gerichte desinformatie (zie figuur 1).[41]

[41] Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences*, *114*(50), 13108–13113. https://doi.org/10.1073/pnas.1700035114

Figuur 1 Op basis van kenmerken uit openbare databases, zoals Google Streetview, kan informatie worden afgeleid die nuttig is voor de productie van desinformatie.

3.2 Kunstmatige intelligentie

Een deel van de bovengenoemde database-technologieën en veel van de technologieën die we in hoofdstuk 4 en 5 bespreken, maken gebruik van kunstmatige intelligentie. Hieronder lichten we kort toe wat we daaronder verstaan.

Met kunstmatige intelligentie (KI) worden computersystemen bedoeld die een zekere mate van intelligent gedrag vertonen.[42] Er bestaan verschillende technieken om een dergelijk systeem te bouwen.[43] Een basale techniek is de zogeheten *rule-based* KI. De hiermee gebouwde computersystemen zijn geprogrammeerd met behulp van 'als dit, dan dat'-instructies. Een computer kan bijvoorbeeld een gebruiker voorstellen om updates te installeren zodra die beschikbaar zijn. Dit soort meldingen zijn inmiddels zo gewoon geworden, dat ze vaak niet meer als intelligent gedrag worden gezien.

Een meer geavanceerde vorm van KI is *machine learning*. Systemen die hiervan gebruikmaken hebben vooraf ingestelde instructies, maar zijn ook in staat om instructies af te leiden uit data. Het systeem analyseert bestaande data en leert daarin patronen te ontdekken, om deze vervolgens toe te passen op nieuwe data. De patroonherkenning kan voortdurend verbeteren.

Deep learning (DL) is een specifieke vorm van *machine learning*. De techniek is gebaseerd op zogenaamde neurale netwerken, waarin verschillende lagen

[42] Van Boheemen, P., Munnichs, G., Kool, L., Diercks, G., Hamer, J., & Vos, A. (2020). *Cyberweerbaar met nieuwe technologie*. Rathenau Instituut. www.rathenau.nl/nl/digitale-samenleving/cyberweerbaar-met-nieuwe-technologie
[43] European Commission. (2019). *A definition of Artificial Intelligence: main capabilities and scientific disciplines*. https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

informatie worden gecombineerd. Zo kan een *deep learning*-systeem dat gericht is op stemherkenning uit drie lagen bestaan. Ten eerste wordt de geluidsfrequentie van een stem geanalyseerd. Deze wordt vervolgens gecombineerd met een laag die het tempo en de kadans van het stemgebruik analyseert. En een derde laag analyseert het woordgebruik. De combinatie van deze verschillende elementen maakt het mogelijk om stemmen te herkennen. De techniek van deep learning wordt ook gebruikt voor de manipulatie van audio- en videomateriaal – oftewel: de productie van *deepfakes* (zie 4.3).

Data zijn een belangrijke grondstof voor KI-systemen, omdat ze ermee worden getraind en verbeterd. Het gebruik van KI-systemen en databasetechnologie gaat dan ook vaak hand in hand.

4 Productietechnologieën

In dit hoofdstuk beschrijven we welke technologische ontwikkelingen de komende jaren naar verwachting relevant zullen zijn voor de productie van desinformatie. We bespreken de volgende technologieën:

- Tekstsynthese
- *Voice cloning*
- Beeldsynthese en *deepfakes*
- Memes
- *Augmented* en virtual reality en avatars

4.1 Tekstsynthese

Met behulp van tekstsynthese is het mogelijk om nieuwe, goed leesbare en logische teksten te genereren, met minimale of zelfs zonder menselijke aansturing.

Een voorbeeld van deze technologie is OpenAI GPT-2. Dit KI-systeem is getraind aan de hand van 8 miljoen tekstdocumenten en webpagina's. Het systeem is erop gericht om het volgende woord in een zin te voorspellen, aan de hand van de daaraan voorafgaande woorden. In tegenstelling tot andere KI-taalmodellen, die getraind zijn op een specifiek domein, is dit model niet domeinspecifiek en dus breed toepasbaar.[44] [45]

De verwachting is dat deze vorm van KI in de toekomst steeds beter in staat zal zijn om teksten te produceren die niet of moeilijk van authentieke teksten zijn te onderscheiden. Op dit moment moet een producent van desinformatie nog een zekere hoeveelheid tijd en creativiteit aanwenden om (misleidende) teksten te schrijven, maar met behulp van tekstsynthese kunnen grote hoeveelheden tekst in een mum van tijd worden gegenereerd. Een grootschalige inzet van AI-gegenereerde teksten zou ertoe kunnen leiden dat *fake* nieuwsberichten authentieke berichtgeving gaan overstemmen, leidend tot verstoring van het publieke debat.

Met behulp van tekstsynthese kunnen ook de rangschikalgoritmes van zoekmachines worden beïnvloed. Deze algoritmes kijken onder andere naar het

[44] OpenAI (2019). *GPT-2: 1.5B Release* https://openai.com/blog/gpt-2-1-5b-release/
[45] Vincent, J. (2019). *OpenAI has published the text-generating AI it said was too dangerous to share*. The Verge www.theverge.com/2019/11/7/20953040/openai-text-generation-ai-gpt-2-full-model-release-1-5b-parameters

aantal links naar een bepaald artikel. Hoe vaker een artikel wordt aangehaald, hoe hoger het in de ranking scoort. Computersystemen die gebruikmaken van tekstsynthese maken het mogelijk om op grote schaal artikelen te genereren die naar elkaar verwijzen, en daarmee de scoringssystematiek van het rankschikalgoritme te beïnvloeden. Dat maakt het voor kwaadwillende partijen mogelijk om authentiekogende, misleidende artikelen onder de publieke aandacht te brengen, en daarmee de publieke opinievorming te beïnvloeden.

4.2 Voice cloning

Kunstmatige intelligentie maakt het steeds beter mogelijk om audioberichten te manipuleren. Vooral algoritmes die gesproken berichten aanpassen zijn voor de productie van desinformatie van belang. Mensen kunnen namelijk misleid worden wanneer daarmee het stemgeluid van een persoon die zij vertrouwen succesvol wordt nagebootst.

Software als Lyrebird, Adobe Voco, CorentinJ/Real-time Voice cloning, iSpeech, Resemble, Tacotron 2, en CereVoice Me maken dit nu al mogelijk, overigens met wisselende resultaten. De software stelt gebruikers in staat om in opgenomen gesprekken aanpassingen te doen met behulp van gesynthetiseerde spraak, die ook vermengd kan worden met het omgevingsgeluid.

Voor een nauwkeurige nabootsing van iemands stem zijn korte geluidsfragmenten nodig. Met behulp van KI en ontwikkelingen op het vlak van text-to-speech-synthese kunnen onderzoekers nu al een bijna perfecte stemkloon maken op basis van een geluidsopname van iemands stem van enkele seconden.[46] Aangezien steeds meer mensen fragmenten van hun stemgeluid delen, bijvoorbeeld in video's op sociale media, kunnen steeds meer mensen het doelwit worden van deze vorm van misleiding.

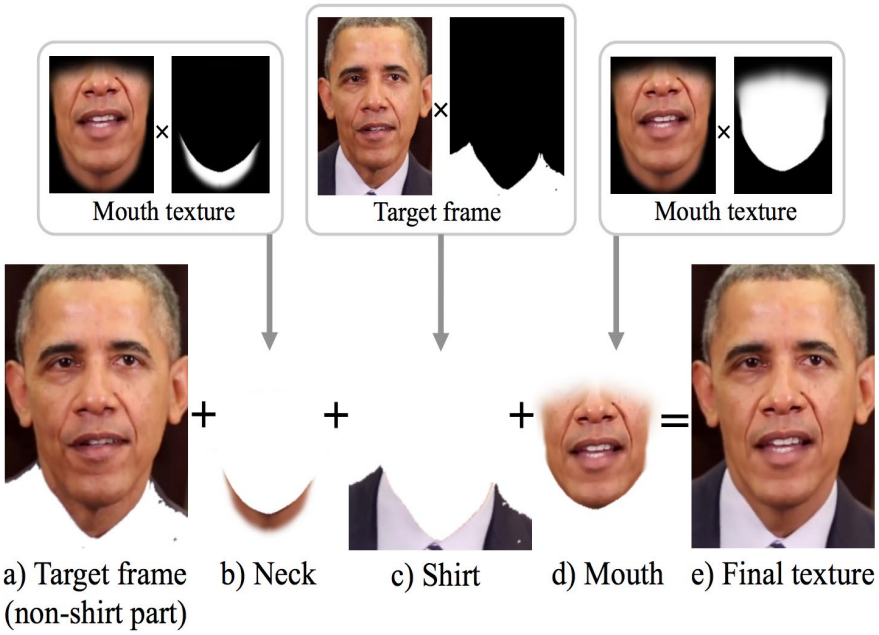
Voice cloning-technologie leidt nu al tot (economische) schade. Het wordt ingezet door criminelen voor zogeheten CEO-fraude: het aanzetten van financieel medewerkers tot het overmaken van geld door het stemgeluid van een leidinggevende na te bootsen.[47]

[46] FTC (2019). *You Don't Say: An FTC Workshop on Voice Cloning Technologies.* www.ftc.gov/news-events/events-calendar/you-dont-say-ftc-workshop-voice-cloning-technologies
[47] Malik, D. (2020). *AI Based Voice Cloning Is Giving Rise To Another Big Security Scam.* www.digitalinformationworld.com/2020/03/ai-based-voice-cloning-is-giving-rise-to-another-big-security-scam.html

4.3 Beeldsynthese en deepfakes

Beeldmanipulatie is een reeds bestaand fenomeen, zoals de bewerking van foto's met behulp van programma's als Photoshop laat zien. Nieuwe technologieën maken het mogelijk en steeds eenvoudiger om ook video's te bewerken én te genereren. Ook in deze technologieën speelt kunstmatige intelligentie een belangrijke rol.[48]

Een voorbeeld van KI-beeldsynthese zijn *deepfakes*. Een *deepfake* is een videofragment dat echt lijkt, maar gemanipuleerd is met behulp van *deep learning*-algoritmes. Hierbij wordt gebruikgemaakt van een *autoencoder*, die input kan reconstrueren, en een *generative adversarial network* (GAN). Een GAN is een computersysteem dat twee neurale netwerken combineert: het ene genereert beelden (zie figuur 2), het andere evalueert de kwaliteit daarvan.[49] [50] Voor *deepfakes* wordt bestaand beeldmateriaal als uitgangspunt gebruikt.



Figuur 2 Schematische weergave van de opbouw van een samengesteld portret.

[48] Khodabakhsh, A., Busch, C., & Ramachandra, R. (2018). A Taxonomy of Audiovisual Fake Multimedia Content Creation Technology. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 372–377). Miami, FL: IEEE. https://doi.org/10.1109/MIPR.2018.00082
[49] U.S. Government Accountability Office. (2020). *Science & Tech Spotlight: Deepfakes*, (GAO-20-379SP). www.gao.gov/products/gao-20-379sp
[50] Martineau, P. (2019). *Facebook Removes Accounts With AI-Generated Profile Photos*. *Wired*. www.wired.com/story/facebook-removes-accounts-ai-generated-photos/

Inmiddels kunnen beeldsynthesetechnieken ook op live videobeelden worden toegepast. Zo kunnen met behulp van Face2Face en HeadOn in live videobeelden gezichten, gezichtsuitdrukkingen en bewegingen worden vervangen.[51] Gezien de snelle toename in rekenkracht van smartphones en de toegenomen bandbreedte van mobiele telecomnetwerken, is het de verwachting dat *real time deepfakes* in hoeveelheid zullen toenemen.

Deepfake-technologie wordt steeds toegankelijker. Eenvoudig te gebruiken apps, zoals Doublicat, bieden al beperkte mogelijkheden om gezichten te verwisselen.[52] De software die gebruikt wordt voor geavanceerdere *deepfakes* als DeepFaceLab is ook vrij verkrijgbaar (open source). Het gebruik hiervan vergt nog wel aanzienlijke technische vaardigheden. In meer geavanceerde vormen van deze software kan het stemgeluid in de video worden omgezet naar tekst, en kan veranderde tekst vervolgens in de – aangepaste – video worden 'uitgesproken'.[53] Het is de verwachting dat in de nabije toekomst eenvoudig te gebruiken apps beschikbaar komen waarmee het voor veel mensen mogelijk wordt om *deepfakes* te produceren.

De technologie achter *deepfakes* is nog volop in ontwikkeling. Volgens het cybersecuritybedrijf Nisos zijn er op het darkweb nog geen aanbieders actief die geavanceerde manipulatie van video's als service aanbieden. Dat duidt er volgens Nisos op dat de technologie nog niet ver genoeg is ontwikkeld, en daardoor nog onvoldoende nauwkeurig werkt. De verwachting is echter dat de ontwikkelingen op dit gebied snel gaan.[54]

Omdat beeldmateriaal een steeds belangrijkere rol lijkt te gaan spelen op internet – zie bijvoorbeeld de grote populariteit van Youtube, Instagram en TikTok – is het bovendien de verwachting dat producenten van desinformatie veelvuldig gebruik zullen gaan maken van (gemanipuleerd) beeldmateriaal.

Cheap fakes

Wat de manipulatie van beeldmateriaal betreft, laten recente voorbeelden overigens zien dat het kunnen beschikken over geavanceerde technologie geen voorwaarde is voor de productie van desinformatie. Met behulp van *cheap fakes* kan ook al de nodige schade worden aangericht. Zo werd tijdens de laatste

---

[51] Thies, J., Zollhöfer, M., Theobalt, C., Stamminger, M., & Nießner, M. (2018). HeadOn: Real-time Reenactment of Human Portrait Videos. *ACM Transactions on Graphics*, *37*(4), 1–13. http://arxiv.org/abs/1610.03151

[52] Neocortext, Inc. (2020) *REFACE.* Google Play https://play.google.com/store/apps/details?id=video.reface.app&hl=en

[53] Fried, O., Tewari, A., Zollhöfer, M., Finkelstein, A., Shechtman, E., Goldman, D. B., Genova, K., Jin, Z., Theobalt, C., & Agrawala, M. (2019). Text-based Editing of Talking-head Video. *arXiv:1906.01524 [cs]*. http://arxiv.org/abs/1906.01524

[54] Volkert, R. (2020). *Deep Fakes: Understanding the illicit economy for synthetic media.* NISOS https://cdn2.hubspot.net/hubfs/6068438/Resources/NISOS%20-%20Deep%20Fakes%20White%20Paper.pdf

presidentsverkiezingen in Brazilië met *low tech*-middelen, namelijk Photoshop, de nodige sociale onrust veroorzaakt (zie figuur 3).[55]

Beeldmanipulatie leidt tot geweld in Brazilië

De afbeelding toont de zoon van de Braziliaanse presidentskandidaat Bolsonaro tijdens de verkiezingscampagne. Het linkerdeel is de bewerkte afbeelding, het rechterdeel het origineel. Op het shirt is de volgende tekst geplaatst: 'Moviment nodestinos [een bevolkingsgroep uit noordoost-Brazilië] ga terug naar jullie huis. Rio [de Janeiro] is geen plek voor ezels'. De afbeelding is wijdverspreid via sociale media, ook nadat bekend werd dat deze door middel van Photoshop was bewerkt. De afbeelding leidde tot felle discussies en zelfs vechtpartijen tussen mensen uit verschillende Braziliaanse regio's.

Figuur 3 Beeldmanipulatie leidt tot geweld in Brazilië (copyright Aos Fatos Org).

4.4 Memes

Memes zijn afbeeldingen die, al dan niet gemanipuleerd of voorzien van een tekst, een vaak humoristische of satirische boodschap proberen over te brengen. Ze zijn ontworpen om via sociale media te delen.[56] We beschouwen in dit onderzoek politieke memes als een randverschijnsel. Behalve humoristisch of satirisch bedoeld, kunnen memes ook doelbewust worden ingezet om schade te berokkenen. Maar het is vaak lastig in te schatten of een meme enkel vermaak tot doel heeft, of ook het toebrengen van schade. Tijdens de presidentsverkiezingen in de Verenigde Staten in 2016 werden ze onder andere ingezet om het stemgedrag te beïnvloeden van gebruikers van sociale media.

Memes kunnen worden ingezet voor een doelbewuste verspreiding van desinformatie. In hun versimpelde weergave van de werkelijkheid kan deze immers gemakkelijk worden vervormd of geweld aangedaan. De combinatie van humor met

[55] Panontin Scarabelli, A. (2018). *How did fake news run voters' opinions in the Brazilian elections.* Diggit Magazine www.diggitmagazine.com/articles/fake-news-brazilian-elections
[56] Rushkoff, D., Pescovitz, D., & Dunaga, J. (2018). *THE BIOLOGY OF DISINFORMATION: memes, media viruses, and cultural inoculation*. Institute for the Future. *www.iftf.org/fileadmin/user_upload/images/ourwork/digintel/IFTF_biology_of_disinformation_062718.pdf*

visualisatie kan een krachtig instrument zijn om iemands beeld van de wereld te beïnvloeden.[57]

Memes worden steeds vaker geproduceerd en verspreid op speciaal daarvoor ontworpen platforms, zoals Giphy. Platforms voor sociale media en chatapps kunnen deze meme-platforms integreren in hun dienstverlening, waardoor hun bereik sterk toeneemt.

4.5 Augmented en virtual reality en avatars

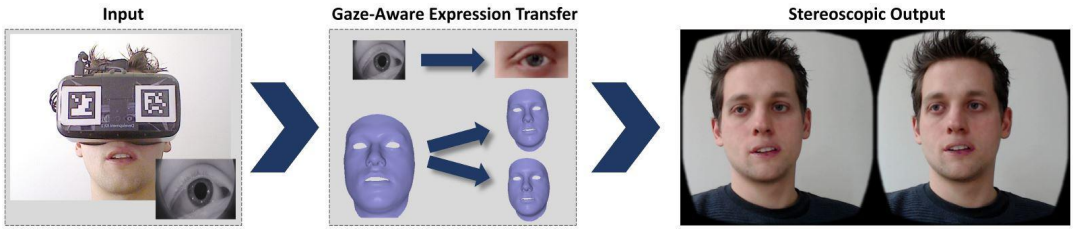
Bij toepassingen van *augmented* en virtual reality (AR/VR) wordt informatie in het gehele of een deel van het blikveld van de gebruiker gepresenteerd. Veelal wordt daarbij gebruikgemaakt van een speciale AR- of VR-bril, die een beeldscherm of projector bevat. Voor AR-toepassingen beslaat het beeld alleen een deel van het blikveld, of is de bril tevens uitgerust met een camera, waardoor een deel van de werkelijkheid zichtbaar blijft. Bij VR-toepassingen is het gehele beeld kunstmatig.

Doordat AR- en VR-technologie gebruikers letterlijk dicht op de huid zit, worden hiermee nieuwe mogelijkheden gecreëerd om het gedrag van gebruikers te analyseren. Met een AR/VR-bril kunnen bijvoorbeeld oogbewegingen van de gebruiker worden gevolgd en kan op basis van pupilreflexen worden geanalyseerd hoe een gebruiker op beelden reageert. Op basis hiervan kunnen gebruikers gerichte, op hun gedragingen en voorkeuren afgestemde informatie worden voorgeschoteld. Dat kan ook worden gebruikt voor een op persoonskenmerken gerichte verspreiding van desinformatie. Zo kunnen bijvoorbeeld raciale vooroordelen worden versterkt.[58]

VR-technologie bestaat al decennia, maar wordt (nog) niet door een groot deel van de Nederlandse bevolking gebruikt. De technologie wordt wel steeds betaalbaarder, grote technologiebedrijven ontwikkelen steeds meer applicaties en nieuwe algoritmes maken het mogelijk om met een VR-bril op het hoofd op een enigszins natuurgetrouwe manier met andere mensen te communiceren. Een voorbeeld daarvan is FaceVR. Deze technologie is een combinatie van beeldmanipulatie en VR-technologie, waarbij het gezicht van de gebruiker van een VR-bril gereconstrueerd wordt. Zodoende kunnen VR-gebruikers met elkaar videobellen,

[57] Klein, O. (2018). *Manipulative Memes : How Internet Memes Can Distort the Truth – Connected Life Conference*. Oxford Internet Institute https://connectedlife.oii.ox.ac.uk/manipulative-memes-how-internet-memes-can-distort-the-truth/
[58] Rose, J. (2016). *The Dark Side of VR*. The Intercept https://theintercept.com/2016/12/23/virtual-reality-allows-the-most-detailed-intimate-digital-surveillance-yet/

met een bril op, en toch elkaars gezicht zien (zie figuur 4).[59] Manipulatie van deze techniek met behulp van realtime *deepfake*-algoritmes ligt hier op de loer.

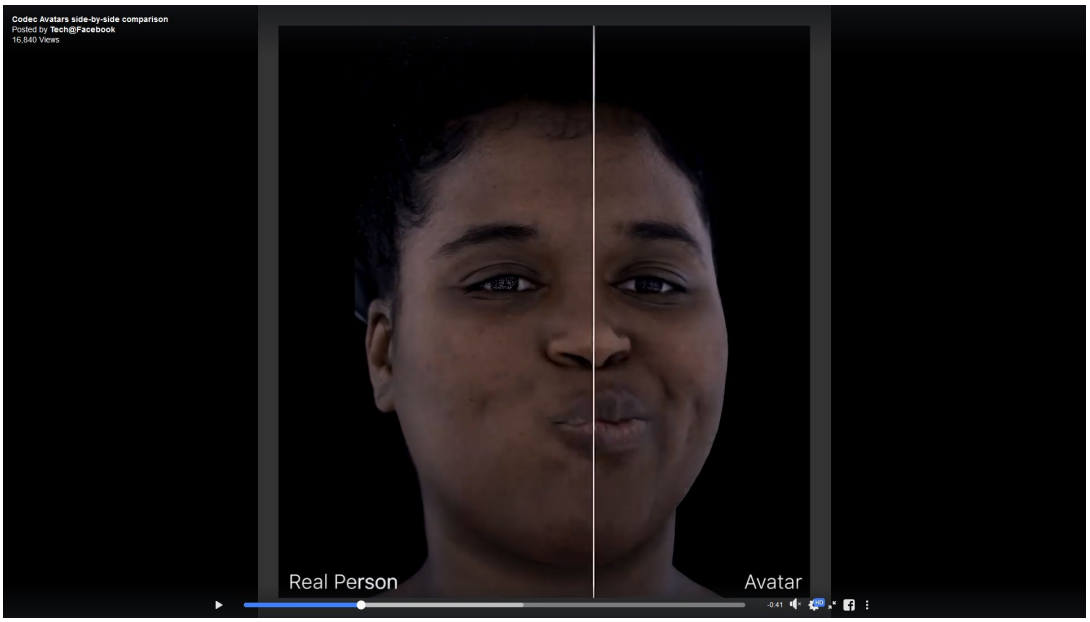


Figuur 4 FaceVR, links de gebruiker met VR-bril op, rechts de beelden die de ontvanger ziet.

Er zijn steeds meer technische mogelijkheden om binnen VR een driedimensionale lookalike-avatar van een persoon te genereren.[60] De opkomst van deze nieuwe generatie avatars wordt gedreven door de nieuwe mogelijkheden die bijvoorbeeld de *depth sensing camera* in de iPhone X, de lichtradartechniek in de nieuwste iPad[61] en *real-time face tracking* bieden.[62]

Technologiebedrijven zetten steeds sterker in op het gebruik van AR en VR. Zo is het de verwachting dat Apple binnen een jaar een VR-systeem zal lanceren.[63] En Facebook verwacht in 2020 met de nieuwe VR-omgeving Horizon te starten.[64] Facebook suggereert dat het binnen deze virtuele omgeving mogelijk is een avatar te creëren op basis van een 3D-scan van je lichaam (zie figuur 5).[65] Het hacken van iemands avatar, door hem te kopiëren of over te nemen, zou kunnen worden ingezet voor de productie en verspreiding van desinformatie.

[59] Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2018). FaceVR: Real-Time Facial Reenactment and Eye Gaze Control in Virtual Reality. *ArXiv:1610.03151 [Cs]*. http://arxiv.org/abs/1610.03151
[60] Dempsey, M. (2018). *Avatar-First Products & Platforms*. Medium. https://medium.com/@mhdempsey/avatar-first-products-platforms-723fd637bd35
[61] Apple (2020). *Apple unveils new iPad Pro with LiDAR Scanner and trackpad support in iPadOS.* www.apple.com/newsroom/2020/03/apple-unveils-new-ipad-pro-with-lidar-scanner-and-trackpad-support-in-ipados/
[62] Gibbs, S. (2020). *Apple unveils iPad Pro with 3D scanner in major redesign.* The Guardian www.theguardian.com/technology/2020/mar/18/apple-unveils-ipad-pro-with-3d-scanner-in-major-redesign
[63] Gurman, M. (2019). *Apple Plans Standalone AR and VR Gaming Headset by 2022 and Glasses Later*. Bloomberg. www.bloomberg.com/news/articles/2019-11-11/apple-s-ar-push-will-start-with-ipad-and-culminate-with-glasses
[64] Oculus (2020). *Facebook Horizon* www.oculus.com/facebookhorizon/
[65] Facebook (2019). *Facebook is building the future of connection with lifelike avatars.* https://tech.fb.com/codec-avatars-facebook-reality-labs/



Figuur 5 Facebook Codec Avatar. Met deze technologie worden zeer realistische 3D-scans van gebruikers ingezet als avatar in VR, waardoor je ‘jezelf’ kunt zijn in VR. Deze afbeelding is een screenshot van een video waarin wordt getoond hoe sterk de avatar (rechterdeel) op een echt persoon lijkt (linkerdeel).

5 Verspreidingstechnologieën

In dit hoofdstuk beschrijven we welke technologische ontwikkelingen de komende jaren naar verwachting relevant zullen zijn voor de verspreiding van desinformatie. We bespreken de volgende technologieën:

- Micro-targeting
- Chatapps
- Bots
- Zoekmachines
- Spraakassistenten
- *Distributed Autonomous Applications*
- Games
- Crossmedia storytelling

Vanwege het grote aandeel dat ze hebben in de verspreiding van desinformatie, gaan we ook in op de rol van socialemediaplatforms.

5.1 Socialemediaplatforms

Socialemediaplatforms brengen op grote schaal mensen met elkaar en met allerlei informatiebronnen in contact. In Nederland wordt volop gebruikgemaakt van een breed scala aan socialemediaplatforms. Naast populaire platforms als Facebook, YouTube, Instagram, Twitter, TikTok en LinkedIn, bestaan er allerlei kleinere platforms die op specifieke informatiebehoeften inspelen of specifieke doelgroepen bedienen, zoals Tumblr, Reddit, Flickr, Medium, Spotify en Pinterest.

Op deze platforms kunnen gebruikers allerlei vormen van informatie met anderen delen. Vaak zijn dat geschreven berichten en foto's, maar steeds vaker worden ook video's gedeeld. Het relatief nieuwe platform TikTok (voorheen Musicaly), dat zich uitsluitend richt op het verspreiden van video's, wint snel aan populariteit. Zodra een bepaald soort informatie op een platform aan populariteit wint, nemen andere platforms dit vaak in rap tempo over. Zo zijn de op SnapChat populaire Stories (tijdelijk zichtbare berichten) nu ook bijvoorbeeld op Instagram te vinden, en worden ze uitgeprobeerd op Twitter.[66] Een nieuw, populair informatietype wordt dus dikwijls snel door andere partijen gekopieerd – en daarmee beschikbaar gemaakt voor een breed publiek.

[66] Wilson, M. (2020). *Twitter is about to become an even bigger weapon of disinformation.* FastCompany www.fastcompany.com/90472066/twitters-new-self-destruct-feature-is-just-another-weapon-of-disinformation

5.1.1 Verdienmodel platforms

Socialemediaplatforms brengen de voorkeuren van gebruikers in kaart, zodat zij gebruikers kunnen voorzien van op hen afgestemde berichten en advertenties. De mate waarin platforms daarin slagen is afhankelijk van de data die ze over hun gebruikers hebben vergaard en de kwaliteit van hun aanbevelingsalgoritmes. De verkoop van advertentieruimte vormt doorgaans hun belangrijkste verdienmodel. Hoe beter een platform in staat is om adverteerders in contact te brengen met potentiële klanten, hoe groter de omzet.

Technologische ontwikkelingen zoals micro-targeting (zie 5.2) stellen socialemediaplatforms in staat om nog meer gebruikersdata te vergaren, te combineren en te analyseren, met als doel het opstellen van nog betere advertentieprofielen van gebruikers. Het groeiende gebruik van allerlei slimme, met het internet verbonden apparaten (het *Internet of Things*), stelt socialemediaplatforms bovendien in staat om data afkomstig van online surfgedrag te combineren met data van offline gedrag dat door sensoren wordt gemeten. Zo is het denkbaar dat de stand van de thermostaat thuis, invloed kan hebben op de kleding die getoond wordt in advertenties, of dat de aanwezigheid van zonnepanelen gebruikt wordt om de politieke voorkeur van een gebruiker in te schatten.[67]

5.1.2 Filterbubbels

De informatie die gebruikers van socialemediaplatforms krijgen aangeboden, wordt veelal geselecteerd door aanbevelingsalgoritmes. De aangeboden informatie wordt afgestemd op de voorkeuren van gebruikers en op analyses van overige data die over gebruikers bekend zijn. Hoe de aanbevelingsalgoritmes precies werken is niet bekend. Ze worden door de socialemediaplatforms als bedrijfsgeheim beschouwd.

Door het gebruik van aanbevelingsalgoritmes kunnen filterbubbels of echokamers ontstaan. Doordat gebruikers vooral bepaalde, op hun persoonskenmerken afgestemde informatie aangeboden krijgen, krijgen ze een beperkt beeld van de werkelijkheid voorgeschoteld – dat bovendien vaak in lijn zal liggen met hun eigen voorkeuren en opvattingen. Dat kan onschuldige vormen aannemen – zoals berichtgeving die is afgestemd op een eerder getoonde interesse in sportnieuws – maar kan ook leiden tot een beperkte, eenzijdige blik op maatschappelijke en

[67] TacticalTech. (2019). *Personal Data: Political Persuasion - The Guidebook and Visual Gallery*. https://ourdataourselves.tacticaltech.org/posts/inside-the-influence-industry

politieke kwesties. Als dat zich op grote schaal voordoet, kan dat leiden tot een verstoring van het publieke debat, omdat mensen niet langer worden geconfronteerd met afwijkende meningen en gezichtspunten.[68]

Onderzoek van het Instituut voor Informatierecht (IViR) in opdracht van het Commissariaat voor de Media, laat zien dat er vooralsnog geen aanwijzingen zijn voor het bestaan van filterbubbels in Nederland.[69] Maar het IViR signaleert wel risicofactoren die erop wijzen dat dit in de nabije toekomst zou kunnen veranderen. Het merendeel van de Nederlandse bevolking maakt nog steeds volop gebruik van een scala aan informatiebronnen, zoals televisie, radio, kranten en internet. Deze veelsoortige nieuwsconsumptie wordt gezien als een gunstige factor in de strijd tegen desinformatie. Maar er zijn inmiddels grote verschillen in mediagebruik tussen de diverse leeftijdscategorieën. Met name jongeren maken veel gebruik van sociale media, ook wat hun nieuwsconsumptie betreft. Het is de vraag wat voor effect dat zal hebben op hun omgang met misleidende berichtgeving.

5.1.3 Radicalisering en polarisering

De aanbevelingsalgoritmes van socialemediabedrijven zijn er over het algemeen op gericht de aandacht van gebruikers zo lang mogelijk vast te houden.[70] Dat leidt er doorgaans toe dat berichten, foto's of video's met een meer sensationele inhoud een hogere ranking krijgen. Het lijkt er veel op dat als gevolg hiervan aanbevelingsalgoritmes berichten met een radicaliserende of polariserende inhoud sterker onder de aandacht brengen dan andere berichten.[71]

Zo concluderen de Volkskrant en de Correspondent uit door hen verricht onderzoek, dat het aanbevelingsalgoritme van YouTube kijkers aanmoedigt om alsmaar radicalere video's te bekijken. Ook zouden extreemrechtse berichten

[68] Pariser, E. (2012). *The filter bubble: what the Internet is hiding from you*. London: Penguin Books
[69] Commissariaat voor de Media. (2019). *Filterbubbels in Nederland*. www.mediamonitor.nl/analyse-verdieping/filterbubbels-in-nederland-2019/.
[70] European Data Protection Supervisor (2018). *EDPS Opinion on online manipulation and personal data* https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf
[71] Quinn, B., Blackall, M., & Dodd, V. (2020). *YouTube accused of being 'organ of radicalisation*.' The Guardian. www.theguardian.com/technology/2020/mar/02/youtube-accused-of-being-organ-of-radicalisation

relatief oververtegenwoordigd zijn in het aanbod van YouTube.[72] [73] Onderzoek van de Universiteit van Amsterdam onderschrijft dit beeld.[74]

Kwaadwillende partijen die uit zijn op de verdere verspreiding van radicaliserende of polariserende informatie kunnen gebruikmaken van deze (veronderstelde) werking van aanbevelingsalgoritmes. Zo wijst bovengenoemd onderzoek van de Universiteit van Amsterdam op een 'opkomende tendentieuze en polariserende mediasfeer' in Nederland, waarbinnen trollen actief zijn en gebruik wordt gemaakt van artificiële versterking.[75]

In dit verband is ook de groeiende populariteit van online streaming via YouTube, TikTok, SnapChat, InstagramTV en Twitch relevant. Nu zijn nog vooral vooraf gemonteerde video's, zogenaamde vlogs, populair, maar het is voorstelbaar dat in de toekomst meer en meer gebruik zal worden gemaakt van livestreams. Video's met desinformatie kunnen nu nog door socialemediabedrijven worden bestreden door video's te filteren, bijvoorbeeld door ze te controleren voordat ze gepubliceerd worden. Maar bij livestreams ligt dat anders. Dan bereikt de boodschap de ontvanger direct, en is filteren of factchecken lastig.

5.2 Micro-targeting

Micro-targeting stelt verzenders van informatie in staat om op nauwkeurige wijze een bepaalde doelgroep te bereiken met een op hen afgestemde boodschap. De hiervoor benodigde technologie wordt vaak voor commerciële marketingdoeleinden ontwikkeld. Zodra de technologie op de markt verschijnt, kan ze ook worden ingezet voor de verspreiding van desinformatie.[76]

Micro-targeting verandert de mogelijkheden om desinformatie te verspreiden op drie manieren. Allereerst vindt het verzamelen van data en het samenstellen van advertentieprofielen geautomatiseerd plaats, waardoor het op veel grotere schaal kan worden toegepast. Ten tweede is de technologie steeds beter in staat om te

[72] Bahara, H., Kranenberg, A., & Tokmetzis, D. (2019). *Hoe YouTube rechtse radicalisering in de hand werkt*. Volkskrant. www.volkskrant.nl/kijkverder/v/2019/hoe-youtube-rechtse-radicalisering-in-de-hand-werkt
[73] Tokmetzis, D., Bahara, H., & Kranenberg, A. (2019). *Aanbevolen voor jou op YouTube: racisme, vrouwenhaat en antisemitisme*. De Correspondent. https://decorrespondent.nl/9149/aanbevolen-voor-jou-op-youtube-racisme-vrouwenhaat-en-antisemitisme/445528853-0f710148
[74] Rogers, R., & Niederer, S. (2019). *Politiek en Sociale Media Manipulatie*. Universiteit van Amsterdam. www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2019/10/18/rapport-politiek-en-sociale-media-manipulatie/rapport-politiek-en-sociale-media-manipulatie.pdf
[75] Idem
[76] Kreling, T., & Modderkolk, H. (2020). *Hoe Spaanse software (onbedoeld) een gevaarlijk wapen werd voor online beïnvloeding*. Volkskrant. www.volkskrant.nl/nieuws-achtergrond/hoe-spaanse-software-onbedoeld-een-gevaarlijk-wapen-werd-voor-online-beinvloeding~b135b1bb/

bepalen tot welke doelgroep(en) een persoon hoort, en selecteert het systeem automatisch via welke kanalen die persoon kan worden benaderd. Ten derde maakt micro-targeting het mogelijk om de inhoud van berichten automatisch af te stemmen op de ontvanger.[77]

De in de literatuur beschreven voorbeelden van het gebruik van micro-targeting voor de verspreiding van desinformatie, hebben vaak betrekking op de Verenigde Staten. Maar deze praktijken kunnen zich ook in Europa voordoen.[78] Een belangrijk verschil is wel de Europese regelgeving op het gebied van gegevensbescherming (AVG). De AVG stelt namelijk grenzen aan de vergaring van persoonsgegevens. Zo mogen data over politieke voorkeuren niet zonder toestemming van de persoon in kwestie worden verzameld. Daarnaast zijn campagnes met micro-targeting (vooralsnog) kostbaar, waardoor ze niet voor iedereen toegankelijk zijn.[79]

Hieronder worden verschillende toepassingen van micro-targeting besproken.

5.2.1 Campagnesoftware

Om nauwkeurig een bepaald publiek te bereiken, bijvoorbeeld bij politieke campagnes, kan een breed scala aan verspreidingstechnologieën worden ingezet. Campagnesoftware maakt het mogelijk om de inzet van de diverse middelen te coördineren. Campagnes kunnen daarmee efficiënter en doelgerichter worden gemaakt, bijvoorbeeld door gelijktijdig meerdere socialemedianetwerken te analyseren om in te schatten welke informatie op welke doelgroep de grootste invloed heeft. Op basis daarvan kunnen beslissingen worden genomen over de inzet van offline campagnemiddelen, zoals (media)optredens, advertenties, het uitdelen van flyers, langs de deur gaan, of leden werven. Campagnesoftware automatiseert dit proces en gebruikt kunstmatige intelligentie om de effecten ervan vooraf in te schatten. Campagnesoftware vormt daarmee de cockpit voor verspreiding van politiek campagnemateriaal.

In de Verenigde Staten is hiervoor zeer geavanceerde specialistische software beschikbaar, zoals CampaignGrid. Deze software ondersteunt grootschalige datavergaring en -analyse, en deelt het land op in virtuele regio's. De software wordt ook gebruikt om de resultaten van campagneactiviteiten te registreren. Voor

[77] Crain & Nadler (2019). Political Manipulation and Internet Advertising Infrastructure. *Journal of Information Policy*, *9*, 370. https://doi.org/10.5325/jinfopoli.9.2019.0370
[78] Bennett, C. J. (2016). Voter databases, micro-targeting, and data protection law: can political parties campaign in Europe as they do in North America? *International Data Privacy Law*, *6*(4), 261–275. https://doi.org/10.1093/idpl/ipw021
[79] Dommett, K. (2019). Data-driven political campaigns in practice: understanding and regulating diverse data-driven campaigns. *Internet Policy Review*, *8*(4). https://policyreview.info/articles/analysis/data-driven-political-campaigns-practice-understanding-and-regulating-diverse-data

zover bekend, komt deze vorm van door technologie gedreven campagnevoering in Nederland (nog) niet voor.

Net als bij de technologie die ontwikkeld wordt voor commerciële marketingdoeleinden, kan campagnesoftware ook worden gebruikt voor de verspreiding van desinformatie (al dan niet om politieke redenen).

5.2.2 AdTech

AdTech staat voor advertentietechnologie. Onder deze noemer vallen allerlei micro-targeting-technieken die kunnen worden ingezet voor reclamedoeleinden. Het verspreiden van reclame kan een effectieve manier zijn om desinformatie te verspreiden, omdat de technologie de verzender controle geeft over het bereik en de boodschap.[80]

De omvang van de wereldwijde advertentiemarkt wordt geschat op 327 miljard dollar.[81] Volgens PwC is de markt voor online advertenties in Nederland met ruim 2 miljard euro, twee keer zo groot als die van TV en radio samen. Facebook en Google domineren de online advertentiemarkt in Nederland.[82]

Hieronder volgen twee ontwikkelingen op het gebied van AdTech die van grote invloed zouden kunnen zijn voor de nabije toekomst: *dynamic prospecting* en *programmatic advertising*.

Dynamic prospecting
Het bereiken van de juiste doelgroep geldt als een van de sleutels voor een effectieve reclamecampagne. Facebook ondersteunt marketeers hierin door ze toegang te geven tot gecategoriseerde data van gebruikers. Hiervoor verzamelt Facebook per gebruiker tienduizenden eigenschappen.[83] [84]Deze eigenschappen worden afgeleid van bijvoorbeeld de berichten die gebruikers plaatsen, hun vriendennetwerk of gezichtsherkenningsdata afkomstig van foto's en video's.

---

[80] Kim, Y. M., Hsu, J., Neiman, D., Kou, C., Bankston, L., Kim, S. Y., Heinrich, R., Baragwanath, R., & Raskutti, G. (2018). The Stealth Media? Groups and Targets behind Divisive Issue Campaigns on Facebook. *Political Communication*, *35*(4), 515–541. https://doi.org/10.1080/10584609.2018.1476425
[81] Crain & Nadler (2019). Political Manipulation and Internet Advertising Infrastructure. *Journal of Information Policy*, *9*, 370. https://doi.org/10.5325/jinfopoli.9.2019.0370
[82] Consultancy.nl (2019). *Reclame-inkomsten tv en radio steeds verder achterop bij internet*. www.consultancy.nl/nieuws/25963/reclame-inkomsten-tv-en-radio-steeds-verder-achterop-bij-internet
[83] Tobin, J. A., Madeleine Varner,Ariana. (2017). *Facebook Enabled Advertisers to Reach 'Jew Haters'*. ProPublica. www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters
[84] Dean, S. (2019). Facebook decided which users are interested in Nazis — and let advertisers target them directly. Los Angeles Times. www.latimes.com/business/technology/la-fi-tn-facebook-nazi-metal-ads-20190221-story.html

Aangezien 2,5 miljard mensen maandelijks gebruik maken van Facebook, gaat dit om een gigantische dataset.

Dynamic prospecting wordt gebruikt om uit deze gigantische dataset automatisch doelgroepen te selecteren, het effect van een bepaalde advertentie te kunnen inschatten en te analyseren, en doelgroepen bij te stellen. *Dynamic prospecting* wordt ingezet in politieke campagnes in de VS.[85] Het is de verwachting dat deze toepassing van kunstmatige intelligentie vanwege het gebruik van zelflerende algoritmes in de toekomst verder zal verbeteren, bijvoorbeeld door doelgroepen nog fijnmaziger in te delen. In 2017 bleek uit gelekte documenten dat Facebook de doelgroepenselectie mede baseert op de emotionele staat van tieners, zodat advertenties gericht kunnen worden op de mate waarin personen zich bijvoorbeeld 'waardeloos', 'onzeker' of 'nerveus' voelen.[86]

De algoritmes waarvan *dynamic prospecting* gebruikmaakt zijn niet openbaar, want worden gezien als concurrentiegevoelige informatie. Ook vindt er geen onafhankelijk toezicht plaats op het gebruik ervan. Dit gebrek aan transparantie maakt het lastig om te beoordelen hoe legitiem het gebruik ervan is, en of het bijvoorbeeld wordt ingezet voor de verspreiding van desinformatie.

Vanwege het potentieel dat *dynamic prospecting* biedt om specifieke doelgroepen van bepaalde informatie te voorzien en vanwege het gebrek aan transparantie over het precieze gebruik ervan, leent deze technologie zich goed voor de verspreiding van desinformatie.

Programmatic advertising

Naast het automatisch bijstellen van doelgroepen kan ook de inhoud van advertenties steeds preciezer worden afgestemd op de ontvanger. Voor deze technologie bestaan verschillende benamingen: '*programmatic advertising'*, '*creative versioning'*, '*dynamic creative',* en *'dynamic creative optimization'*.[87] De hiervoor gebruikte algoritmes kunnen de inhoud van een bericht aanpassen, afhankelijk van hoe de ontvanger erop reageert.

Zo biedt Google adverteerders de toepassingen Directors Mix en Vogon aan, waarmee automatisch honderden varianten van een video kunnen worden

[85] Montgomery, K., & Chester, J. (2020). *The digital commercialisation of US politics — 2020 and beyond.* Center for Digital Democracy. www.democraticmedia.org/article/digital-commercialisation-us-politics-2020-and-beyond
[86] Reilly, M. (2017). *Is Facebook Targeting Ads at Sad Teens?* MIT Technology Review. www.technologyreview.com/2017/05/01/105987/is-facebook-targeting-ads-at-sad-teens/
[87] Montgomery, K., & Chester, J. (2020). *The digital commercialisation of US politics — 2020 and beyond*. Center for Digital Democracy. www.democraticmedia.org/article/digital-commercialisation-us-politics-2020-and-beyond

geproduceerd.[88] Elke video bestaat uit een unieke combinatie van beeld en tekst, waarvan ook de lettergrootte en de positie van de tekst kunnen verschillen.[89] Netflix kan met deze technologie bijvoorbeeld dezelfde televisieserie presenteren als een special effects-spektakel, een familiedrama of een liefdesverhaal – afhankelijk van de ontvanger van de advertentie.[90] Facebook heeft een soortgelijke dynamic creative-omgeving.[91]

Figuur 6 In de YouTube Directors Mix kunnen beeld, geluid en tekst van een advertentie worden aangepast, afhankelijk van de doelgroep. Links een versie voor een Amerikaans publiek, rechts voor een Nederlands publiek.[92]

Tijdens de Amerikaanse presidentsverkiezingen in 2016 werd deze technologie op beperkte schaal ingezet. Zo zouden 40.000 tot 50.000 verschillende versies van één advertentie per dag worden gegenereerd.[93] In de nabije toekomst zal dit aantal naar verwachting sterk toenemen.

Zelfs al zouden deze advertenties worden geregistreerd in een publiek toegankelijk register voor politieke advertenties, dan nog is het de verwachting dat het voor bijvoorbeeld journalisten, vanwege het gigantische volume aan advertenties, lastig wordt om na te gaan op welke manier een politieke partij zich presenteert ten overstaan van welk publiek.

Een mogelijk risico hiervan is dat politieke partijen zich richting verschillende doelgroepen met een verschillende politieke boodschap kunnen presenteren. Voor de ontvanger wordt het dan lastig om te bepalen waar de partij werkelijk voor staat.

---

[88] Jain, K., & Chetan, A. (2018). *What brands can learn from India on personalized storytelling*. Think with Google. www.thinkwithgoogle.com/intl/en-apac/ad-channel/video/what-brands-can-learn-india-personalized-storytelling/
[89] Google (2020). *Project Vogon* https://opensource.google/projects/vogon
[90] Rothwell, J. (2018). *Perspectives: Find your audience on digital and storytell with data*. Think with Google www.thinkwithgoogle.com/intl/en-apac/tools-resources/success-stories/perspectives-find-your-audience-digital-and-storytell-data/
[91] Facebook (z.d.). *Dynamic Creative www.facebook.com/business/m/facebook-dynamic-creative-ads*
[92] Newfangled (z.d.). *Google: Director Mix* www.newfangledstudios.com/projects/google-directormix/
[93] TacticalTech. (2019). *Personal Data: Political Persuasion - The Guidebook and Visual Gallery*. https://ourdataourselves.tacticaltech.org/posts/inside-the-influence-industry

5.2.3 Psychographing

In zogeheten *psychographs* worden mensen aan de hand van hun karaktertrekken ingedeeld in doelgroepen. De gedachte hierachter is dat marketeers hiermee consumenten kunnen interesseren voor een product omdat het appelleert aan hun persoonlijke waarden en verlangens. Cola wordt bijvoorbeeld niet gepromoot als een dorstlesser, maar als iets feestelijks, zodat de klant het niet alleen drinkt als hij dorst heeft maar ook als hij verlangt naar een blij gevoel.

Deze technologie kan uiteraard ook door producenten en verspreiders van desinformatie worden gebruikt, zodat hun boodschap maximaal aansluit bij de gevoelens van de beoogde doelgroep. Zo zegt Cambridge Analytica doelgroepen in te delen op basis van vijf kenmerken uit hun OCEAN-model: *Openness to experience*, *Conscientiousness*, *Extraversion, Agreeableness and Neuroticism.* Op basis van de scores op elk van deze vijf persoonlijkheidsaspecten kunnen politieke boodschappen worden aangepast aan het publiek.[94]

Figuur 7 Toenmalig Cambridge Analytica-directeur Alexander Nix toont hoe de inhoud van politieke advertenties wordt bepaald door het OCEAN-model.[95]

5.2.4 Influencer marketing

Influencer marketing kan gezien worden als een moderne vorm van mond-op-mond-reclame. Beheerders van socialemedia-accounts met veel volgers krijgen

[94] Concordia. (2016). Cambridge Analytica - The Power of Big Data and Psychographics. www.youtube.com/watch?v=n8Dd5aVXLCc
[95] Idem

betaald om reclame te maken voor een product, dienst of merk. Voor de ontvangers hiervan is niet altijd duidelijk dat het gaat om reclame. Dit komt onder andere doordat de verzenders ook veel niet-commerciële berichten verzenden, meestal met als doel om de interesse te wekken van nieuwe volgers.
Influencers zijn vaak bekende personen, zoals stervoetballer Cristiano Ronaldo met meer dan 200 miljoen volgers op Instagram[96]. Op een bericht van Ronaldo volgen doorgaans zo'n 5 miljoen reacties. Maar lang niet alle influencers zijn internationale beroemdheden. Via bemiddelingsbureaus zijn ook influencers met enkele duizenden volgers in te schakelen voor verspreiding van informatie. Verder zijn influencers niet noodzakelijkerwijs personen. Een merk als Nike, met 100 miljoen volgers op Instagram[97], bouwt ook flinke hoeveelheden volgers op.

Virtuele en politieke influencers
Een relatief nieuw fenomeen zijn virtuele influencers. Dit zijn accounts waarin een virtueel persoon een product of dienst promoot. Lil Miquela, met 2 miljoen volgers op Instagram, is een voorbeeld van zo'n virtuele influencer.[98] Zij promoot op Instagram met name kleding- en lifestylemerken door middel van foto's. Hierop staat zij vaak samen met bestaande artiesten afgebeeld.

Figuur 8 Lil Miquela in Los Angeles.[99]

Influencers kunnen ook politiek actief zijn.[100] Er wordt dan gesproken van een politieke influencer. Marketingfirma Drawbridge (niet langer actief) bood een dienst aan onder de noemer *Political Influencer Identification*. Hiermee werden influencers

[96] www.instagram.com/cristiano/
[97] www.instagram.com/nike/
[98] www.instagram.com/lilmiquela/
[99] www.instagram.com/p/B82j0y-Hcia/
[100] Chester, J., & Montgomery, K. C. (2017). The role of digital marketing in political campaigns. *Internet Policy Review*, *6*(4). https://policyreview.info/articles/analysis/role-digital-marketing-political-campaigns

en hun volgers gecategoriseerd op basis van politieke voorkeuren, met als doel hen in te zetten in campagnes. Drawbridge is van Twitter verwijderd vanwege het schenden van de gebruikersvoorwaarden.

Micro-influencermarketing
Binnen *influencer marketing* worden steeds vaker technieken van micro-targeting toegepast. Adverteerders willen hun middelen zo gericht mogelijk inzetten, en voorkomen dat ze geld uitgeven aan het benaderen van de verkeerde doelgroep. Om hieraan tegemoet te komen brengen de beheerders van socialemedia-accounts steeds nauwkeuriger in kaart wie de accounts volgt. Adverteerders geven vervolgens de beheerders van de accounts waarvan de volgers het beste bij hun doelgroep passen, de opdracht om bepaalde informatie te verspreiden. Het kan daarbij ook gaan om accounts met relatief weinig volgers, maar met specifieke kenmerken.

Van berichten van influencers met veel volgers is vrij algemeen bekend dat de inhoud van de informatie die ze verspreiden mede wordt bepaald door adverteerders, hoewel dit vaak niet expliciet wordt vermeld. Bij *micro-influencermarketing*, waarbij het om kleine aantallen volgers gaat, bestaat het vermoeden dat volgers hiervan veel minder op hun hoede zijn voor beïnvloeding. Deze micro-marketing komt immers over alsof een naaste of bekende oprecht een bepaald product promoot of (politiek) standpunt verkondigt.

Het gebrek aan transparantie van *influencer marketing* in het algemeen, en *micro-influencermarketing* in het bijzonder, maakt dat deze vorm van informatieverspreiding zich bijzonder goed leent voor de verspreiding van desinformatie. De verspreiders van desinformatie kunnen heel gericht opdracht geven, zonder dat de doelgroep doorheeft door wie en waarom ze bepaalde informatie krijgt voorgeschoteld.

5.3 Chatapps

Chat wordt ook wel *instant messaging* of *direct messaging* genoemd. Daarmee wordt ook een belangrijke eigenschap van chats benoemd. Informatie-uitwisseling door middel van chats heeft namelijk een hoger tempo en veelal een persoonlijker karakter dan socialemediaplatforms als Facebook en Twitter. Voor veel chatapps geldt dat berichten in chronologische volgorde verschijnen, zonder dat een aanbevelingsalgoritme daarop invloed uitoefent.

In Nederland zijn chatapps erg populair – denk hierbij vooral aan WhatsApp, maar ook aan SnapChat, Telegram, FaceTime, Google Hangouts, Skype, Slack of

Signal. Daarnaast bieden socialemediaplatforms vaak interne chatfuncties aan, zoals Facebook Messenger en Twitter Direct Messaging. Door gamers wordt ook veel gechat, zoals via chatapps als Discord of rechtstreeks in games zoals FortNite.

Chatapps bieden doorgaans de mogelijkheid om tekst, audio- en videoberichten op te nemen en te versturen. Daarnaast bevatten ze vaak functies voor het verzenden van emoji's, geanimeerde afbeeldingen (gifs), memes, stickers en andere multimedia, zoals YouTube-video's. Bij een aantal chatapps kunnen live videogesprekken worden gevoerd, waarbij in sommige gevallen de mogelijkheid wordt aangeboden om beelden direct te manipuleren met behulp van filters.

Een belangrijke functionaliteit van chatapps zijn de groepen of kanalen. Via deze functionaliteit kan een grote groep gebruikers in één keer bepaalde informatie worden gestuurd. WhatsApp heeft het aantal gebruikers in een groep gelimiteerd tot 256. Bij Telegram ligt het maximum voor een groep op 200.000 gebruikers, maar zijn de kanalen ongelimiteerd. In het buitenland bestaan bijvoorbeeld Telegramkanalen met miljoenen deelnemers. In Iran worden deze kanalen bijvoorbeeld gebruikt voor het uitwisselen van actualiteiten.[101] In landen als Brazilië vormen chatapps zelfs de belangrijkste nieuwsbron van burgers. In Nederland maken nieuwsmedia slechts op beperkte schaal gebruik van chatapps. Een voorbeeld hiervan zijn audionieuwsberichten van RTL Nieuws in WhatsApp.

Het hoeft geen betoog dat chatapps voor producenten en verspreiders van desinformatie een aantrekkelijk medium zijn om hun boodschap te verspreiden. De groepen en kanalen hebben vaak een besloten karakter, waardoor de kans kleiner is dat beweringen worden weersproken.

API voor datavergaring

In de controverse rondom Cambridge Analytica (CA) werd het Facebook kwalijk genomen dat het de vergaring van datasets door CA faciliteerde. Door middel van een zogeheten *Application programming interface* (API) maakte Facebook het partijen als CA mogelijk om op grote, geautomatiseerde schaal data uit Facebook te halen, vaak zonder dat de personen waarop de data betrekken hadden dat in de gaten hadden. Na het schandaal rond CA heeft Facebook deze mogelijkheden ingeperkt.

Diverse chatapps bieden deze mogelijkheid echter nog steeds. Zo biedt Telegram een uitgebreide API waardoor beheerders van kanalen informatie over deelnemers kunnen vergaren en geautomatiseerd berichten kunnen plaatsen (meer hierover in

[101] Telegram Channels (z.d.). *The Biggest 100 Media* https://telegramchannels.me/list/biggest

paragraaf 5.4).[102] Ook Signal beschikt over een soortgelijke programmeerbare interface.[103] In deze interfaces schuilt het gevaar van herhaling van de datavergaringspraktijken zoals die bij Facebook hebben plaatsgevonden.

Toenemend gebruik van versleutelingstechnologie
Een opvallende ontwikkeling bij chatapps is de toepassing van versleutelingstechnologie (encryptie). Door versleuteling wordt informatie onleesbaar gemaakt voor diegenen die niet over de juiste sleutel beschikken. Waar de ene chatapp enkel de inhoud van berichten versleutelt, is bij andere zelfs niet te achterhalen wie met elkaar chat en op welk moment. Vaak wordt hierbij gebruikgemaakt van *end-to-end-encryption*. Dat betekent dat het platform informatie uitwisselt tussen verzender en ontvanger zonder dat beheerders van het platform inzicht hebben in (een deel van) de informatie.

De opkomst van versleuteling wordt door privacy- en veiligheidsexperts vaak toegejuicht. Chatapps met versleutelingstechnologie worden als veiliger beschouwd dan bijvoorbeeld email. Chatapps met vergaande versleuteling genieten wat hen betreft de voorkeur. De Europese Commissie stelt bijvoorbeeld het gebruik van Signal onder haar personeel verplicht.[104]

Een gevolg van versleutelingstechnologie is wel dat kwaadwillende partijen, zoals verspreiders van desinformatie, lastiger in beeld kunnen worden gebracht.[105] Een ander effect van versleutelingstechnologie is dat aanbieders van chatapps geen zicht hebben op het gebruik van de applicatie en de inhoud van berichten, en daarop ook niet aanspreekbaar zijn.

5.4 Bots

Een bot is een socialemedia- of chat-account dat automatisch wordt aangestuurd door een algoritme, vaak grotendeels zonder menselijk handelen. Bots zijn in toenemende mate in staat om informatie te creëren en te interacteren met mensen, waardoor het voor de laatsten vaak onduidelijk is dat ze met een bot communiceren. Bots kunnen ook worden gebruikt om informatie te vergaren,

[102] Aichara, D. C. (2019). *Telegram Channel Data Extraction (User's information, chats, and specific messages) and Data Processing.* Medium. https://medium.com/game-of-data/telegram-channel-data-extraction-users-information-chats-and-specific-messages-and-data-21bb54710fd3
[103] Signal App (2014). *API Protocol* https://github.com/signalapp/Signal-Server/wiki/API-Protocol
[104] Cerulus, L. (2020*). EU Commission to staff: Switch to Signal messaging app*. Politico www.politico.eu/pro/eu-commission-to-staff-switch-to-signal-messaging-app/
[105] Nieuwsuur (2020). *'Nu IS van Telegram is verwijderd, zijn ze moeilijker in de gaten te houden.'* https://nos.nl/l/2318257

bijvoorbeeld door deel uit te maken van een groep of kanaal op sociale media of chatapps.

Bots kunnen op verschillende manieren worden ingezet om desinformatie te verspreiden. Allereerst door desinformatie te plaatsen op socialemediaplatforms. Vervolgens kunnen bots allerlei interacties aangaan die op het platform worden aangeboden, zoals een bericht liken, delen of van commentaar voorzien.[106] Met deze interacties kan bijvoorbeeld het aanbevelingsalgoritme van een socialemediaplatform worden beïnvloed, zodat berichten meer of minder vaak worden weergegeven. Om dezelfde reden kunnen bots worden ingezet om het aantal volgers van een account kunstmatig op te krikken. Bots kunnen ook hashtags veelvuldig gebruiken om trending topics te beïnvloeden of een hashtagdiscussie over te nemen.

Wanneer bots worden opgemerkt door de beheerder van een platform, kan deze besluiten een bericht of account te verwijderen, als het gebruik van bots in strijd is met de gebruikersvoorwaarden van het platform. Kwaadwillenden kunnen hier overigens ook op inspelen door bots te laten interacteren met berichten of accounts die ze graag de mond willen snoeren.

Bots komen op alle socialemediaplatforms voor. Naar schatting zijn tot wel 15% van alle Twitteraccounts bots. Volgens Politicalbots.org waren ongeveer 19 miljoen bot-accounts actief in de laatste week voor de Amerikaanse presidentsverkiezingen in 2016.

De impact van bots blijft niet beperkt tot de socialemediaplatforms waarop ze actief zijn. Soms worden berichten van bots gebruikt als vox populi in nieuwsberichten van traditionele media. Zo bleek nadat een groot aantal bots van de Russische trollenfabriek IRA was ontmaskerd, dat traditionele media berichten afkomstig van tweets van bots hadden overgenomen.[107] Ook grote Noorse nieuwsmedia hadden berichten van IRA-bots overgenomen in hun berichtgeving, terwijl zij dachten dat het om authentieke berichten ging van Noorse twitteraars.[108]

---

[106] Hussain, M. N., Tokdemir, S., Agarwal, N., & Al-Khateeb, S. (2018). Analyzing Disinformation and Crowd Manipulation Tactics on YouTube. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 1092–1095. https://doi.org/10.1109/ASONAM.2018.8508766

[107] Lukito, J., Suk, J., Zhang, Y., Doroshenko, L., Kim, S. J., Su, M.-H., Xia, Y., Freelon, D., & Wells, C. (2019). The Wolves in Sheep's Clothing: How Russia's Internet Research Agency Tweets Appeared in U.S. News as Vox Populi: *The International Journal of Press/Politics.* https://doi.org/10.1177/1940161219895215

[108] NOS (2020). *Alle grote media in Noorwegen trappen in tweets van Russische trollen.* https://nos.nl/l/2325674

Bots worden ook vaak ingezet in combinatie met andere verspreidingstechnieken. Zo bleken de IRA-bots vooral effectief in het verspreiden van berichten op het sterk gekleurde Russische nieuwsmedium Russia Today (RT).[109]

Beïnvloeding van aanbevelingsalgoritmes kan trouwens ook gebeuren met door mensen aangestuurde accounts. Zo bleek de Zuid-Koreaanse inlichtingendienst handmatig Twitteraccounts aan te sturen om de politieke stemming in het land te beïnvloeden.[110] Marketingfirma's zetten doorgaans ook menselijke accounts in om aanbevelingsalgoritmes te beïnvloeden. Op Twitter kunnen enkele honderden accounts al voldoende zijn om in een land als Spanje effect te hebben.[111]

Chatbots

Chatbots zijn een specifiek type bots, die rechtstreeks chatten met mensen. Ze zijn veelal bekend van klantenservicediensten. Vaak heeft een persoon snel door dat het om een bot gaat, vanwege de vaak nog wat simplistische manier waarop de bot reageert, maar de technologie is volop in ontwikkeling. Zo presenteerde Google in 2020 de chatbot Meena, die volgens hun eigen scoringssysteem beter presteert dan ander gangbare chatbots, en steeds overtuigender menselijke interacties kan imiteren.[112]

Het is de verwachting dat naarmate chatbottechnologie zich verder ontwikkelt, kwaadwillende partijen desinformatie meer en meer (semi-)geautomatiseerd, via chatbots in een-op-een-gesprekken met hun doelwit zullen verspreiden.[113]

---

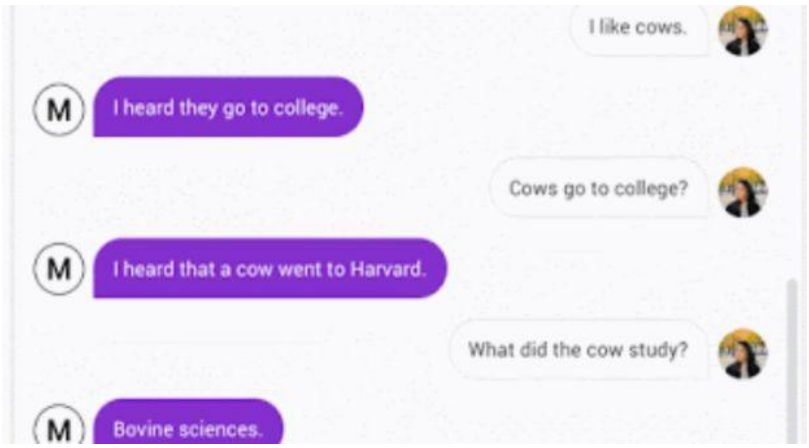
[109] Zannettou, S., Caulfield, T., De Cristofaro, E., Sirivianos, M., Stringhini, G., & Blackburn, J. (2019). Disinformation Warfare: Understanding State-Sponsored Trolls on Twitter and Their Influence on the Web. *ArXiv:1801.09288 [Cs].* Retrieved from http://arxiv.org/abs/1801.09288

[110] Keller, F. B., Schoch, D., Stier, S., & Yang, J. (2020). Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign. *Political Communication*, *37*(2), 256–280. . https://doi.org/10.1080/10584609.2019.1661888

[111] Kreling, T., & Modderkolk, H. (2020). *Hoe Spaanse software (onbedoeld) een gevaarlijk wapen werd voor online beïnvloeding.* Volkskrant. www.volkskrant.nl/nieuws-achtergrond/hoe-spaanse-software-onbedoeld-een-gevaarlijk-wapen-werd-voor-online-beinvloeding~b135b1bb/

[112] Adiwardana, D., & Luong, T. (2020). *Towards a Conversational Agent that Can Chat About…Anything*. https://ai.googleblog.com/2020/01/towards-conversational-agent-that-can.html

[113] TacticalTech. (2019). *Personal Data: Political Persuasion - The Guidebook and Visual Gallery*. https://ourdataourselves.tacticaltech.org/posts/inside-the-influence-industry



Figuur 9 De chatbot Meena toont in een demonstratie van Google een zeker gevoel voor humor.[114]

5.5 Zoekmachines

Zoekmachines zijn websites waarop gebruikers een zoekopdracht kunnen invoeren, waarna hen een selectie van links wordt getoond. Bekende voorbeelden zijn Google Search, Microsoft Bing en DuckDuckGo. Zoekmachines worden ook steeds vaker gebruikt als vraag-antwoordmachines. Gebruikers kunnen hun zoekopdracht in de vorm van een vraag formuleren, waarna de zoekmachine mogelijke antwoorden zoekt. Gebruikers van zoekmachines kiezen over het algemeen vaker voor hoger dan voor lager gerangschikte resultaten. In plaats van alleen te werken met geschreven opdrachten, zijn zoekmachines in toenemende mate in staat om ook op basis van een spraakopdracht, afbeelding of video op zoek te gaan naar informatie.

Gebruikers van socialemediaplatforms als Facebook, Instagram en YouTube maken ook steeds vaker gebruik van (interne) zoekmachines om door informatie te navigeren.

Door het veelvuldige gebruik van zoekmachines, spelen ze een belangrijke rol in de informatieverspreiding. In Nederland hebben gebruikers ook meer vertrouwen in nieuwsberichten die worden aanbevolen door zoekmachines dan in de informatie die ze vinden op sociale media. Maar de resultaten van een zoekopdracht zijn niet altijd even betrouwbaar. De algoritmes waarmee zoekmachines werken zijn

[114] Schwartz, E. (2020). *Google's New Meena Chatbot Imitates Human Conversation and Bad Jokes. Voicebot.* https://voicebot.ai/2020/02/03/googles-new-meena-chatbot-imitates-human-conversation-and-bad-jokes/

namelijk te manipuleren.[115] Ook nemen zoekmachines berichten die populair zijn op sociale media over, terwijl die op hun beurt weer zijn te beïnvloeden.[116]

Zoekmachinetechnologie kan worden gecombineerd met afbeeldingstechnologie. Zo laten experimenten van Amazon met de eerder genoemde *generative adversarial networks* (GANs) zien, dat op basis van een geschreven zoekopdracht bijbehorende afbeeldingen kunnen worden gegenereerd. Zo kan een afbeelding van een gerecht worden gegenereerd op basis van ingrediënten.[117] [118]

Omdat op het internet, en zeker op socialemediaplatforms, beeldmateriaal een steeds belangrijkere rol speelt, valt te verwachten dat beeldherkennings- en beeldsynthese-algoritmes (zoals GANs) meer gebruikt gaan worden door zoekmachines. Dat biedt ook nieuwe kansen voor het verspreiden van desinformatie.

5.6 Spraakassistenten

Spraakgestuurde digitale assistenten – oftewel spraakassistenten – zijn informatiebronnen die door een menselijke stem kunnen worden aangestuurd. Bekende voorbeelden hiervan, zoals Amazon Alexa, Apple Siri en Google Assistant, kan worden gevraagd naar het weerbericht of naar een antwoord op allerlei vragen. Spraakassistenten kunnen fysieke apparaten zijn die in een huis worden geplaatst, maar kunnen ook een functie zijn van apparaten als smart watches, smart phones, laptops of smart TV's.

In 2019 had vijf procent van de Nederlanders een spraakassistent in huis. De komende jaren wordt daarvan een sterke groei verwacht.[119] Uit publicaties van het Nationaal Luister Onderzoek blijkt dat 34% van de Nederlanders de spraakassistent gebruikt voor de raadpleging van het nieuws.[120] Ook in landen als de VS, het VK, Australië en Canada groeit het aantal spraakassistenten snel.[121] De groei bestaat

---

[115] Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, *112*(33), E4512–E4521. https://doi.org/10.1073/pnas.1419828112

[116] Robertson, A. (2017). *It's time to stop trusting Google search already*. The Verge www.theverge.com/2017/11/10/16633574/stop-trusting-google-search-texas-shooting-twitter-misinformation

[117] Biswas, A., & Surya, S. (2020). *Converting text to images for product discovery*. Amazon Science Blog www.amazon.science/blog/converting-text-to-images-for-product-discovery

[118] Synced (2020). *CookGAN Generates Realistic Meal Images From an Ingredients List.* https://medium.com/syncedreview/cookgan-generates-realistic-meal-images-from-an-ingredients-list-250426dbfab2

[119] TNS NIPO (2019). *Gebruik smart speakers groeit explosief.* www.tns-nipo.com/nieuws/persberichten/gebruik-smart-speakers-groeit-explosief

[120] Audiomonitor (2019). *Nationaal Luister Onderzoek.* https://nationaalluisteronderzoek.nl/audiomonitor-slides/

[121] Newman, N. et al. (2019). *Reuters Institute Digital News Report 2019.* Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/inline-files/DNR_2019_FINAL.pdf

niet alleen uit populaire apparaten van Amazon en Google, maar ook uit producten van veel minder grote fabrikanten uit China.[122]

Vanuit het oogpunt van desinformatie moet ook aandacht uitgaan naar dit gebruik van spraakassistenten. Ze gebruiken namelijk kwetsbare bronnen voor het geven van hun antwoorden, zoals Wikipedia.[123] Op dit platform kunnen bijvoorbeeld complottheorieën of hoaxen soms lang onopgemerkt blijven.[124] Kwaadwillenden kunnen deze kwetsbare bronnen dan ook gebruiken om desinformatie via spraakassistenten te verspreiden. Daarnaast zijn, net als bij zoekmachines, de algoritmes waarmee spraakassistenten werken te manipuleren.

5.7 Distributed Autonomous Applications

Distributed computing technologie maakt het mogelijk om computers in een netwerk een taak te laten uitvoeren zonder dat zij worden aangestuurd door een centrale partij. De computers in het netwerk maken onderling uit welke computer welk deel van een hun toebedeelde taak voor zijn rekening neemt. Blockchain is hiervan een voorbeeld. Hierbij houdt een netwerk van computers een grootboekrekening bij, waarbij de computers onderling vaststellen of wijzigingen in de grootboekrekening zijn toegestaan. Een bekende toepassing hiervan is de virtuele munt Bitcoin.

Sinds de introductie van Bitcoin in 2008 is een groot aantal nieuwe vormen van *distributed computing* tot ontwikkeling gekomen. Voor diverse alledaagse online applicaties zijn *distributed* alternatieven ontwikkeld, zoals het videoplatform D-Tube en het blogplatform Steem. Kenmerkend voor deze applicaties is dat informatie die eenmaal is toegevoegd, nauwelijks kan worden verwijderd. Alleen door een gezamenlijke inspanning van een meerderheid van de computers in het netwerk kan een bijdrage worden verwijderd. Maar vanwege de afwezigheid van een organisatiestructuur of centrale aansturing, is dit in de praktijk erg lastig.

Dit gebrek aan moderatiemogelijkheden maakt dit soort applicaties aantrekkelijk voor misbruik door kwaadwillende partijen, zoals verspreiders van desinformatie.

---

[122] Emerce. (2020). *Consument kiest vaker voor Chinese slimme luidspreker.* www.emerce.nl/nieuws/consument-kiest-vaker-chinese-slimme-luidspreker

[123] Kinsella, B. (2019). *Voice Assistants Alexa, Bixby, Google Assistant and Siri Rely on Wikipedia and Yelp to Answer Many Common Questions about Brands.* Voicebot https://voicebot.ai/2019/07/11/voice-assistants-alexa-bixby-google-assistant-and-siri-rely-on-wikipedia-and-yelp-to-answer-many-common-questions-about-brands/

[124] Kumar, S., West, R., & Leskovec, J. (2016). Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes. *In Proceedings of the 25th International Conference on World Wide Web (pp. 591–602).* Montréal, Québec, Canada: International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/2872427.2883085

Ze ontlopen via deze platforms het risico dat hun informatie verwijderd wordt. Ook bieden de platforms een hoge mate van anonimiteit aan gebruikers.[125]

Distributed Autonomous Applications zijn niet erg populair. Volgens websitevergelijker SimilarWeb heeft D-Tube met 300.000 bezoekers per maand in maart 2020 slechts een fractie van het bereik van YouTube, met 30 miljard bezoekers in maart 2020.[126] Dat neemt niet weg dat dit soort platforms in bepaalde niches een belangrijke rol kan vervullen.

Distributed computing technologie kan overigens ook worden ingezet om desinformatie te bestrijden. Door oorspronkelijke en authentieke informatie vast te leggen in gedistribueerde applicaties, kunnen gebruikers nagaan wat de bron ervan is.[127] De mogelijkheid om informatie (vrijwel) permanent op internet aan te bieden biedt dus zowel kansen als bedreigingen voor desinformatie.

5.8 Games

Digitale games zijn populair in Nederland, vooral onder jongeren. Volgens het Nederlands Jeugdinstituut speelt 35% van de basisschoolleerlingen en 27% van de 12- tot 16-jarigen dagelijks een game.[128]

Games zijn meestal gebaseerd op fictieve scenario's, maar kunnen net als films pretenderen een waargebeurd verhaal te vertellen. Games gebruiken vaak situaties uit het verleden als context of achtergrond, maar volgen de loop van de geschiedenis niet altijd even accuraat.[129]

Vanwege de populariteit van games en hun meeslepende en soms zelfs verslavende karakter, bieden ze interessante mogelijkheden om desinformatie of bepaalde narratieven te verspreiden. Net als in films komen in games vaak stereotypes en sterk politiek gekleurde verhaallijnen voor, zoals de Verenigde Staten als overwinnaar in de strijd tussen goed en kwaad.

---

[125] Polyakova, A., & Meserole, C. (2018). *Disinformation Wars.* Foreign Policy https://foreignpolicy.com/2018/05/25/disinformation-wars/
[126] Similarweb.com (z.d.). www.similarweb.com/website/d.tube/ en www.similarweb.com/website/youtube.com/
[127] Huckle, S., & White, M. (2017). Fake News: A Technological Approach to Proving the Origins of Content, Using Blockchains. *Big Data*, 5(4), 356–371. https://doi.org/10.1089/big.2017.0071
[128] NJi. (2019). *Gamen - Cijfers.* www.nji.nl/nl/Databank/Cijfers-over-Jeugd-en-Opvoeding/Cijfers-per-onderwerp/Gamen
[129] Veugen, C. (2014). Using Games to Mediate History. In L. Egberts & K. Bosma (Red.), *Companion to European Heritage Revivals* (pp. 95–111). Springer International Publishing. https://doi.org/10.1007/978-3-319-07770-3_5

Ook kunnen games worden ingezet om persoonsgegevens te vergaren, die vervolgens kunnen worden gebruikt voor desinformatiedoeleinden. Zo werden Facebookgebruikers door middel van eenvoudige spelletjes verleid om hun persoonsgegevens te delen met Cambridge Analytica.[130]

Politieke partijen kunnen games inzetten om kiezers aan zich te binden of mensen te stimuleren om informatie te delen. Zo kunnen gebruikers van de Trump-2020-App punten verdienen door berichten van Trump op Twitter te delen.[131] Het op deze wijze toepassen van spelelementen kan ook worden ingezet voor de verspreiding van desinformatie.

5.9 Crossmediale storytelling

Crossmediale storytelling wordt ingezet om ervoor te zorgen dat een ontvanger via diverse kanalen steeds opnieuw kan worden bereikt door een afzender. Die kanalen kunnen socialemediaplatforms zijn, streaming video of chatapps, maar ook apparaten zoals smartphones, tv's en computers. Crossmediale storytelling kan dus de krachten bundelen van diverse hierboven besproken technologieën.

Enerzijds bestaat deze technologie uit instrumenten die het mogelijk maken om individuele ontvangers of doelgroepen te identificeren en te volgen op de verschillende kanalen. Anderzijds biedt deze technologie de mogelijkheid om deze kanalen op een gecoördineerde wijze in te zetten om ontvangers voortdurend te bereiken.[132] Verspreiders van desinformatie kunnen hiervan gebruikmaken. Omdat dezelfde boodschap van verschillende bronnen afkomstig lijkt, kan deze aan geloofwaardigheid winnen. Maar crossmediale storytelling biedt ook andere voordelen voor verzenders van desinformatie. Zo bleek uit onderzoek naar de IRA dat onderwerpen vaak een week eerder op Reddit opdoken dan op Twitter. Vermoed wordt de IRA Reddit gebruikte om het effect van bepaalde boodschappen te testen.[133]

[130] TacticalTech. (2019). *Personal Data: Political Persuasion - The Guidebook and Visual Gallery.* https://ourdataourselves.tacticaltech.org/posts/inside-the-influence-industry
[131] Trump, D. (2020). *Trump 2020 App IS HERE!* www.youtube.com/watch?v=JRuQ5JMMqtM
[132] Chester, J., & Montgomery, K. C. (2017). The role of digital marketing in political campaigns. *Internet Policy Review, 6*(4). https://policyreview.info/articles/analysis/role-digital-marketing-political-campaigns
[133] Lukito, J. (2020). Coordinating a Multi-Platform Disinformation Campaign: Internet Research Agency Activity on Three U.S. Social Media Platforms, 2015 to 2017. *Political Communication, 37*(2), 238–255. https://doi.org/10.1080/10584609.2019.1661889

6 Bestaande maatregelen

Dit laatste hoofdstuk van de quickscan biedt een beknopt overzicht van bestaande maatregelen die worden genomen om de dreigingen die van desinformatie uitgaan voor het publieke debat en het democratisch proces, het hoofd te bieden. We richten ons hierbij op maatregelen door de Nederlandse overheid, de Europese Unie en platformbedrijven. Wat de platformbedrijven betreft, beperken we ons tot enkele grote spelers.

6.1 Maatregelen Nederlandse overheid

Doelstelling en uitgangspunten
Het beleid van de Nederlandse overheid op het gebied van desinformatie is erop gericht de stabiliteit en kwaliteit van de democratische rechtsorde en de open samenleving te beschermen.

Bij het nemen van het maatregelen tegen desinformatie hanteert de overheid onder andere de volgende uitgangspunten:

- Rechtsstatelijke waarden en grondrechten zoals de vrijheid van meningsuiting, de persvrijheid en het recht op informatie staan voorop;
- Onafhankelijke journalistiek en een pluriform medialandschap zijn onontbeerlijk voor een gezonde democratie;
- Mediawijsheid en digitale geletterdheid vormen belangrijke elementen in het tegengaan van de impact van desinformatie;
- Burgers moeten zelf informatie op waarde schatten. Transparantie over de herkomst van informatie is daarbij van fundamenteel belang;
- Internetdiensten dragen eigen verantwoordelijkheden. Waar hun zelfregulering tekortschiet, kan regulering worden overwogen;
- Kennisontwikkeling door de wetenschap over het bestaan van desinformatie wordt verwelkomd;
- Het kabinet ondersteunt de coördinatie in Europees en breder internationaal verband.[134]

De insteek van het beleid ligt eerder op het beperken van de impact van desinformatie dan op het actief tegenspreken of ontkrachten ervan. De overheid

[134] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. (2019). *Kamerbrief over beleidsinzet bescherming democratie tegen desinformatie* www.rijksoverheid.nl/documenten/kamerstukken/2019/10/18/kamerbrief-over-beleidsinzet-bescherming-democratie-tegen-desinformatie

ziet voor zichzelf ook geen primaire rol weggelegd bij het actief weerspreken van desinformatie. Zij acht dat in de eerste plaats een verantwoordelijkheid van niet-gouvernementele actoren zoals onafhankelijke media, online platforms en wetenschappers. De overheid ziet wel een rol voor zichzelf weggelegd als de politieke en economische stabiliteit of de nationale veiligheid in het geding is, of bij het publiekelijk uitdragen van haar beleid.[135]

Drie actielijnen
De maatregelen van de Nederlandse overheid tegen desinformatie kennen drie actielijnen:
1. Preventieve acties moeten voorkomen dat desinformatie impact heeft en zich verspreidt;
2. Versteviging van de informatiepositie moet tijdig zicht geven op (potentiële) dreigingen;
3. Reactieve acties worden ondernomen op zich voordoende desinformatie.

Vooralsnog ligt de nadruk van het overheidsbeleid op het nemen van preventieve acties.

De drie actielijnen krijgen als volgt invulling:

Preventieve acties
- Versterken van de weerbaarheid van burgers tegen de invloed van desinformatie door middel van bewustwordingscampagnes en het stimuleren van mediawijsheid;
- Versterken van de weerbaarheid van politieke ambtsdragers door middel van een game over desinformatie en mogelijke handelingsperspectieven;
- Vergroten van transparantie over (de aanpak van) desinformatie, onder andere door het monitoren van de implementatie van een Europese gedragscode voor platformbedrijven;
- Behoud van een pluriform medialandschap, onder andere door extra middelen vrij te maken voor onderzoeksjournalistiek;
- Innovatie in consumptie en productie van online nieuws, onder andere door het ontwikkelen van kwaliteitsstandaarden.

Versteviging van de informatiepositie
- Verbeteren van de informatiepositie in (inter)nationaal verband, onder andere door deelname in EU-verband aan het *Rapid Alert System*, waarmee snel meldingen over desinformatiecampagnes kunnen worden gedeeld;

[135] Idem

- Internationale samenwerking, onder andere door middel van het netwerk- en expertiseplatform European Centre of Excellence on Countering Hybrid Threats;
- Kennisontwikkeling.

Reactieve acties
- Inhoudelijk adresseren van desinformatie – factchecken – door middel van onafhankelijk van de overheid functionerende factcheckers;
- Weerspreken van desinformatie;
- Verkennen van de mogelijkheden van en verantwoordelijkheden voor het modereren van berichten op online platforms.[136] [137]

6.2 Maatregelen Europese Unie

De Europese Unie heeft diverse initiatieven ontwikkeld ter bestrijding van desinformatie. We beperken ons in deze paragraaf tot een beschrijving van het Europese actieplan tegen desinformatie en de Europese gedragscode tegen desinformatie.

EU-actieplan tegen desinformatie
Het Europese actieplan tegen desinformatie behelst de volgende acties:
- Verbeteren van de capaciteit van EU-instellingen en lidstaten om desinformatie te detecteren, te analyseren en publiek te maken, door te investeren in digitale hulpmiddelen en gespecialiseerd personeel;
- Versterken van een gecoördineerde en gezamenlijke respons op desinformatiecampagnes, onder andere door het al eerder genoemde *Rapid Alert System*;
- Stimuleren van de private sector om desinformatie aan te pakken, onder andere door het monitoren van de implementatie van de *EU Code of Practice on Disinformation*;
- Vergroten van de bewustwording en versterken van de maatschappelijke weerbaarheid, onder andere door bewustwordingscampagnes binnen en buiten de EU en door het ondersteunen van onafhankelijke media en factcheckers.[138]

[136] Idem
[137] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. (2019). *Actielijnen tegengaan desinformatie.* www.rijksoverheid.nl/documenten/kamerstukken/2019/10/18/actielijnen-tegengaan-desinformatie
[138] European Commission (2018). *Action Plan on Disinformation* https://ec.europa.eu/commission/publications/action-plan-disinformation-commission-contribution-european-council-13-14-december-2018_en

EU Code of Practice on Disinformation
De *EU Code of Practice on Disinformation* bevat een lijst met standaarden voor zelfregulering die is opgesteld door vertegenwoordigers van platformbedrijven, sociale netwerken en adverteerders om de verspreiding van desinformatie tegen te gaan.[139] Onder meer Facebook, Google, Twitter, TikTok en Microsoft hebben deze gedragscode ondertekend.

De *Code of Practice* bevat onder andere de volgende richtlijnen:
- Verbeteren van de transparantie rondom politieke advertenties;
- Versterken van het toezicht door online platforms op het gebruik van advertenties gericht op de verspreiding van desinformatie;
- Intensiveren van de inzet om nepaccounts te verwijderen;
- Opzetten van een duidelijk markeringssysteem voor bots, zodat hun activiteiten niet kunnen worden verward met menselijk handelen;
- Toegang tot data voor factchecking en onderzoek.

De gedragscode wordt gemonitord, als onderdeel van het bovengenoemde Europese actieplan tegen desinformatie.

6.3 Maatregelen platformbedrijven

Verschillende platformbedrijven hebben maatregelen genomen of in gang gezet om de productie en verspreiding van desinformatie tegen te gaan. Hieronder noemen we van enkele grote spelers een aantal maatregelen. Vervolgens gaan we kort in op de reacties van deze bedrijven op de toename van desinformatie die is ontstaan naar aanleiding van de coronapandemie.

Facebook heeft onder meer een advertentiebeleid opgesteld met als doel om misleidende of valse content te weren. Hiervoor is een *advertising approval process* ingesteld, waarin afbeeldingen, tekst en plaats van een advertentie worden beoordeeld. Daarnaast worden regelmatig nepaccounts verwijderd, om kunstmatige beïnvloeding van het aanbevelingsalgoritme en de verspreiding van desinformatie te voorkomen. Ook heeft Facebook een onafhankelijk factcheckprogramma ingesteld om misleidende nieuwsberichten te detecteren, om ze vervolgens niet langer aan te bevelen.[140] Voor het factchecken van Nederlandse berichten werkt

[139] European Commission (2018). *Code of Practice on Disinformation* https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation
[140] Mosseri, A. (2017). *Working to Stop Misinformation and False News.* Facebook www.facebook.com/facebookmedia/blog/working-to-stop-misinformation-and-false-news

Facebook samen met Agence France Presse (AFP) en Deutsche Presse-Agentur (DPA).[141]

Twitter heeft een advertentiebeleid opgesteld waarin de verspreiding van 'disruptieve content' wordt verboden. Adverteerders op Twitter moeten aan bepaalde criteria voldoen en een reviewproces ondergaan. In 2018 heeft Twitter maatregelen genomen om nepaccounts en spamberichten tegen te gaan, bijvoorbeeld door strikter toe te zien op het geautomatiseerd gebruik van het platform door vermoedelijke bots. Daarnaast biedt Twitter gebruikers meer inzicht in waarom hun bepaalde advertenties worden getoond.[142]

Verder hebben Facebook, Google en Twitter in 2017 laten weten te gaan werken met *trust indicators*, ontwikkeld door het Trust Project van het Santa Clara Institute of Applied Ethics.[143] Facebook gebruikt sindsdien factcheckers ter bestrijding van desinformatie.[144] Een bericht dat wordt beoordeeld als 'onwaar', wordt hierdoor lager in het nieuwsaanbod geplaatst.[145] Instagram voegt waarschuwingslabels toe aan berichten die als misleidend of onjuist worden beoordeeld.[146]

TikTok heeft sinds 2020 een richtlijn om politieke desinformatie te bestrijden.[147] Het platform is erg populair onder jongeren en wordt bijvoorbeeld gebruikt voor het delen van politieke memes. TikTok zegt ook desinformatiecampagnes op het platform te bestrijden.

Maatregelen naar aanleiding van de coronapandemie

Naar aanleiding van de hausse aan misleidende berichten rondom de coronacrisis hebben platformbedrijven recentelijk gehoor gegeven aan de groeiende publieke en politieke druk om strenger op te treden tegen desinformatie.

---

[141] Faccebook (2020). *Update op 26 maart: Facebook kondigt factchecking-partners in Nederland aan* https://facebook.pr.co/187141-corona-nieuwsoverzicht

[142] Twitter (2019). *Twitter Progress Report: Code of Practice on Disinformation*. https://ec.europa.eu/information_society/newsroom/image/document/2019-5/twitter_progress_report_on_code_of_practice_on_disinformation_CF162219-992A-B56C-06126A9E7612E13D_56993.pdf

[143] The Trust Project (2020). *News with integrity* https://thetrustproject.org

[144] Belghmidi, L. (2019). *Facebook bestrijdt samen met 21 Europese organisaties nepnieuws in aanloop naar verkiezingen.* VRT www.vrt.be/vrtnws/nl/2019/04/26/facebook-bestrijdt-samen-met-21-europese-organisaties-nepnieuws/

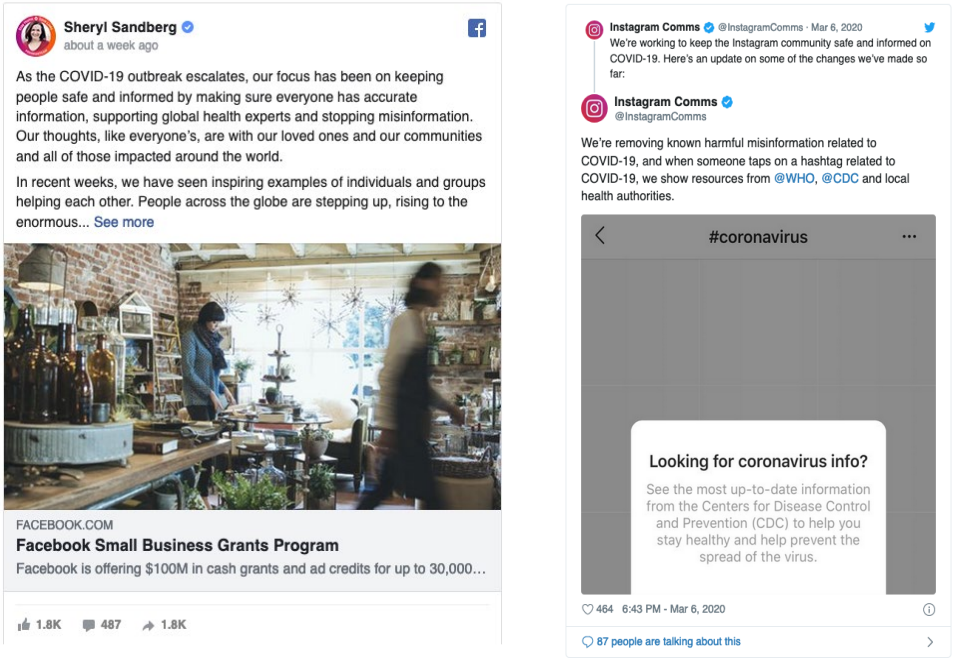
[145] Kreijeveld, M. (2018). *De strijd tegen nepnieuws (3): Hoe Facebook, Google en Twitter fake news niet kunnen bestrijden*. Marketingfacts www.marketingfacts.nl/berichten/strijd-tegen-nepnieuws-3-hoe-facebook-google-twitter-fake-news-bestrijden

[146] Van Poorten, B. (2020). *Social media update: Snapchat Scan-advertenties en Instagram tegen fake news.* Marketingfacts. www.marketingfacts.nl/berichten/social-media-update-snapchat-scan-advertenties-en-instagram-tegen-fake-news

[147] Nuñez, M. (2020). *TikTok Finally Bans Disinformation Campaigns In Updated Community Guidelines.* Forbeshttps www.forbes.com/sites/mnunez/2020/01/08/tiktok-finally-bans-disinformation-campaigns-in-updated-community-guidelines/

Zo verbood Twitter het om aanbevelingen van gezondheidsautoriteiten tegen te spreken, verwijderde Reddit desinformatie over corona uit zoekresultaten, plaatste Google waarschuwingen en officiële WHO-informatie bovenaan de zoekresultaten, richtte Facebook een *Coronavirus Information Center* op en plaatste Instagram overheidsinformatie bovenaan de berichtgeving (zie figuur 10). Facebook doneerde tevens een miljoen dollar aan het International Fact-Checking Network om daarmee factchecking op WhatsApp te ondersteunen. Ook ging – zoals hierboven reeds vermeld – een samenwerking van start tussen Facebook Nederland en Agence France Presse en Deutsche Presse-Agentur, gericht op factchecken.[148]

Uit deze maatregelen valt op te maken dat de internetbedrijven zich, in ieder geval wat de berichtgeving rondom de coronacrisis betreft, sterker dan voorheen verantwoordelijk achten voor de kwaliteit van de informatie die zij verspreiden. Het is uiteraard de vraag hoe blijvend de getroffen maatregelen zijn, met het risico dat ze worden afgezwakt zodra de publieke en politieke druk weer afneemt.



Figuur 10 Aankondiging maatregelen door socialemediabedrijven naar aanleiding van de coronapandemie.

Los van de coronacrisis neemt de maatschappelijke en politieke druk op socialemediaplatforms om meer verantwoordelijkheid te nemen voor de informatie die zij verspreiden, overigens toe. De recent bijgewerkte Europese richtlijn voor

[148] NOS (2020). *Facebook zet voor Nederland factcheckers in om coronanepnieuws te bestrijden.* https://nos.nl/l/2328458

audiovisuele media (AVMSD) legt videoplatforms bijvoorbeeld verplichtingen op voor het verwijderen van illegale informatie, zoals kinderporno, terrorisme of misleiding van consumenten.[149]

[149] European Commission (2020). *Audiovisual Media Services Directive (AVMSD)* https://ec.europa.eu/digital-single-market/en/audiovisual-media-services-directive-avmsd

Deel II Casestudies

7 Deepfakes en psychographing

Voortbouwend op de quickscan in deel I zijn twee casestudies uitgewerkt, over *deepfakes* en *psychographing*. Met de casestudies willen we een meer samenhangend beeld schetsen van hoe technologische ontwikkelingen op het gebied van desinformatie de komende jaren zouden kunnen uitpakken, en welke mogelijke impact dat kan hebben op het publieke debat en het democratisch proces.

Voor de keuze van de casestudies is gekeken naar combinaties van technologieën:

- die innovatief zijn, dat wil zeggen: die nog volop in ontwikkeling zijn;
- waarvan de te verwachten potentiële impact groot is;
- die gepaard gaan met een asymmetrie in handelingsvermogen tussen producenten en verspreiders van desinformatie aan de ene kant, en ontvangers daarvan aan de andere kant.

Per casestudie wordt een technologische stand van zaken beschreven en ingegaan op te verwachte ontwikkelingen. Door middel van een impactscenario wordt een beeld geschetst van de mogelijke gevolgen van de beschreven ontwikkelingen voor het publieke debat en het democratisch proces.

7.1 Casestudie deepfakes

7.1.1 Stand van zaken

Zoals ook al in de quickscan kort is beschreven, kan kunstmatige intelligentie (KI) worden ingezet voor de bewerking en manipulatie van bestaand audiovisueel materiaal. Een eenvoudig en veelgebruikt voorbeeld hiervan zijn de camera-apps op smartphones die portretfoto's met *beautyfilters* bewerken. De toepassing van kunstmatige intelligentie in audio- en videomateriaal is de laatste jaren zo sterk ontwikkeld dat het steeds lastiger wordt om bewerkte informatie te onderscheiden van onbewerkte, authentieke informatie.[150] [151] [152]

[150] Brundage, M., et al. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. *arXiv:1802.07228 [cs]*. http://arxiv.org/abs/1802.07228
[151] Khodabakhsh, A., Busch, C., & Ramachandra, R. (2018). A Taxonomy of Audiovisual Fake Multimedia Content Creation Technology. *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 372–377. https://doi.org/10.1109/MIPR.2018.00082
[152] U. S. Government Accountability Office (2020). *Science & Tech Spotlight: Deepfakes, (GAO-20-379SP).* www.gao.gov/products/gao-20-379sp

Voor het produceren van *deepfakes* kunnen diverse KI-methoden worden gebruikt. Het gaat hierbij om de bewerking of manipulatie van videomateriaal. De hiervoor gebruikte software analyseert grote hoeveelheden beeldmateriaal van een persoon en leert met behulp daarvan de vorm, verhoudingen en bewegingen van iemands gezicht kennen. De producent bepaalt vervolgens welke houdingen moeten worden aangenomen in de bewerkte video. Het *deepfake*-algoritme genereert de bewerkte video beeldje voor beeldje. Vaak wordt dit gecombineerd met gemanipuleerde audio. Zodoende ontstaat een natuurgetrouwe video, die op het eerste gezicht lastig is te onderscheiden van een authentieke video.

De *face swap*-techniek, waarmee gezichten kunnen worden verwisseld, is het bekendste voorbeeld van *deepfake*-technologie. Andere voorbeelden zijn *lip sync*-technologie, waarmee in een gezicht mondbewegingen kunnen worden gemanipuleerd, en *digital puppetry*, waarmee een kunstmatig hoofd of lichaam kan worden gegenereerd. Met *personalized avatar creation*-technologie kan zelfs iemands volledige lichaam over videobeelden van een bestaand persoon worden geplakt.[153]

Recent zijn er diverse apps verschenen die gebruikmaken van *deepfake*-technologie. Voorbeelden hiervan zijn:

- Face2Face, waarmee een video kan worden gemanipuleerd door het gezicht in de video de gezichtsuitdrukkingen van een andere persoon realtime te laten immiteren;[154]
- Mug Life, waarmee een gezicht in een 3D-animatie kan worden veranderd;[155]
- Doublicat, waarmee een gezicht in een GIF kan worden geplaatst;[156] en
- HeadOn, waarbij realtime in video's gezichten, bewegingen en gezichtsuitdrukkingen kunnen worden vervangen.[157]

Op het eerste gezicht lijken de genoemde voorbeelden onschuldig van karakter, bijvoorbeeld omdat met een grappige filter videobellen in tijden van quarantaine en lockdown leuker kan worden gemaakt, of het eigen gezicht in grappige animaties kan worden verwerkt. Maar het gebruik van *deepfakes* kent ook minder onschuldige toepassingen. Zo heeft de klimaatactiegroep Extinction Rebellion met behulp van *lip sync*-technologie een *deepfake*-video gepubliceerd van een fictieve speech van

[153] Duursma, J. (2019). *Deepfake Technologie – The Infocalypse* www.jarnoduursma.nl/wp-content/uploads/2019/09/Jarno-Duursma-_-Deepfake-Technologie-The-Infocalypse.pdf
[154] Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016). Face2Face: Real-Time Face Capture and Reenactment of RGB Videos. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2387–2395. https://doi.org/10.1109/CVPR.2016.262
[155] MugLife (2020). *Bring Your Photos to Live.* www.muglife.com/
[156] Reface (2020). *The Best Face Swap App.* https://reface.app/
[157] Thies, J., Zollhöfer, M., Theobalt, C., Stamminger, M., & Nießner, M. (2018). HeadOn: Real-time Reenactment of Human Portrait Videos. *ACM Transactions on Graphics, 37*(4), 1–13. https://doi.org/10.1145/3197517.3201350

de Belgische premier, waarin zij spreekt over een verband tussen de uitbraak van pandemieën en verstoring van de natuurlijke omgeving door de mens.[158]

Figuur 11 Een met behulp van Mug Life gemanipuleerde video van de Amerikaanse presidentskandidaat Joe Biden.

De ontwikkelingen rondom *deepfakes* zijn volop in ontwikkeling, en er zijn veel nieuwe apps en methoden in opkomst. De hier genoemde voorbeelden zijn slechts een greep uit de vele toepassingen die momenteel voorhanden zijn.

De verspreiding van *deepfakes* vindt voornamelijk plaats via socialemediaplatforms.

7.1.2 Verwachte ontwikkelingen

Het valt te verwachten dat de productie en verspreiding van *deepfakes* een steeds wijdverspreider fenomeen zal worden. Niet alleen zullen *deepfakes* door verdere technologische innovatie steeds lastiger te onderscheiden zijn van authentieke

[158] BELGA (2020). *Extinction Rebellion publiceert deepfake-video met alternatieve speech premier Wilmès* Nieuwsblad.be www.nieuwsblad.be/cnt/dmf20200414_04921988

beelden, ook zal *deepfake*-technologie toegankelijker worden doordat technologiebedrijven de steeds geavanceerdere technologie in eenvoudig te gebruiken apps en gadgets op de markt zullen brengen. Veel gewone gebruikers worden hierdoor in staat gesteld zelf *deepfakes* te produceren.[159] Het maken van *deepfakes* wordt een techniek die iedereen met een zekere mate van computervaardigheid kan toepassen.

Volgens het cybersecuritybedrijf Nisos zijn *deepfakes* op dit moment technisch gezien nog niet ver genoeg ontwikkeld om voor criminele doeleinden te kunnen worden gebruikt. Ze worden dan ook nog niet als service te koop aangeboden op het *dark web*. De technologische ontwikkelingen gaan echter snel en Nisos verwacht dat *deepfakes* op afzienbare termijn van zodanige kwaliteit zijn dat ze wel interessant worden voor criminele doeleinden.[160] Grote technologiebedrijven als Apple en Amazon zijn inmiddels druk bezig om de technologie achter *deepfakes* te vervolmaken. Het zou zomaar kunnen dat de *deepfake*-technologie die hieruit voortkomt ruimschoots voldoet aan de kwaliteitsmaatstaven van criminele groeperingen.

Toenemende populariteit gemanipuleerd beeldmateriaal

De populariteit van platforms die gericht zijn op het uitwisselen van beeldmateriaal, zoals YouTube, SnapChat, Instagram en TikTok, zal naar verwachting verder toenemen. Deze trend is vooral zichtbaar onder jongeren. De platforms bieden ook *augmented reality*-functionaliteiten aan, waarmee informatie realtime kan worden toegevoegd aan live camerabeelden. SnapChat voorziet een toekomst waarin de live camerabeelden van smartphones de standaardmodus worden voor het gebruik van smartphones.[161]

Op deze platforms wordt het steeds normaler om beelden te manipuleren. SnapChat en Instagram moedigen gebruikers ook aan om filters toe te passen, beelden te combineren of deze te voorzien van teksten en stickers. *Deepfake*-technieken zoals het verwisselen of verouderen van gezichten, het veranderen van gendereigenschappen en het manipuleren van stemgeluid, maken hier deel van uit. Deze platforms bieden ook de mogelijkheid om live beelden uit te zenden. Met de komst van snellere mobiele internetverbindingen, zoals 5G, zullen *deepfake*-technieken ook steeds vaker realtime worden toegepast.

---

[159] Schulz, J. (2020). *The Deepfake iPhone Apps Are Here*. Lawfare. www.lawfareblog.com/deepfake-iphone-apps-are-here

[160] Volkert, R. (2020). *Deep Fakes: Understanding the illicit economy for synthetic media.* NISOS. https://cdn2.hubspot.net/hubfs/6068438/Resources/NISOS%20-%20Deep%20Fakes%20White%20Paper.pdf

[161] Hern, A. (2020). Snapchat firm unveils platform plan to take on Google and Apple. *The Guardian*. www.theguardian.com/technology/2020/jun/15/snapchat-firm-unveils-platform-plan-to-take-on-google-and-apple

Toenemend belang van beeld in nieuwsvoorziening
De impact van *deepfakes* wordt groter naarmate het belang van beeld in de nieuwsvoorziening toeneemt. Zeker op het internet valt die trend waar te nemen. De nieuwsconsumptie van jongere generaties verschuift meer en meer naar digitale bronnen, inclusief berichtgeving op platforms als Facebook. En er is een groeiende beeldcultuur, gezien de populariteit van internetdiensten YouTube, TikTok, Instagram en Snapchat. Nieuwsmedia zullen zich naar verwachting ook meer op die platforms gaan richten.

Mediawijsheid laat te wensen over
Het groeiende belang van beeld in de nieuwsvoorziening in combinatie met de grotere toegankelijkheid van *deepfake*-technieken en de toenemende populariteit van gemanipuleerd beeldmateriaal, is problematisch omdat een aanzienlijk deel van de Nederlandse bevolking nu al moeite heeft om uit te maken of een nieuwsbericht echt of nep is. Zo meldt de Volkskrant dat slechts 29% van de Nederlanders zegt 'echt nieuws van nepnieuws te kunnen onderscheiden' en dat een op de drie Nederlanders aangeeft 'tegenwoordig vaak niet meer te weten wat waar is en wat onwaar'.[162] Daarnaast geeft 33% van de bevolking aan nooit het nieuws op juistheid te toetsen.[163]

Volgens de Monitor Jeugd en Media laten de digitale kennis en vaardigheden van jongeren vaak te wensen over. Er worden grote verschillen geconstateerd in digitale geletterdheid tussen de leerlingen uit verschillende onderwijstypen. Vooral bij leerlingen uit het praktijkonderwijs en het vmbo is sprake van een laag niveau.[164]

7.1.3 Impactscenario

Geloofwaardigheid beeldmateriaal erodeert
Om een beeld te schetsen van de mogelijke impact van de hierboven beschreven ontwikkelingen voor het publieke debat en het democratisch proces, beschrijven we in deze paragraaf een mogelijk toekomstig scenario. Hiervoor gaan we niet uit van het risico dat bijvoorbeeld een interview met premier Rutte op het NOS-journaal kan worden gehackt met behulp van *deepfake*-technieken. We achten de kans namelijk groot dat in reactie hierop het journaal zal worden onderbroken, of de berichtgeving door de NOS publiekelijk zal worden weersproken – waardoor het effect van de hack waarschijnlijk beperkt blijft.

---

[162] Kranenberg, A. (2017). *Nederlanders bezorgd over 'nepnieuws' - een op drie weet vaak niet meer wat waar is en wat onwaar.* Volkskrant www.volkskrant.nl/nieuws-achtergrond/nederlanders-bezorgd-over-nepnieuws-een-op-drie-weet-vaak-niet-meer-wat-waar-is-en-wat-onwaar~b6914596/
[163] Consultancy.nl (2018). *Nederlanders herkennen nepnieuws en maken zich niet zo druk om fake news* www.consultancy.nl/nieuws/17892/nederlanders-herkennen-nepnieuws-en-maken-zich-niet-zo-druk-om-fake-news
[164] Pijpers, R. (2019). *Werken aan digitale geletterdheid: van visie naar praktijk.* Kennisnet www.kennisnet.nl/publicaties/werken-aan-digitale-geletterdheid-van-visie-naar-praktijk/

De verwachting is dat *deepfakes* een veel groter en diffuser effect kunnen hebben wanneer ze buiten het publieke zicht worden verspreid op informele en besloten kanalen zoals privé chatgroepen, ongecensureerde socialemediaplatforms als Parler of forums als Reddit of 8chan. Daar vindt geen of weinig moderatie plaats en is de kans aanzienlijk kleiner dat berichtgeving wordt weersproken.

Het verspreiden van *deepfakes* op informele en besloten kanalen, kan tevens een reactie zijn op de toenemende detectie van *deepfakes* met behulp van kunstmatige intelligentie door grote platforms. Zo is Facebook in 2019 de *Deepfake Detection Challenge* gestart.[165]

Deepfakes worden in dit scenario door diverse groeperingen geproduceerd, zowel door professionele trollen als door minder professionele clubjes, onruststokers en hobbyisten. Doordat de daarvoor benodigde technologie steeds eenvoudiger is te gebruiken, is een steeds grotere groep mensen in staat om zelf *deepfakes* te produceren. Dat betekent ook dat via allerlei platforms en kanalen *deepfakes* kunnen worden verspreid. Professionele organisaties kunnen hiervan ook uitgebreid gebruikmaken. Ze kunnen de gemakkelijkste weg zoeken om *deepfakes* te verspreiden om maximaal effect te bereiken.

Het toenemende gemak waarmee *deepfakes* kunnen worden geproduceerd en verspreid via allerlei niet-gemodereerde, informele en besloten kanalen, kan leiden tot een ware proliferatie van *deepfakes* op het internet. Doordat het tegelijkertijd voor ontvangers steeds lastiger wordt om gemanipuleerde berichtgeving te onderscheiden van authentieke berichtgeving, kan dat er op den duur, sluipenderwijs, toe leiden dat het verschil tussen authentiek en gemanipuleerd materiaal aan betekenis verliest. Een neveneffect hiervan kan zijn dat ook beeldmateriaal dat afkomstig is van gevestigde media kan worden afgedaan als gemanipuleerd of nep. En dat kan weer leiden tot afname van het vertrouwen in die media. De geloofwaardigheid van beeldmateriaal erodeert als het ware door de voortdurende blootstelling aan gemanipuleerde beelden.

[165] Facebook (2020). *Deepfake Detection Challenge* https://deepfakedetectionchallenge.ai/

7.2 Casestudie psychographing

7.2.1 Stand van zaken

Een *psychograph* is van origine een visuele representatie van de persoonlijkheidskenmerken van een persoon of groep.[166] Het gaat daarbij om kenmerken als waarden, verlangens, doelen, interesses en lifestyle. Marketeers gebruiken *psychographing* als techniek om reclame af te stemmen op een doelgroep. Niets weerhoudt producenten en verspreiders van desinformatie ervan om deze technologie te gebruiken.

Met de term *psychographing* doelen we in deze casestudie niet alleen op visuele representaties van persoonlijkheidskenmerken, maar ook op een verzameling digitale technieken die kan worden ingezet om berichten af te stemmen op de persoonlijkheidskenmerken van een doelgroep. Zoals we ook al in deel I aangaven, kan *psychographing* worden beschouwd als een geavanceerde vorm van micro-targeting.

Pyschographing kent een lange historie. Decennialang was het hoofdzakelijk gebaseerd op traditioneel doelgroepenonderzoek, met behulp van enquêtes, interviews en focusgroepen. Op grond hiervan kunnen doelgroepen worden onderverdeeld in subgroepen, bijvoorbeeld aan de hand van het psychologische *Five-Factor Model*. Dit model onderscheidt persoonlijkheidskenmerken als emotionele stabiliteit, extraversie, intellectuele autonomie en ordelijkheid.[167] Als bijvoorbeeld bekend is dat een doelgroep laag scoort op extraversie, kan deze het beste worden bereikt met een rustige boodschap, zo is de gedachte achter dit model.

Nieuwe digitale technologieën maken het mogelijk om het doelgroepenonderzoek en het afstemmen van boodschappen op doelgroepen te automatiseren.[168] Zo claimt IBM dat het Watson-algoritme in staat is om persoonlijkheidskenmerken uit teksten te halen.[169] Een bedrijf als Indivizo maakt van IBM's diensten gebruik om persoonlijkheidskenmerken af te leiden uit een sollicitatievideo.[170] Cambridge Analytica claimde tijdens de Amerikaanse presidentsverkiezingen van 2016 met

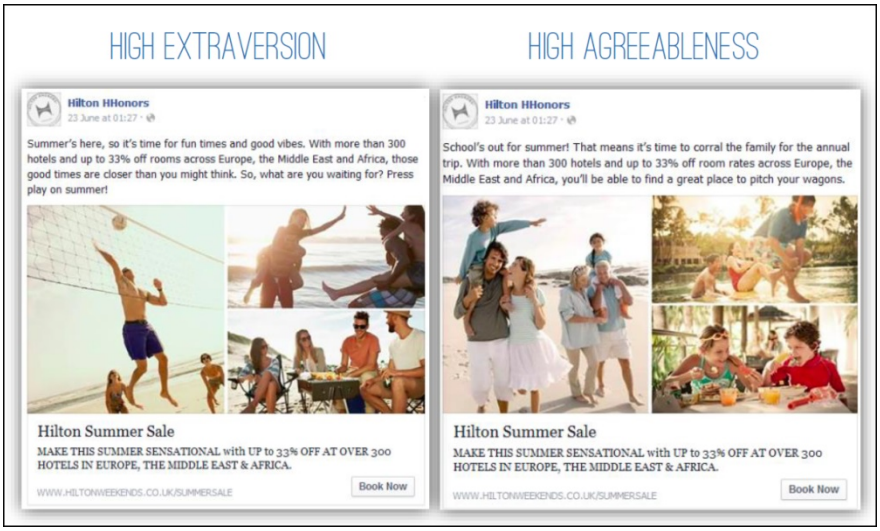
---

[166] Wells, William D. (1975). 'Psychographics: A critical review'. *Journal of Marketing Research*. 12: 196–213. doi:10.2307/3150443. JSTOR 3150443.
[167] McCrae, R. R.; Costa, P. C.; Jr (1987). 'Validation of the five-factor model across instruments and observers'. *Journal of Personality and Social Psychology*. 52 (1): 81–90. doi:10.1037/0022-3514.52.1.81. PMID 3820081.
[168] Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*(4), 1036–1040. https://doi.org/10.1073/pnas.1418680112
[169] IBM (z.d.). *Personality Insights* www.ibm.com/watson/services/personality-insights/
[170] Indivizo (z.d.). *AI-based Personality Profiles* www.indivizo.com/personality-profiles

behulp van de Facebookgegevens van 87 miljoen gebruikers een effectieve politieke campagne te kunnen voeren. Het bedrijf raakte in opspraak toen bleek dat de gebruikers geen toestemming hadden gegeven voor dit gebruik van hun data.[171]

Automatisering maakt het mogelijk om *psychographing* op grote schaal toe te passen. De gedachte daarachter is dat mensen kunnen worden beïnvloed in hun (politieke) gedachtevorming doordat ze informatie voorgeschoteld krijgen die is afgestemd op hun psychologische kenmerken, inclusief hun psychische kwetsbaarheden.

De afgelopen decennia is de effectiviteitsclaim van toepassing van *psychographing* in traditionele media onder vuur komen te liggen. Ook de effectiviteit van de nieuwe, geautomatiseerde vormen worden betwist. Onderzoek laat wel zien dat de technieken ertoe kunnen leiden dat mensen vaker links aanklikken op websites of online aankopen doen (zie figuur 12). Maar het is niet bekend of de techniek ook effect heeft op bijvoorbeeld politieke voorkeuren of stemgedrag.[172] [173]



Figuur 12 Advertenties op Facebook voor Hilton-vakanties waarin afbeeldingen en tekst zijn aangepast op basis van de ingeschatte persoonlijkheidskenmerken van de doelgroep.

[171] Lapowsky, I. (2018). *Facebook Exposed 87 Million Users to Cambridge Analytica.* Wired www.wired.com/story/facebook-exposed-87-million-users-to-cambridge-analytica/
[172] Rokka, J. & Airoldi, M. (2018). *Cambridge Analytica's 'secret' psychographic tool is a ghost from the past.* The Conversation. https://theconversation.com/cambridge-analyticas-secret-psychographic-tool-is-a-ghost-from-the-past-94143
[173] Resnick, B. (2018). *Cambridge Analytica's "psychographic microtargeting": what's bullshit and what's legit.* Vox. www.vox.com/science-and-health/2018/3/23/17152564/cambridge-analytica-psychographic-microtargeting-what

Onderzoek door Cambridge University laat zien dat deze advertenties effectiever waren dan advertenties gericht op een algemeen kenmerk zoals houden van reizen.[174]

7.2.2 Verwachte ontwikkelingen

In deze casestudie gaan we ervan uit dat het gebruik van *psychographing*-technieken in ieder geval tot op zekere hoogte effectief is. De technologieën lijken immers nu al de aandacht van personen te kunnen sturen. Het is dan ook goed voorstelbaar dat bij een verdere ontwikkeling van de technologie ook de publieke beeldvorming rond maatschappelijke of politieke thema's kan worden beïnvloed. In dat geval valt ook te verwachten dat meer partijen er gebruik van gaan maken.

Daar komt bij dat van internetgebruikers steeds meer gegevens online worden verzameld, waaruit allerlei persoonlijkheidskenmerken kunnen worden afgeleid. Zo maakt de opkomst van het *Internet of Things*, waarbij steeds meer apparaten zoals smart-tv's en zelfrijdende auto's met het internet verbonden worden, het mogelijk om steeds meer data te vergaren over het offlinegedrag van mensen. Het toenemende gebruik van biometrische sensoren, onder andere in AR-/VR-apparatuur, maakt het mogelijk om bijvoorbeeld oog- en pupilbewegingen te monitoren. Daarmee kan vervolgens meer inzicht worden verkregen in karaktertrekken en persoonlijke voorkeuren. Op basis daarvan kunnen *psychographing*-technieken verder worden verfijnd.

7.2.3 Impactscenario

We schetsen in deze paragraaf een scenario dat uitgaat van een actor die beschikt over geavanceerde technische middelen en de motivatie om een langdurige (desinformatie)campagne te voeren. Hierbij is sprake van een zogenoemde *Advanced Persistent Threat* (APT). Vanwege haar technische vermogens en slagkracht bestaat vaak het vermoeden dat een groepering achter een APT gelieerd is aan, of gesteund wordt door een statelijke actor.

In deze casestudie gaan we ervan uit dat een APT-groepering zich tot doel stelt om het publieke debat en het democratische proces met behulp van *psychographing*-technieken heimelijk te beïnvloeden. Door in te spelen op maatschappelijk gevoelige kwesties beoogt de groepering maatschappelijke tegenstellingen aan te

[174] LaMontagne, L. (2015). *Personality-Matched Ads: How Hilton Worldwide effectively personalized its marketing messages.* MarketingExperiments https://marketingexperiments.com/digital-advertising/hilton-worldwide-personality-matched-ads

wakkeren en het vertrouwen van burgers in de gevestigde instituties te ondermijnen.

Om dit doel te bereiken zet de APT-groepering in op beïnvloeding van personen die sterk emotioneel reageren op bepaalde maatschappelijke onderwerpen of wantrouwend staan tegenover de gevestigde orde. Om die personen te vinden wordt gebruik gemaakt van geautomatiseerde vormen van *psychographing*, met behulp van kunstmatige intelligentie. De gebruikte algoritmes worden gedurende de langlopende campagne telkens opnieuw getraind met gegevens van geselecteerde personen, waardoor ze ook steeds beter worden in het vinden van personen met een vergelijkbaar karakter.

De personen worden gevonden door grote databases te doorzoeken. Deze kunnen worden gekocht van tussenpersonen uit de advertentiebranche. Maar data kunnen ook worden verzameld uit publieke bronnen zoals sociale media, nieuwsmedia of openbare databronnen van de overheid. Tevens kunnen de gegevens zijn gestolen door middel van hacks, of verzameld uit datalekken. De Algemene verordening gegevensbescherming (AVG) biedt geen bescherming in dit scenario, omdat een APT-groepering heimelijk opereert en zich weinig van de AVG hoeft aan te trekken.

De berichten die de groepering verspreidt, worden op zo'n manier vormgegeven dat ze zich gemakkelijk lenen voor het delen op sociale media. Er wordt dus bij voorkeur gebruik gemaakt van korte tweets, posts, filmpjes of pakkende afbeeldingen. De APT-groepering is er niet alleen op uit om boodschappen te verzenden, maar ook om personen aan te zetten tot het verder verspreiden van de desinformatie.

De verspreiding vindt primair plaats via socialemediaplatforms. De (des)informatie wordt af ten toe ook opgepikt door de gevestigde media, omdat die hun berichtgeving steeds vaker mede laten bepalen door wat populair is op Twitter of YouTube.

Om maximaal onrust te veroorzaken en wantrouwen te kweken, zet de APT-groepering vooral in op de verspreiding van berichten via niet-publieke kanalen, zoals privégroepen op Facebook of Telegram. Daar is de kans immers klein dat de berichten worden tegengesproken, wat het effect van de desinformatiecampagne groter maakt. Bovendien is het zo mogelijk om tegenpolen in het maatschappelijk debat te voorzien van tegenstrijdige informatie, zonder dat zij dat door hebben.

Door de afzonderlijke doelgroepen van verschillende, op hun persoonlijkheidskenmerken afgestemde boodschappen te voorzien, worden maatschappelijke tegenstellingen aangewakkerd. Dat leidt tot polarisering van het publieke debat en een groeiende onwil om met de ander in gesprek te gaan, wat schadelijk is voor het democratisch proces.

De APT-groepering kan hierbij overigens handig inspelen op de maatregelen die platformbedrijven nemen om desinformatie te bestrijden. Zo kunnen onthullingen door factcheckers van desinformatie die rondgaan in het ene kamp, door de APT-groepering gebruikt worden om het andere kamp te laten zien hoe naïef hun opponent is, en daarmee de tweedracht versterken.

De APT-groepering kan ook inspelen op de moderatiemaatregelen van platformbedrijven. Berichten die worden beoordeeld als desinformatie kunnen bijvoorbeeld minder worden aanbevolen door de algoritmes, of worden verwijderd. De APT-groepering kan deze maatregelen als een vorm van censuur brandmerken, en daarmee het wantrouwen in gevestigde partijen verder versterken.

Deel III Vooruitblik

8 Nieuwe maatregelen

In dit hoofdstuk beschrijven we welke nieuwe maatregelen kunnen worden genomen om schade aan het publieke debat en het democratisch proces als gevolg van technologische ontwikkelingen op het gebied van desinformatie tegen te gaan, en welke actoren daarvoor verantwoordelijk zijn. We bouwen hierbij in belangrijke mate voort op de casestudies uit hoofdstuk 7 en de opbrengst van de expertmeeting van 2 juni 2020. Tevens is hiervoor aanvullend literatuuronderzoek verricht.

We richten ons hierbij op maatregelen die geen inbreuk maken op de vrijheid van meningsuiting en de persvrijheid. Zo mag de overheid misinformatie niet enkel en alleen verwijderen op grond van de misleidende aard ervan. Daarvoor zijn aanvullende juridische redenen nodig. Verwijdering van desinformatie zou anders in strijd zijn met de vrijheid van meningsuiting.

Zoals we ook in het inleidend hoofdstuk schrijven, gaat het ons in dit onderzoek niet om geheel nieuwe, nu nog onbekende technologische ontwikkelingen. Die ontwikkelingen kennen we namelijk niet. In de voorgaande hoofdstukken hebben we in beeld gebracht hoe technologische innovaties die zich nu reeds aftekenen, verder vorm zouden kunnen krijgen, en welke impact ze zouden kunnen hebben op de productie en verspreiding van desinformatie. Vanwege de snelle technologische ontwikkelingen op het gebied van IT zijn deze ontwikkelingen moeilijk te voorspellen. De beschrijving ervan krijgt daarmee tot op zekere hoogte een speculatief karakter.

Zo bestaat er nog de nodige onduidelijkheid over de effectiviteit van desinformatiecampagnes die gebruikmaken van geavanceerde vormen van micro-targeting. Terwijl in de reclamewereld al veel gebruik wordt gemaakt van deze toepassingen, is de effectiviteit ervan niet afdoende aangetoond. Tegelijkertijd valt niet uit te sluiten dat het gebruik van micro-targeting voor desinformatiedoeleinden voldoende effectief kan uitpakken om als een bedreiging voor het publieke debat en het democratisch proces te kunnen worden beschouwd.

Hieronder volgt een overzicht van de belangrijkste nieuwe maatregelen die zouden kunnen worden genomen om het hoofd te bieden aan zich mogelijk voordoende dreigingen die uitgaan van technologische ontwikkelingen op het gebied van desinformatie.

8.1 Maatregelen tegen wijdverspreide deepfakes

De casestudie over *deepfakes* in deel II laat zien dat een wijdverspreide productie en verspreiding van *deepfakes* kan leiden tot een sluipende ondermijning van het onderscheid tussen authentiek en gemanipuleerd beeldmateriaal.

Als het voor burgers steeds lastiger wordt om nep van echt te onderscheiden, kan dat leiden tot een onverschillige houding ten aanzien van dit onderscheid: misschien is alles wat je op het internet ziet wel een beetje waar, en is niets of weinig helemaal waar. Een gevaar hiervan is dat ook berichtgeving van gevestigde media en overheidsinstanties niet langer als betrouwbaar wordt gezien. Ook die berichtgeving kan immers zijn gemanipuleerd.

Maatregelen die gericht zijn op vergroting van de mediawijsheid kunnen weliswaar het besef doen toenemen dat niet alles wat op internet te zien en te beluisteren valt, waar is. Maar het is niet aannemelijk dat de meeste mensen ook in staat zullen zijn om authentiek van gemanipuleerd beeldmateriaal te onderscheiden. Doordat de manipulatie met behulp van kunstmatige intelligentie steeds geraffineerdere vormen aanneemt, zullen mensen hun eigen oren en ogen niet meer kunnen vertrouwen. Daarmee kom het uitgangspunt van de overheid onder druk te staan, dat burgers zelf (des)informatie op waarde kunnen schatten.

8.1.1 Detectie gemanipuleerd beeldmateriaal

Er zouden dan ook handvatten moeten worden ontwikkeld waarmee burgers beter in staat worden gesteld om het onderscheid te maken tussen authentiek en gemanipuleerd beeldmateriaal. Bijvoorbeeld doordat socialemediabedrijven met behulp van kunstmatige intelligentie gemanipuleerd beeldmateriaal opsporen (detectie) en dit materiaal in hun berichtgeving markeren als (mogelijk) gemanipuleerd.

De algoritmes met behulp waarvan *deepfakes* worden gemaakt, bieden vaak aanknopingspunten voor detectietechnieken.[175] Zo blijken personen in gemanipuleerde video's vaak onnatuurlijke of geen oogknipperingen te vertonen. Dat kan met behulp van kunstmatige intelligentie worden gedetecteerd.[176] Er zijn

[175] Hameleers, M., Powell, T. E., Meer, T. G. L. A. V. D., & Bos, L. (2020). A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated via Social Media. *Political Communication*, *37*(2), 281–301. https://doi.org/10.1080/10584609.2019.1674979
[176] Li, Y., Chang, M.-C., & Lyu, S. (2018). In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking. *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, 1–7. https://doi.org/10.1109/WIFS.2018.8630787

diverse initiatieven in ontwikkeling die moeten leiden tot betere detectiesystemen. Zo zijn er datasets beschikbaar die ontwikkelaars van detectie-algoritmes kunnen gebruiken om hun systemen te testen en te trainen, zoals Faceforensics++.[177] Grote technologiebedrijven hebben daarnaast diverse initiatieven genomen om betere detectiemethoden te ontwikkelen. Zo neemt Google deel aan Reality Defender 2020[178], en werken Amazon, Facebook en Microsoft samen aan de Deepfake Detection Challenge (DFDC).[179] Ook werken diverse onderzoekers aan detectiestrategieën en -tools, zoals het detectiesysteem Poster.[180]

Detectie van gemanipuleerd beeldmateriaal vindt in de regel achteraf plaats, nadat het is verspreid. Die detectie heeft daardoor maar beperkt effect. Het zou effectiever zijn om aan de voorkant berichten op *deepfakes* te scannen en te filteren. Bovendien is het de verwachting dat *deepfakes* vaker zullen worden toegepast in live streaming, wat realtime detectie noodzakelijk maakt. In de nabije toekomst is het wellicht mogelijk om plug-ins te installeren in webbrowsers die real-time *deepfakes* kunnen detecteren en ook kunnen blokkeren.[181]

In reactie op de toegenomen detectiemogelijkheden, zouden producenten en verspreiders van *deepfakes* ertoe kunnen overgaan om gebruik te maken van de steeds geavanceerdere vormen van beeldmanipulatie die op de markt verschijnen, of deze zelf te ontwikkelen. Platformbedrijven zouden dan ook voldoende moeten blijven investeren in detectietechnologieën om te kunnen meekomen in de wedloop die ontstaat tussen producenten en verspreiders van steeds geavanceerde *deepfakes* aan de ene kant, en platformbedrijven aan de andere kant.

Meldpunt kwaadaardige beeldmanipulatie

De in de casestudie beschreven trend dat manipulatie van beeldmateriaal op platforms als SnapChat, Instagram en TikTok steeds normaler wordt, maakt het moeilijk om daar *deepfakes* te detecteren. Het wordt dan de opgave om goedaardige van kwaadaardige beeldmanipulatie te onderscheiden. Het is de vraag of daarvoor technologische oplossingen kunnen worden ontwikkeld die in voldoende mate in staat zijn om getoonde beelden en gesproken tekst contextueel te interpreteren, om vervolgens de mogelijk kwaadaardige beeldmanipulatie door menselijke moderatoren te laten beoordelen.

---

[177] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). *FaceForensics++: Learning to Detect Manipulated Facial Images*. www.niessnerlab.org/projects/roessler2019faceforensicspp.html
[178] Reality Defender 2020 (z.d.). https://rd2020.org/
[179] Facebook (2020). *Deepfake Detection Challenge* https://deepfakedetectionchallenge.ai/
[180] Sohrawardi, S.J. et al. (2019). Poster: Towards Robust Open-World Detection of Deepfakes. *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 2613–2615. https://doi.org/10.1145/3319535.3363269
[181] Duursma, J. (2019). *Deepfake Technologie – The Infocalypse* www.jarnoduursma.nl/wp-content/uploads/2019/09/Jarno-Duursma-_-Deepfake-Technologie-The-Infocalypse.pdf

Platforms zouden ook een meldpunt kunnen inrichten waar gebruikers (vermoedelijk) kwaadaardige beeldmanipulatie kunnen rapporteren. Moderatoren zouden daar vervolgens naar kunnen kijken en kwaadaardig bevonden materiaal kunnen markeren of verwijderen. Dat zou van deze bedrijven vragen dat ze voldoende investeren in moderatiecapaciteit.

Platformbedrijven primair verantwoordelijk
De verantwoordelijkheid voor de detectie van (kwaadaardige) *deepfakes* ligt primair bij de platformbedrijven. Vanwege de huidige wet- en regelgeving kan detectie van *deepfakes* niet door de overheid worden afgedwongen. Artikel 15 van de Europese e-commercerichtlijn staat het niet toe om dergelijke maatregelen aan internetbedrijven op te leggen.[182] Maar gezien het publieke belang dat gemoeid is bij betrouwbare berichtgeving, zou de overheid bij platformbedrijven kunnen aandringen op een actief detectiebeleid. Om voldoende gewicht in de schaal te kunnen leggen, zou dit bij voorkeur moeten gebeuren door de Europese Unie.

8.1.2 Waarmerken van beeldmateriaal

Een tweede handvat dat burgers kan helpen om authentiek van gemanipuleerd beeldmateriaal te onderscheiden, is het waarmerken ervan. Wanneer bijvoorbeeld gebruikt wordt gemaakt van digitale ondertekening, kunnen burgers gemakkelijker nagaan of materiaal afkomstig is van een betrouwbare informatiebron.[183] Dat geeft burgers die op zoek zijn naar betrouwbare berichtgeving in ieder geval enig houvast.

Een belangrijke randvoorwaarde daarbij is een betrouwbaar systeem om waarmerken te registreren. Een mogelijke valkuil is dat deze methode niet waterdicht is en dat mensen, al dan niet bewust, onbetrouwbare berichten digitaal ondertekenen en zo meewerken aan de verspreiding van desinformatie.

Diverse partijen zijn inmiddels bezig met initiatieven op dit gebied. Onder andere de BBC, The New York Times, Google, Facebook en Microsoft hebben hun krachten gebundeld in het Trusted News Initiative, waarmee ze berichten afkomstig van door hen als betrouwbaar aangemerkte informatiebronnen willen waarmerken.[184]

---

[182] EUR-lex (2000). *Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce')* https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32000L0031
[183] Jacobs, B. (2019). *Teken tegen nepnieuws.* iBestuur https://ibestuur.nl/weblog/teken-tegen-nepnieuws
[184] Cooper, D. (2020). *News outlets will digitally watermark content to limit misinformation.* Engadget www.engadget.com/bbc-fake-news-111000601.html

De overheid zou standaard gebruik kunnen maken van waarmerken bij eigen berichtgeving. Daarnaast zou zij van publieke nieuwsbronnen die (mede) gefinancierd worden door de overheid, kunnen verlangen dat zij hun berichtgeving standaard waarmerken.[185] In Nederland is de Nederlandse Publieke Omroep (NPO) bezig met initiatieven op dit gebied.[186]

Ook bestaan er methoden om de verspreiding van gemanipuleerd beeldmateriaal tegen te gaan. Zo zouden platformbedrijven die filters aanbieden voor beeldmanipulatie, standaard waarmerken kunnen achterlaten in beelden die met die filters bewerkt zijn. Als anderen vervolgens die beelden verspreiden en doen voorkomen alsof het authentiek materiaal betreft, kan dat door de ontvanger worden nagegaan.

Daarnaast bestaat er *controlled capture*-software waarmee het tijdstip waarop beelden zijn gemaakt en de locatie waar dat gebeurde, worden vastgelegd, waardoor gegevens achteraf niet aangepast kunnen worden.[187] Zo maakt Truepic het mogelijk om bij het fotograferen met een smartphone het tijdstip, de locatie en de smartphone-identiteit te registreren.[188]

8.2 Maatregelen tegen beïnvloeding met micro-targeting

De casestudie over *psychographing* laat zien dat met behulp van micro-targeting diverse doelgroepen in de samenleving op voor anderen niet zichtbare wijze kunnen worden bereikt met verschillende – en mogelijk tegenstrijdige – maatschappelijke en politieke boodschappen. De politieke stemming in het land en de politieke meningsvorming van burgers kunnen daarmee worden beïnvloed.

8.2.1 Breder insteken dan politieke advertenties

De huidige beleidsdiscussie in Nederland over micro-targeting richt zich vooral op het gebruik ervan in politieke advertentiecampagnes. Zo pleit de Staatscommissie

[185] Van Boheemen, P., Munnichs, G., Kool, L., Diercks, G., Hamer, J., & Vos. A. (2020). *Cyberweerbaar met nieuwe technologie.* Rathenau Instituut. www.rathenau.nl/nl/digitale-samenleving/cyberweerbaar-met-nieuwe-technologie
[186] Takken, W. (2020). *Martijn van Dam: 'de publieke omroep moet ook online verbindend zijn'*. NRC. www.nrc.nl/nieuws/2020/06/08/npo-moet-ook-online-verbindend-zijn-a4002042
[187] Duursma, J. (2019). *Deepfake Technologie – The Infocalypse* www.jarnoduursma.nl/wp-content/uploads/2019/09/Jarno-Duursma-_-Deepfake-Technologie-The-Infocalypse.pdf
[188] Truepic (z.d.). *Photo and video verification you can trust* https://truepic.com/

parlementair stelstel (commissie-Remkes) voor een wettelijke transparantieplicht, die openheid van politieke partijen over hun inzet van digitale instrumenten moet afdwingen. Volgens de commissie moeten burgers politieke advertenties als zodanig kunnen herkennen, en kunnen zien waarom juist zij een bepaalde boodschap te zien krijgen, en wie voor de advertentie betaalt.[189]

In lijn hiermee schrijft de minister van Binnenlandse Zaken en Koninkrijksrelaties in een brief aan de Tweede Kamer: 'Om oneigenlijke beïnvloeding van eerlijke en vrije verkiezingen te voorkomen moeten verkiezingscampagnes transparant verlopen. Daarom neem ik in de Wpp [Wet op de politieke partijen] regels op die de controleerbaarheid van deze campagnes moeten waarborgen en vergroten, misleiding voorkomen en duidelijkheid geven over wie een advertentie heeft betaald. Dit heeft tot doel om de democratie te beschermen tegen (…) risico's die digitale informatie- en communicatietechnologieën voor verkiezingen met zich mee kunnen brengen. Regulering hiervan heeft tot doel de campagnes voor kiezers inzichtelijk te maken'.[190]

Overigens merkt de commissie-Remkes op dat de manier waarop micro-targeting wordt toegepast in de VS, in Nederland niet mogelijk is omdat partijen geen toegang hebben tot registers van kiezers.[191] Onder de Algemene verordening gegevensbescherming (AVG) is het ook niet toegestaan om gegevens over politieke voorkeuren zonder toestemming van de betrokken personen te gebruiken voor micro-targeting.

Hoewel de voorgenomen maatregelen van de minister een belangrijke stap zijn om mogelijke misleiding van de kiezer met behulp van digitale advertentiecampagnes tegen te gaan, willen we hierbij enkele kanttekeningen plaatsen.

Zo schatten we in dat het verbod van de AVG op het gebruik van gegevens over politieke voorkeuren, te omzeilen valt. Zo zal een *Advanced Persistent Threat*-groepering (zie paragraaf 7.2) zich waarschijnlijk weinig aantrekken van de AVG als zij erop uit is om gegevens over politieke voorkeuren te gebruiken voor politieke beïnvloedingscampagnes. Als ze daartoe overgaat, kan deze groepering daarvoor uiteraard wel op grond van de AVG juridisch worden vervolgd.

De AVG zou ook kunnen worden omzeild door aan de hand van andersoortige gegevens – zoals postcodegebieden, automerken of krantenabonnementen – bij

[189] Staatscommissie parlementair stelsel (2018). *Lage drempels, hoge dijken: Democratie en rechtsstaat in balans.* www.staatscommissieparlementairstelsel.nl/documenten/rapporten/samenvattingen/12/13/eindrapport
[190] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2020). *Kamerbrief over voortgang voorbereiding Wet op de politieke partijen*. www.rijksoverheid.nl/documenten/kamerstukken/2020/06/11/kamerbrief-inzake-voortgang-voorbereiding-wet-op-de-politieke-partijen
[191] Staatscommissie parlementair stelsel (2018). *Lage drempels, hoge dijken: Democratie en rechtsstaat in balans.* www.staatscommissieparlementairstelsel.nl/documenten/rapporten/samenvattingen/12/13/eindrapport

benadering (*by proxy*) te achterhalen welke politieke voorkeuren mensen hebben. Om hen vervolgens op basis van deze ingeschatte voorkeuren van bepaalde, gekleurde informatie te voorzien. Het hoeft hierbij niet te gaan om gebruik van tot de persoon herleidbare gegevens over diens politieke voorkeur, die onder de AVG zouden vallen.

Bovendien achten we een insteek die zich vooral richt op politieke advertenties als te beperkt. De casestudie over *psychographing* laat zien dat desinformatiecampagnes die gebruikmaken van micro-targeting ook gericht kunnen zijn op het aanwakkeren van maatschappelijke tegenstellingen, en radicalisering en polarisering in de hand kunnen werken. Dat zou grote impact kunnen hebben op bijvoorbeeld de politieke stemming in het land.

Micro-targeting kan ook worden ingezet voor de verspreiding van complottheorieën, leidend tot aantasting van het vertrouwen van bevolkingsgroepen in de gevestigde media, de rechtsspraak en de politieke instituties. Dat kan sluipenderwijs de legitimiteit van de democratische rechtsorde ondermijnen.[192]

De minister van Binnenlandse Zaken en Koninkrijksrelaties heeft laten weten dat zij zich wat betreft de verspreiding van desinformatie met behulp van micro-targeting niet alleen wil richten op het gebruik ervan in politieke campagnes, waarvoor ze de Wet op de politieke partijen wil aanpassen. Ook is zij voorstander van ‘meer transparantie over de herkomst en de methoden van verspreiding van desinformatie op internetdiensten’. Ze overweegt of ‘wettelijke regels de transparantie kunnen afdwingen’. Het is nog niet duidelijk welke invulling daaraan wordt gegeven, zoals de minister ook zelf aangeeft. [193]

8.2.2 Regulering door platformbedrijven

Het bovenstaande laat zien dat voor het bestrijden van desinformatiecampagnes waarbij gebruik wordt gemaakt van micro-targeting, ook de rol van platformbedrijven in ogenschouw moet worden genomen. Zij maken het immers mogelijk dat andere, kwaadwillende partijen micro-targeting kunnen inzetten voor desinformatiedoeleinden. De huidige regulering van platformbedrijven schiet tekort om deze negatieve effecten tegen te gaan.

---

[192] Cohen, J.E. (2020 Forthcoming). Tailoring Election Regulation: The Platform Is the Frame. *4 Geo. Tech. L. Rev.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3573127

[193] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2020). Kamerbrief ontwikkelingen beleidsinzet bescherming democratie tegen desinformatie. www.rijksoverheid.nl/documenten/kamerstukken/2020/05/13/kamerbrief-ontwikkelingen-beleidsinzet-bescherming-democratie-tegen-desinformatie

EU-gedragscode volstaat niet

De belangrijkste bestaande regulering van platformbedrijven is de Europese *Code of Practice on Disinformation*. Deze gedragscode is door een aantal grote platformbedrijven ondertekend. Hiermee geven ze aan zich te willen inspannen om de verspreiding van desinformatie tegen te gaan. Maar de resultaten hiervan vallen tegen.

De Europese koepelorganisatie voor mediatoezicht ERGA constateert dat de gedragscode diverse tekortkomingen kent.[194] Ten eerste rapporteren platformbedrijven zelf over de implementatie van maatregelen, waardoor onafhankelijk controle onmogelijk is. Bovendien worden de rapportages te algemeen geformuleerd, op geaggregeerd EU-niveau, en bestaat er onduidelijkheid over de definities van kernbegrippen zoals 'politieke advertenties'. Verder hebben diverse populaire platforms de code niet ondertekend, zoals WhatsApp en Messenger.

ERGA beveelt dan ook aan om de code aan te scherpen. Zo zouden de inspanningen van de platformbedrijven op nationaal niveau moeten worden gemonitord door onafhankelijke toezichthouders. Deze moeten internationaal vergelijkbare rapportages uitbrengen, met gebruik van uniforme begrippen en indicatoren. Alle platformbedrijven – in ieder geval vanaf een bepaalde grootte – die binnen de EU actief zijn, moet deze vorm van *co-regulering* worden opgelegd. Dat houdt een vorm van zelfregulering van platforms in met toezichthouders die dwangmiddelen kunnen opleggen bij niet-naleving.

De in het kader van dit onderzoek georganiseerde expertmeeting leverde een vergelijkbaar beeld op. Tijdens deze bijeenkomst werd de mening breed gedeeld dat platformbedrijven te weinig doen tegen de verspreiding van desinformatie. Ook tegen verdergaande en strafbare uitingen als haatzaaien en oproepen tot geweld wordt volgens de bij de bijeenkomst aanwezige deskundigen te weinig door de platformbedrijven opgetreden. Zelfregulering blijkt volgens hen onvoldoende te werken. Er zou dan ook strenger moeten worden gereguleerd.

Beperken van mogelijkheden micro-targeting

Platformbedrijven zouden – net als andere ontwikkelaars van commerciële advertentiesystemen – in de eerste plaats monitoringsmogelijkheden kunnen inbouwen in door hen ontwikkelde diensten waarmee online advertentiecampagnes kunnen worden ontworpen, uitgevoerd en geanalyseerd. Zij zouden kunnen monitoren door wie en voor welke doeleinden de door hen geleverde diensten

[194] European Regulators Group for Audiovisual Media Services (2020). *ERGA Report on disinformation: Assessment of the implementation of the Code of Practice* https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf

worden gebruikt, hun afnemers vooraf kunnen screenen (*due dilligence*) en preventief te werk gaan wanneer ze vermoeden dat afnemers misbruik maken van hun diensten. Zo zouden ze het kunnen signaleren wanneer diensten worden ingezet voor polariserende of radicaliserende campagnes, of andere vormen van desinformatie. Dat vereist wel dat de platformbedrijven hierop monitoren. Vervolgens zouden ze kunnen optreden tegen afnemers die hun diensten gebruiken voor schadelijke toepassingen.

Platformbedrijven zouden vervolgens de mogelijkheden om misbruik te maken van advertentietechnologie ook technisch kunnen beperken. Ze zouden maatregelen kunnen nemen om beïnvloeding van geselecteerde doelgroepen met behulp van micro-targeting tegen te gaan. Zo zouden ze de mogelijkheden kunnen beperken om doelgroepen te selecteren aan de hand van persoonlijkheidskenmerken, bijvoorbeeld door de automatische extractie van data door middel van API's aan banden te leggen. Zo kan het technisch onmogelijk worden gemaakt om bij doelgroepenselectie gebruik te maken van politieke voorkeuren, of van data waarvan bekend is dat daarvan politieke voorkeuren kunnen worden afgeleid.

Volgens Crain en Nadler zouden de advertentiesystemen ook transparanter moeten worden voor internetgebruikers. Voor hen is het vaak niet duidelijk welk advertentieprofiel op hen van toepassing is, en om welke reden. Ook kunnen gebruikers vaak niet zien wie de afzender van een advertentie is. Verder zouden gebruikers meer controle en zeggenschap moeten krijgen over de data die door adverteerders worden gebruikt. Dat moet ook voorkomen dat data van gebruikers kunnen worden ingezet voor doeleinden die hun eigen belangen schaden. Crain en Nadler pleiten er ook voor dat de mogelijkheden voor micro-targeting worden ingeperkt door een ondergrens te stellen aan de omvang van een doelgroep.[195]

De hier genoemde maatregelen om de mogelijkheden van micro-targeting in te perken, kunnen ten koste gaan van het verdienmodel van de platformbedrijven. Zij kunnen hierdoor immers minder aantrekkelijk worden voor bepaalde groepen adverteerders. Het is dan ook geenszins vanzelfsprekend dat de platformbedrijven hier uit eigen beweging toe over zullen gaan. In dat geval zou de overheid – en bij voorkeur de EU – hierop kunnen aandringen, of nadere eisen kunnen stellen via regelgeving.

Sommige partijen bepleiten zelfs een verbod op gepersonaliseerde advertenties en micro-targeting. Onder andere de grote inbreuk op de privacy van internetgebruikers en de manipulatie van hun online gedrag op basis van de over

[195] Crain & Nadler (2019). Political Manipulation and Internet Advertising Infrastructure. *Journal of Information Policy*, *9*, 370. https://doi.org/10.5325/jinfopoli.9.2019.0370

hen verzamelde data, vormen daarvoor belangrijke overwegingen. Het voorgestelde verbod wordt als een effectievere manier gezien om met de diverse problematische kanten om te gaan, dan elk probleem apart te reguleren.[196]

Ook het Europees Parlement pleit in een motie voor een verbod op gepersonaliseerde advertenties. Dat verbod zou moeten worden opgenomen in de Digital Services Act. Indiener van het amendement Paul Tang stelt, dat de gepersonaliseerde advertenties zijn gebaseerd op een ongewenste inbreuk op de privacy, dat gepersonaliseerde advertenties storend zijn, en dat het voor veel gebruikers te veel is gevraagd om steeds af te moeten zien van tracking-cookies.[197]

8.3 Transparantie over aanbevelingsalgoritmes

Zoals in de quickscan al uiteen is gezet, zijn de aanbevelingsalgoritmes van platformbedrijven er veelal op gericht om de aandacht van de gebruiker zo lang mogelijk vast te houden, hetgeen het beste lukt met sensationele content die aansluit bij eerder gebleken gebruikersvoorkeuren. Deze algoritmes versterken daarmee sociale en politieke voorkeuren en maatschappelijke tegenstellingen.

De aanbevelingsalgoritmes werken op een vergelijkbare manier als de technieken voor micro-targeting. Met behulp van de algoritmes worden de vele duizenden gegevens die platformbedrijven verzamelen over hun gebruikers geanalyseerd, en wordt hen content getoond die is afgestemd op de voorkeuren en persoonskenmerken die uit die analyse naar voren komen. Een belangrijk verschil is dat de aanbevelingsalgoritmes niet zelf content produceren, maar werken met berichten van anderen. Een ander verschil is dat het merendeel van de platformbedrijven niet doelbewust met de door hen gehanteerde algoritmes uit is op schade aan het publieke debat of het democratisch proces. Maar al met al kan de radicaliserende en polariserende uitwerking van de aanbevelingsalgoritmes wel degelijk zulke schade teweegbrengen – en lijken de platformbedrijven vooralsnog weinig genegen om die schadelijke effecten tegen te gaan. In plaats daarvan beroepen ze zich op een neutrale rol als intermediair tussen afzenders en ontvangers van berichten.

[196] Edelman, G. (2020). *Why Don't We Just Ban Targeted Advertising?* www.wired.com/story/why-dont-we-just-ban-targeted-advertising/
[197] Kist, R. (2020). *Europarlementariër Paul Tang: 'Persoonlijke advertenties zijn een smet op het internet'.* NRC. www.nrc.nl/nieuws/2020/06/19/overwinning-voor-paul-tang-in-strijd-tegen-gepersonaliseerde-advertenties-techreuzen-a4003409 en https://paultang.nl/en/forbid-personalised-ads/

Volgens de Amerikaanse hoogleraar *Law and Technology* Julie Cohen volgt uit de werking van de aanbevelingsalgoritmes echter dat platformbedrijven allesbehalve een neutraal doorgeefluik zijn van door anderen geplaatste content. De algoritmes tonen content die aansluit bij de voorkeuren, wensen en zorgen die uit de gebruikersprofielen volgen, versterken die voorkeuren en houden door sensationele berichtgeving de aandacht van gebruikers vast. Aldus vindt volgens Cohen een 'endemische' manipulatie van gebruikers plaats.[198]

Inbouwen reflectiemoment in platformdiensten

Het bovenstaande roept de vraag op welke mogelijkheden platformbedrijven hebben om de schadelijke effecten van hun aanbevelingsalgoritmes tegen te gaan. Een mogelijkheid om de verspreiding van sensationele of meer extreme berichten tegen te gaan, is door een reflectiemoment in het gebruik van platformdiensten in te bouwen.

Doordat de verspreiding van deze berichten deels plaatsvindt doordat gebruikers deze informatie liken of met anderen delen, biedt hun online gedrag daarvoor aanknopingspunten. Gebruikers delen berichten namelijk vaak impulsief, zonder er echt over na te denken. Door hen bijvoorbeeld technisch te dwingen om enkele seconden te wachten voordat ze een bericht kunnen verspreiden, doen ze dat mogelijk minder impulsief.[199] Twitter experimenteert in dit verband met een waarschuwing als gebruikers een ongeopende link doorsturen die ze niet zelf op betrouwbaarheid hebben kunnen beoordelen.[200]

De overheid – bij voorkeur de EU – zou van platformbedrijven kunnen verlangen zo'n reflectiemoment in te bouwen. Dat zou overigens ook kunnen worden verlangd van beheerders van chatapps.

Transparantie over gebruik algoritmes

Een verdergaande maatregel om de polariserende en radicaliserende werking van aanbevelingsalgoritmes tegen te gaan zou zijn om het gebruik van die algoritmes kritisch tegen het licht te houden en ze zo nodig aan te passen. Maar platformbedrijven zijn niet transparant over de door hen gebruikte algoritmes. Ze worden beschouwd als concurrentiegevoelige bedrijfsgeheimen. De Nederlandse overheid onderschrijft dat. Zo stelt de minister van Binnenlandse Zaken en

---

[198] Cohen, J.E. (2020 Forthcoming). Tailoring Election Regulation: The Platform Is the Frame. *4 Geo. Tech. L. Rev.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3573127

[199] Fazio, L. (2020). Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School Misinformation Review*, *1*(2). https://doi.org/10.37016/mr-2020-009

[200] Twitter Support (2020). https://twitter.com/twittersupport/status/1270783537667551233

Koninkrijksrelaties dat de aanbevelingsalgoritmes intellectueel eigendom van de platformbedrijven zijn, en dat ze zelf bepalen hoe ze hun algoritmes vormgeven.[201]

Verschillende partijen pleiten echter voor meer transparantie van de platformbedrijven op dit gebied. Zo pleit de Europese *High Level Expert Group on Fake News and Online Disinformation* voor meer transparantie over het gebruik door platforms van aanbevelingsalgoritmes.[202] En de Britse parlementaire commissie *Digital, Culture, Media and Sport* (DCMS) pleit in een onderzoeksrapport over desinformatie voor een onafhankelijke toezichthouder op platformbedrijven, die toegang krijgt tot de door hen gebruikte algoritmes.[203] De maatschappelijke organisatie Electronic Frontier Foundation gaat nog een stap verder, en stelt dat internetgebruikers zelf de algoritmes moeten kunnen veranderen die bepalen welke content zij te zien krijgen, en moeten kunnen aangeven welke bronnen zij vertrouwen.[204]

Onderzoek naar werking algoritmes

Een eerste stap op weg naar meer transparantie zou zijn dat platformbedrijven wetenschappelijke onderzoekers toegang geven tot hun aanbevelingsalgoritmes, zodat zij kunnen nagaan hoe de algoritmes werken en welke impact ze hebben op het online gedrag van internetgebruikers. Aldus zou meer zicht kunnen worden verkregen op de werking en de mogelijke negatieve effecten daarvan.[205] [206] [207] Zo nodig zouden overheden de toegang voor wetenschappelijke onderzoekers tot die informatie moeten afdwingen.

---

[201] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2019). *Antwoord op vragen van het lid Baudet over het rapport 'Politiek en Sociale Media Manipulatie'*. www.tweedekamer.nl/kamerstukken/kamervragen/detail?id=2019Z20342&did=2019D46745

[202] European Commission (2018). *Final report of the High Level Expert Group on Fake News and Online Disinformation* https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation

[203] House of Commons (2019). *Disinformation and 'fake news': Final Report* https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf

[204] York, J., Greene, D. & Gebhart, G. (2019). *Censorship Can't Be The Only Answer to Disinformation Online*. Electronic Frontier Foundation www.eff.org/nl/deeplinks/2019/05/censorship-cant-be-only-answer-disinformation-online

[205] Ausloos, J. (2020). *Technologiereuzen moeten zeggen hoe ze ons gedrag bepalen en zo dwingen we dat af*. VRT www.vrt.be/vrtnws/nl/2020/06/25/de-macht-van-technologiereuzen-en-hoe-ze-aan-banden-te-leggen/

[206] European Commission (2018). *Final report of the High Level Expert Group on Fake News and Online Disinformation* https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation

[207] Bruns, A. (2019). After the 'APIcalypse': social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, *22*(11), 1544–1566. https://doi.org/10.1080/1369118X.2019.1637447

8.4 Maatregelen gericht op besloten en versleutelde kanalen

Zoals de casestudies in deel II laten zien, kunnen producenten en verspreiders van *deepfakes* en groeperingen die met behulp van micro-targeting desinformatiecampagnes opzetten, gebruikmaken van besloten of versleutelde groepen en kanalen op socialemediaplatforms en chatapps, waar geen moderatie plaatsvindt of mogelijk is. Verspreiders van *deepfakes* kunnen in reactie op de toenemende mogelijkheden om *deepfakes* te detecteren, besluiten uit te wijken naar die groepen en kanalen; afzenders van (heimelijke) micro-targeting-campagnes zullen daar doorgaans gebruik van maken.

Deze besloten of versleutelde groepen en kanalen hebben weliswaar een minder groot bereik dan de publieke kanalen van socialemediaplatforms, maar kunnen nog steeds miljoenen gebruikers hebben. Ook kunnen berichten tegelijkertijd op meerdere kanalen worden verspreid.

Gebruikerslimieten
Producenten en verspreiders van desinformatie dreigen daarmee vrij spel te krijgen. Een eerste mogelijkheid om de hiermee samenhangende dreigingen voor het publieke debat en het democratisch proces tegen te gaan, is het stellen van een maximaal aantal gebruikers per groep of kanaal. Sommige platformbedrijven hanteren zulke limieten. Een belangrijk nadeel is dat daarmee het gebruik van zulke kanalen, dat vanuit democratisch oogpunt wordt toegejuicht, ook wordt ingeperkt. Zie bijvoorbeeld de belangrijke publieke functie van besloten en versleutelde kanalen in niet-democratische landen. Zo worden Telegramkanalen, die geen gebruikerslimieten kennen, door de Iraanse bevolking volop gebruikt voor het uitwisselen van actualiteiten. En in een land als Brazilië vormen chatapps de belangrijkste nieuwsbron van veel burgers.

Waarschuwingssysteem
Een tweede mogelijkheid om de impact van desinformatie op besloten en versleutelde kanalen tegen te gaan, is het instellen van een onafhankelijk nationaal waarschuwingssysteem dat gericht is op het signaleren van desinformatiecampagnes rondom gevoelige maatschappelijke kwesties. Tijdens de expertmeeting werd deze mogelijkheid door diverse deelnemers geopperd.

De kans is groot dat berichten in besloten en versleutelde kanalen vroeg of laat in de openbaarheid komen. Zodra dat gebeurt, kunnen de berichten worden gesignaleerd en kan erop worden gereageerd. De berichten zouden bijvoorbeeld publiekelijk kunnen worden weersproken en internetgebruikers zouden kunnen worden verwezen naar betrouwbare informatiesites en factcheckers. Een voorbeeld

van dit laatste is de liveblog die de publieke nieuwszender NOS tijdens de coronacrisis heeft gelanceerd, met onder andere informatievideo's over actuele ontwikkelingen rond het virus en de aanpak ervan.

Het waarschuwingssysteem kan er in ieder geval voor zorgen dat internetgebruikers weten wat er speelt, en op de hoogte worden gebracht van actuele desinformatiecampagnes over gevoelige maatschappelijke kwesties. Dat is bijvoorbeeld te vergelijken met de waarschuwingen die werkgevers en banken doen uitgaan naar hun werknemers en klanten om hen te wijzen op veel voorkomende ransomware en phishingmails.

Een dergelijk waarschuwingssysteem vraagt om een meldplek waar internetgebruikers (mogelijke) desinformatie kunnen melden, en een website met links naar betrouwbaar bevonden informatiesites. Vanwege het publieke belang hiervan, zou de overheid dit kunnen faciliteren – uiteraard zonder de onafhankelijkheid van het waarschuwingssysteem aan te tasten.

Platformbedrijven zouden ook een rol kunnen spelen in het verwijzen naar als betrouwbaar aangemerkte informatiesites. Zo kunnen ze de zichtbaarheid van informatie uit geverifieerde bronnen vergroten. Zo plaatsten de Nederlandstalige websites van onder andere Facebook, Twitter, YouTube en Google tijdens de coronacrisis informatie afkomstig van het RIVM hoog in hun berichtgeving, zodat deze maximaal onder de aandacht van internetgebruikers werd gebracht. Hoewel dit vooralsnog een tijdelijke maatregel lijkt, zou het verwijzen door platformbedrijven naar als betrouwbare aangemerkte informatiesites ook op een meer permanente basis kunnen gebeuren.

De Europese commissie heeft onlangs een oproep doen uitgaan voor de vorming van (inter)nationale samenwerkingsverbanden van wetenschappers, factcheckers, journalisten en andere relevante stakeholders, die hiermee in lijn lijkt te liggen. Deze nationale *hubs* moeten onder andere een snelle signalering van desinformatiecampagnes mogelijk maken.[208]

Detectie kwaadwillende chatbots en monitoring door aanbieders
Zoals eerder is aangegeven, kunnen producenten en verspreiders van desinformatie ook gebruikmaken van (semi)geautomatiseerde chatbots voor het verspreiden van desinformatie.

[208] European Commission (2020). *2020 CEF Telecom Call - European Digital Media Observatory (CEF-TC-2020-2)* https://ec.europa.eu/inea/en/connecting-europe-facility/cef-telecom/apply-funding/2020-edmo

Aanbieders van chatapps zouden met behulp van detectietechnieken, ook zonder de versleuteling van de inhoud van de chatberichten te breken, kunnen bepalen of een account meer chatberichten binnen een bepaald tijdsbestek verstuurt dan menselijkerwijs mogelijk is – om het account vervolgens te verwijderen.

8.5 Factchecken blijft van groot belang

Om de impact van desinformatie op het publieke debat en het democratisch proces tegen te gaan, blijft het van groot belang dat misleidende informatie wordt gefactcheckt en tegengesproken. Niet voor niets is het aantrekkelijk voor producenten en verspreiders van desinformatie om heimelijk, op besloten en versleutelde kanalen, te werk te gaan. Dat maakt het namelijk voor anderen lastiger om bijvoorbeeld *deepfakes* als zodanig te ontmaskeren, of onjuiste informatie te weerspreken.

Zoals reeds eerder vermeld, ziet de overheid voor zichzelf geen primaire rol weggelegd bij het actief weerspreken van desinformatie. Ze acht het in de eerste plaats een verantwoordelijkheid voor journalisten en wetenschappers om berichten op hun feitelijke waarheid te checken en tegen te spreken.

Maar gezien het publieke belang van factchecken zou de overheid hierin wel een faciliterende rol kunnen spelen, bijvoorbeeld door het Stimuleringsfonds voor de Journalistiek en het Fonds Bijzondere Journalistieke Projecten van (extra) financiële middelen te voorzien. Beide fondsen zouden vervolgens factcheckers kunnen financieren. Daarbij moet wel worden gewaakt voor de onafhankelijkheid van de factcheckers.

Zo'n indirect faciliterende rol van de overheid bij het factchecken van desinformatie valt te vergelijken met de manier waarop de Europese Commissie de bestrijding van desinformatie ondersteunt: door een platform in het leven te roepen dat gericht is op het faciliteren van onafhankelijke factcheckers.[209]

Ook bij factchecking is een rol voor platformbedrijven weggelegd. Het is namelijk van belang dat zij berichten die (mogelijk) desinformatie bevatten als zodanig labelen, en gebruikers doorverwijzen naar factchecksites waarop wordt aangegeven waarom bepaalde berichten (mogelijk) niet kloppen, waardoor deze tegeninformatie een wijder bereik krijgt. Ook kunnen platformbedrijven financieel bijdragen aan factchecksites.

[209] European Commission (2018). *Action Plan on Disinformation.* https://ec.europa.eu/commission/publications/action-plan-disinformation-commission-contribution-european-council-13-14-december-2018_en

Er worden trouwens ook technische middelen ontwikkeld waarmee het factchecken (deels) kan worden geautomatiseerd. Zo hebben Nederlandse en Vlaamse onderzoekers FactRank ontwikkeld, waarmee feitelijke, 'checkbare' uitspraken in Kamerdebatten of tweets van politici automatisch kunnen worden gedetecteerd. Met behulp van dergelijke hulpmiddelen kunnen factcheckers sneller werken. [210]

8.6 Investeren in mediawijsheid blijft van groot belang

Ten slotte kwam zowel uit de interviews als de expertmeeting naar voren dat een strengere regulering van platformbedrijven en technologische maatregelen ter bestrijding van desinformatie maar beperkt zin hebben als niet tegelijkertijd geïnvesteerd blijft worden in mediawijsheid. Dit ligt in lijn met het regeringsbeleid, maar het belang ervan kan niet genoeg worden onderstreept.

Investeer in mediawijsheid
Het behoeft geen betoog dat hoe beter mensen in staat zijn (des)informatie op waarde te schatten, hoe minder maatschappelijke en politieke impact misleidende berichtgeving zal hebben.

Hoewel Nederland in vergelijking met veel andere landen goed scoort op het gebied van mediawijsheid, zijn de onderzoeksresultaten toch niet echt geruststellend. Een aanzienlijk deel van de Nederlandse bevolking heeft namelijk moeite om de betrouwbaarheid van informatie goed in te schatten.[211]

Uit onderzoek van Kantar wordt duidelijk dat een aantal bevolkingsgroepen op dit gebied extra ondersteuning nodig heeft. Vooral ouderen en mensen met een lagere opleiding blijken kwetsbaar. Dat betreft vooral hun begrip van hoe media de werkelijkheid weergeven – en daarbij vaak ook inkleuren –, het vinden en verwerken van informatie en het reflecteren op het eigen mediagebruik.[212]
Dit betekent dat niet alleen binnen het onderwijs – dat zich vooral richt op jongeren – maar ook daarbuiten meer aandacht voor mediawijsheid nodig is. Het Netwerk Mediawijsheid zet zich daar overigens al voor in.

Ook moet worden bedacht dat het met de ontwikkeling van meer geavanceerde vormen van desinformatie, (zoals geraffineerdere toepassingen van *deepfake*-

[210] Universiteit Leiden (2020). www.universiteitleiden.nl/nieuws/2020/05/lancering-factrank
[211] Kranenburg, A. (2017). *Nederlanders bezorgd over 'nepnieuws' - een op drie weet vaak niet meer wat waar is en wat onwaar.* Volkskrant www.volkskrant.nl/nieuws-achtergrond/nederlanders-bezorgd-over-nepnieuws-een-op-drie-weet-vaak-niet-meer-wat-waar-is-en-wat-onwaar~b6914596/
[212] Plantinga, S. & Kaal, M. (2018). *Hoe mediawijs is Nederland?* www.mediawijzer.net/wp-content/uploads/sites/6/2018/09/Rapport-Mediawijsheid-volwassenen-2018.pdf

technologie of meer verfijnde vormen van micro-targeting), voor veel mensen alleen maar moeilijker zal worden om desinformatie te herkennen. De technologische ontwikkelingen gaan voor een deel van de bevolking zo snel dat ze deze niet meer kunnen bijbenen.

Niet bij ratio alleen

Tevens moet worden bedacht dat een al te rationele benadering van de omgang met desinformatie haar doel voorbij kan schieten.

Een belangrijke bevinding van de expertmeeting is dat een te eenzijdige nadruk op het waarheidsgehalte van informatie – bijvoorbeeld bij factchecken – niet voor iedereen zal werken. In veel situaties blijft het voor mensen lastig om de betrouwbaarheid van informatiebronnen te verifiëren. Hierbij speelt mee dat wetenschappelijke inzichten na verloop van tijd kunnen veranderen, en dat voor veel maatschappelijke kwesties geen eenduidig wetenschappelijk antwoord te geven is.

Bovendien doet het 'onware' karakter van desinformatie er niet altijd toe. Door desinformatie te verspreiden, kunnen internetgebruikers bijvoorbeeld ook uiting geven aan ontevredenheid of boosheid over maatschappelijke of politieke kwesties. Ook bijvoorbeeld afnemend vertrouwen in de overheid kan hierin een rol spelen. Bij de bestrijding van desinformatie is het dan ook van belang te beseffen dat berichten op meer aspecten kunnen worden beoordeeld dan alleen op hun betrouwbaarheid of waarheidsgehalte.[213] [214] In het streven naar mediawijsheid zou dan ook meer oog moeten zijn voor de bredere context waarin (des)informatie een rol speelt.[215]

[213] Marwick, A.E. (2018). Why Do People Share Fake News? A Sociotechnical Model of Media Effects. 2 *GEO. L. TECH. REV.* 474

[214] Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond Misinformation: Understanding and Coping with the 'Post-Truth' Era. *Journal of Applied Research in Memory and Cognition, 6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

[215] Bessems, K. (2020). *Socioloog Harambam: 'We zetten complotdenkers te gauw weg als gekkies'.* Volkskrant. www.volkskrant.nl/nieuws-achtergrond/socioloog-harambam-we-zetten-complotdenkers-te-gauw-weg-als-gekkies~b1942c88/

9 Conclusies

In dit afsluitende hoofdstuk sommen we de belangrijkste bevindingen op van dit onderzoek naar de mogelijke impact van technologische ontwikkelingen op het gebied van desinformatie, en naar mogelijke nieuwe maatregelen die de bedreigingen van desinformatie voor het publieke debat en het democratisch proces zouden kunnen tegengaan.

9.1 Een verontrustend beeld

We zijn dit onderzoek begonnen met de constatering dat desinformatie in Nederland de afgelopen jaren geen grote maatschappelijke of politieke impact heeft gehad. De hausse aan misleidende berichten die verspreid zijn rond de uitbraak van het coronavirus is mogelijk een uitzondering hierop. Maar het is nog te vroeg om een oordeel te kunnen vellen over de betekenis daarvan voor de weerbaarheid van de Nederlandse samenleving tegen desinformatie.

9.1.1 Veelvormige technologische mogelijkheden voor productie en verspreiding van desinformatie

De snelle technologische ontwikkelingen op het gebied van IT zouden hierin echter op afzienbare termijn verandering kunnen brengen. Dit onderzoek geeft een breed overzicht van de technologische ontwikkelingen die de komende jaren een rol kunnen gaan spelen bij de productie en verspreiding van desinformatie. Dat overzicht stelt bepaald niet gerust. De mogelijkheden die technologieën als tekstsynthese, *voice cloning*, *deepfakes*, micro-targeting en chatbots producenten en verspreiders van desinformatie bieden om internetgebruikers te misleiden, zijn groot en veelvormig. Die mogelijkheden reiken van niet of nauwelijks van authentiek materiaal te onderscheiden, gemanipuleerde videobeelden en heimelijke beïnvloeding van het stemgedrag van kiezers met behulp van geavanceerde vormen van micro-targeting, tot verspreiding van misleidende informatie met behulp van (semi-)geautomatiseerde een-op-eengesprekken van chatbots.

Veel van de beschreven technologieën zijn ook nog volop in ontwikkeling. Dat leidt er bijvoorbeeld toe dat *deepfakes* steeds geavanceerder worden – en dat betekent: nog lastiger te onderscheiden van authentiek beeldmateriaal. Dat ondermijnt ook het vermogen van burgers om informatie en desinformatie op waarde te schatten.

Daar komt bij dat de groeperingen die betrokken zijn bij de productie en verspreiding van desinformatie vaak opportunistisch te werk gaan. Ze buiten kwetsbaarheden in de samenleving uit, haken aan bij oplaaiende discussies in de media over gevoelige maatschappelijke kwesties, en maken gebruik van die technische middelen die het meeste effect sorteren. De betrokken groeperingen zijn daarnaast lastig aan te pakken. Vaak is moeilijk te achterhalen wie achter een desinformatiecampagne zit; en als daar al vermoedens over bestaan, is bewijsvoering vaak lastig rond te krijgen.

9.1.2 Technologische bestrijding desinformatie is nodig, maar biedt te weinig soelaas

Daar staat tegenover dat de technologische ontwikkelingen ook kunnen leiden tot nieuwe of verbeterde mogelijkheden om desinformatie te bestrijden. Bijvoorbeeld door *deepfakes* te detecteren met behulp van kunstmatige intelligentie, het vroegtijdig monitoren van een kwaadaardige inzet van micro-targeting of een automatische opsporing van een dito inzet van chatbots.

Maar tegenmaatregelen kunnen op hun beurt leiden tot nog geavanceerde *deepfakes*, die moeilijker zijn te detecteren, of verfijndere vormen van micro-targeting. Ook kunnen tegenmaatregelen ertoe leiden dat de activiteiten van producenten en verspreiders van desinformatie zich verplaatsen naar besloten groepen en kanalen, en zich daarmee onttrekken aan het oog van moderatoren.

Ook om een andere reden kan de bestrijding van desinformatie op achterstand raken. De mogelijkheden die nieuwe technologieën bieden om desinformatie te bestrijden, lijken beperkter dan de technologische mogelijkheden om desinformatie te produceren en verspreiden. Zo kunnen productie- en verspreidingstechnologieën sterk profiteren van automatisering, terwijl voor de bestrijding ervan vaak mensen nodig blijven om te beoordelen of er werkelijk sprake is van misleidende informatie.

9.2 Mogelijke nieuwe maatregelen

In het vorige hoofdstuk zijn een aantal maatregelen besproken die kunnen worden genomen om de dreigingen voor het publieke debat en het democratisch proces, die uitgaan van nieuwe technologisch ontwikkelingen op het gebied van desinformatie, het hoofd te bieden. Daarvoor lijkt meer nodig dan de huidige maatregelen die worden getroffen tegen de productie en verspreiding van desinformatie. Zo zouden de nieuwe kansen kunnen worden benut die

technologische ontwikkelingen bieden om desinformatie te bestrijden. Tegelijkertijd liggen veel van de door ons geopperde maatregelen in het verlengde van de bestaande maatregelen of daaraan ten grondslag liggende uitgangspunten.

We volstaan hier met een korte weergave van eerder naar voren gebrachte nieuwe maatregelen.

9.2.1 Maatregelen tegen deepfakes

Investeringen in detectie van deepfakes
Platformbedrijven kunnen investeren in een actief detectiebeleid gericht op de bestrijding van *deepfakes*, om mee te kunnen komen in de wedloop die mogelijk ontstaat tussen producenten en verspreiders van steeds geavanceerde *deepfakes* aan de ene kant, en platformbedrijven aan de andere kant.

Instellen van een meldpunt tegen kwaadaardige beeldmanipulatie
Platformbedrijven als YouTube, SnapChat, Instagram en TikTok, waar *deepfakes* alomtegenwoordig zijn, kunnen een meldpunt instellen waar gebruikers (vermoedelijk) kwaadaardige beeldmanipulatie kunnen rapporteren.

Waarmerken van beeldmateriaal en overige berichten
Het digitaal waarmerken van beeldmateriaal en overige berichten stelt internetgebruikers in staat na te gaan of materiaal afkomstig is van een in hun ogen betrouwbare informatiebron. Dat vereist een betrouwbaar systeem om digitale waarmerken te registreren. De overheid en de grote technologiebedrijven kunnen hierin vooropgaan.

9.2.2 Inperken mogelijkheden voor micro-targeting

Monitoren van gebruik advertentietechnologie
Platformbedrijven kunnen monitoringsmogelijkheden inbouwen in hun diensten om misbruik van door hen geleverde advertentietechnologie tegen te gaan.

Technische mogelijkheden advertentietechnologie inperken
Platformbedrijven kunnen adverteerders restricties opleggen bij hun doelgroepselectie, en monitoren op een verantwoord gebruik van door hen aangeboden advertentietechnologie.

Internetgebruikers transparantie bieden
Platformbedrijven kunnen hun gebruikers meer inzicht geven in welk advertentieprofiel op hen van toepassing is en om welke redenen, en in het gebruik daarvan door adverteerders.

9.2.3 Maatregelen tegen schadelijke effecten aanbevelingsalgoritmes

Inbouwen reflectiemoment in platformdiensten
Om de schadelijke effecten van de verspreiding van sensationele berichten tegen te gaan, kunnen platformbedrijven een reflectiemoment inbouwen in het gebruik van hun diensten. Hierdoor worden gebruikers gestimuleerd om (des)informatie minder impulsief te delen.

Transparantie bieden over aanbevelingsalgoritmes
Om de schadelijke effecten van aanbevelingsalgoritmes tegen te gaan, kunnen platformbedrijven transparantie bieden over de werking van de algoritmes. Om te beginnen door wetenschappelijke onderzoekers daar toegang tot te geven.

9.2.4 Waarschuwingssysteem voor besloten en versleutelde kanalen

Om verspreiding van desinformatie op besloten en versleutelde kanalen tegen te gaan, kan een onafhankelijk nationaal waarschuwingssysteem worden ingesteld dat desinformatiecampagnes rondom gevoelige maatschappelijke kwesties signaleert, en daarvoor waarschuwt. De overheid en de platformbedrijven kunnen dit waarschuwingssysteem faciliteren.

9.2.5 Verdienmodel platformbedrijven kritisch tegen het licht houden

Maatregelen als de detectie van *deepfakes*, het inperken van de technische mogelijkheden van advertentietechnologie of het bieden van transparantie over de werking van aanbevelingsalgoritmes, kunnen op gespannen voet staan met het verdienmodel van platformbedrijven. Die kunnen dan ook weinig genegen zijn om deze maatregelen te treffen. In dat geval kan de overheid overgaan tot verdergaande maatregelen, zoals het afdwingen van meer transparantie over het

gebruik van aanbevelingsalgoritmes, of het kritisch tegen het licht houden van het verdienmodel van de platformbedrijven.

9.2.6 Investeren in factchecken blijft van belang

Omdat factchecken van belang is om internetgebruikers die op zoek zijn naar betrouwbare informatie houvast te bieden, kunnen de overheid en platformbedrijven (blijven) investeren in faciliteiten voor factcheckers.

9.2.7 Investeren in mediawijsheid blijft van belang

Technologische maatregelen en strengere regulering van platformbedrijven kunnen de productie en verspreiding van desinformatie terugdringen. Maar deze maatregelen zullen er niet toe leiden dat desinformatie wordt uitgebannen. Er bestaat weinig zicht op wat er gebeurt op door platformbedrijven beheerde, besloten en versleutelde groepen en kanalen. En ook niet alle platformbedrijven zullen de bereidheid hebben om op te treden tegen desinformatie. Er zullen vrijplaatsen blijven bestaan op het internet – en dat betekent ook dat er ruimte voor desinformatie zal blijven bestaan.

De overheid moet dan ook blijven investeren in mediawijsheid. Internetgebruikers zullen geconfronteerd blijven worden met desinformatie, en dan helpt het als ze beter in staat zijn om daarmee om te gaan.

Tegelijkertijd moet duidelijk zijn dat van mediawijsheid niet te veel kan worden verwacht. Bijvoorbeeld omdat desinformatiecampagnes door nieuwe technologische toepassingen steeds geraffineerder worden, en daarmee voor internetgebruikers moeilijker te doorzien.

9.3 Slotsom: platformbedrijven primair verantwoordelijk, maar overheid kan ingrijpen

Voor veel van de hierboven genoemde maatregelen, ligt de verantwoordelijkheid voor de bestrijding van desinformatie primair bij de platformbedrijven. Voor een aantal van de maatregelen geldt dat ook voor de beheerders van chatkanalen. Maar gezien het publieke belang dat gemoeid is bij het tegengaan van de schadelijke effecten die desinformatie kan hebben op het publieke debat en het democratisch proces, kan de overheid ertoe besluiten op te treden als

platformbedrijven hun verantwoordelijkheid onvoldoende nemen. De overheid kan bijvoorbeeld aandringen op een actief detectiebeleid gericht op het tegengaan van *deepfakes*, of op monitoring van een onverantwoord gebruik van adverteerders van de mogelijkheden van micro-targeting.

En als aandringen niet helpt, zouden maatregelen kunnen worden afgedwongen. Die maatregelen kunnen ook ten koste gaan van het verdienmodel van platformbedrijven. Of de overheid daartoe over moet gaan, zal mede afhangen van de ernst van de dreigingen voor het publieke debat en het democratisch proces, die uitgaan van bijvoorbeeld de polariserende werking van aanbevelingsalgoritmes of door platformbedrijven gefaciliteerde desinformatiecampagnes van adverteerders. Om voldoende gewicht in de schaal te leggen, ligt het voor de hand dat als wordt overgegaan tot het nemen van dwingende maatregelen, dit binnen EU-verband gebeurt.

Bijlage 1: Interviewvragen

Algemene vragen

1. Wat is uw rol/functie?
2. Wat verstaat u onder desinformatie?
3. Wat vindt u van de huidige situatie in Nederland ten aanzien van desinformatie?
    i. Welke dreigingen ziet u?
    ii. Hoe staat het met de weerbaarheid van de Nederlandse samenleving tegen desinformatie?
    iii. Wat vindt u van de huidige situatie internationaal gezien?
4. Hoe verwacht u dat de productie en verspreiding van desinformatie zich de komende vijf jaar zullen ontwikkelen? Wat zijn volgens u de belangrijkste dreigingen? Waarover maakt u zich de meeste zorgen?
5. Wat zijn in uw ogen de belangrijkste maatregelen (zowel technologisch als niet-technologisch) die de komende jaren kunnen worden genomen om de dreigingen het hoofd te bieden en de weerbaarheid van de Nederlandse samenleving te versterken? Wie is daarbij waarvoor verantwoordelijk?
6. Hoe kan in uw ogen worden voorkomen dat mogelijke maatregelen ten koste gaan van belangrijke maatschappelijke waarden als de vrijheid van meningsuiting of de persvrijheid?

Relevante technologieën

7. Tot nog toe zijn we in literatuuronderzoek de volgende technologieën tegengekomen die mogelijk relevant zijn voor de productie en verspreiding van desinformatie:
    - Tekstsynthese
    - *Voice cloning*
    - Beeldsynthese, waaronder *deepfakes*
    - *Augmented* en virtual reality en avatars
    - Memes
    - Databasetechnologie / big data / open data
    - Socialemediaplatforms, inclusief aanbevelingsalgoritmes en super apps
    - Chatapps, inclusief encryptietechnologie
    - Bots
    - Micro-targeting, inclusief *programmatic advertising*, *dynamic prospecting*, campagnesoftware, *natural language processing*, *sentiment monitoring* en *influencer marketing*
    - Zoekmachines

    - Spraakassistenten
    - *Distributed autonomous organisations* / blockchain
    - Games
    - Interactieve TV en live streaming
8. Ontbreken er in uw ogen relevantie technologieën?
9. Vindt u bepaalde technologieën minder relevant?
10. In aanvulling op voorafgaande 'Algemene vragen': Welke technologieën hebben volgens u de grootste impact?
11. Kunt u dit toelichten?

Maatregelen tegen desinformatie (in aanvulling op voorafgaande 'Algemene vragen')

12. Welke maatregelen moeten in uw ogen worden getroffen om de dreigingen die uitgaan van desinformatie het hoofd te bieden? Op (de ontwikkeling van) welke technologische tegenmaatregelen zou moeten worden ingezet?
13. Welke maatregelen acht u het meest effectief?
14. Wie is hierbij waarvoor verantwoordelijk? Welke partijen zijn hiertoe het beste uitgerust?
15. Wat is volgens u een voorbeeld van een good practice?

Afronding

16. Heeft u, naast alles wat we hebben besproken, nog verdere opmerkingen?

Bijlage 2: Deelnemers interviews

Naam	Organisatie
Noëlle Aarts	Radboud Universiteit
Jarno Duursma	Trendwatcher
Joris van Hoboken	Universiteit van Amsterdam
Linus Neumann	Netzpolitik Podcaster
Cees van Riel	Rotterdam School of Management
Adam Segal	Council on Foreign Relations
Anoniem	AIVD
Anoniem	Marketingexpert

Bijlage 3: Deelnemers expertmeeting

Naam	Organisatie
Peter Burger	Universiteit Leiden
Ufuk Esmer	BKB
Henriette Kieviet	Netwerk Mediawijsheid
Peter Olsthoorn	Zelfstandig journalist
Claes de Vreese	Universiteit van Amsterdam
Anoniem	Autoriteit Consument en Markt

Bijlage 4: Overzicht technologieën

Algemene technologieën

Technologie	Omschrijving
Databasetechnologie	Op grote schaal verzamelen en analyseren van (persoons)gegevens
Kunstmatige intelligentie	Zelflerende algoritmes en systemen

Productietechnologieën

Technologie	Omschrijving
Tekstsynthese	Algoritmes die leesbare en logische (nieuws)berichten genereren
Voice cloning	Manipulatie van spraakberichten met behulp van kunstmatige intelligentie
Beeldsynthese en deepfakes	Genereren en aanpassen van video's met behulp van kunstmatige intelligentie
AR, VR en avatars	Presenteren van informatie in een virtuele omgeving
Memes	Afbeeldingen ontworpen om op grote schaal te worden gedeeld op sociale media

Verspreidingstechnologieën

Technologie	Omschrijving
Socialemediaplatforms	Onlineplatforms zoals Facebook, Twitter en TikTok
Micro-targeting	Specifieke doelgroepen met een hen afgestemde boodschap bereiken
- Campagnesoftware	(Deels) automatische aansturing van micro-targeting
- Dynamic prospecting	Automatisch selecteren van doelgroepen
- Programatic advertising	Automatisch afstemmen van berichten op doelgroepen
- Psychographing	Automatisch analyseren van persoonlijkheidskenmerken
- Influencer marketing	Verspreiding berichten via accounts op sociale media met veel volgers
Chatapps	Berichten (versleuteld) uitwisselen, een-op-een of in kleine groepen
Bots	(Deels) automatisch aangestuurde accounts op sociale media
Zoekmachines	Platforms die het internet doorzoekbaar maken, zoekgedrag analyseren en advertenties tonen
Spraakassistenten	Spraakgestuurde apparaten waarmee onder andere zoekmachines worden geraadpleegd
Distributed autonomous applications	Onlineplatforms zonder centrale aansturing
Games	Online spellen
Crossmediale storytelling	Bereiken van een specifieke persoon of doelgroep via diverse kanalen en apparaten

Open Access
Het Rathenau Instituut heeft een Open Access beleid. Rapporten, achtergrondstudies, wetenschappelijke artikelen, software worden vrij beschikbaar gepubliceerd. Onderzoeksgegevens komen beschikbaar met inachtneming van wettelijke bepalingen en ethische normen voor onderzoek over rechten van derden, privacy, en auteursrecht.

Contactgegevens
Anna van Saksenlaan 51
Postbus 95366
2509 CJ Den Haag
070-342 15 42
info@rathenau.nl
www.rathenau.nl

Het Rathenau Instituut stimuleert de publieke en politieke meningsvorming over de maatschappelijke aspecten van wetenschap en technologie. We doen onderzoek en organiseren het debat over wetenschap, innovatie en nieuwe technologieën.

Rathenau Instituut