Aanvalsplan desinformatie.

Kamerlid Kees Verhoeven

D66

1. Introductie: desinformatie als maatschappelijk probleem.

Onze levens spelen zich steeds meer online af. Ons intensieve gebruik van social mediaplatformen is onder meer terug te zien in de 1 miljard Facebook berichten die dagelijks worden gepost, of de 500 uur film die iedere minuut op YouTube wordt geplaatst[1]. Deze berg aan berichten, video's en audio die mensen elke dag online consumeren geeft veel informatie en plezier, maar betekent ook een toenemend maatschappelijk risico.

Na een periode van hoge verwachtingen en hoopvolle visioenen is het tijd te erkennen dat social media ook scherpe schaduwzijden kennen. Zomaar een greep uit de negatieve gevolgen van techreuzen als Facebook, Twitter en Youtube: privacy-schendingen, informatiesturing, misleidende clickbaits, verkiezingsbeïnvloeding, maatschappelijke polarisatie, discriminatie en –steeds geavanceerdere- desinformatie over vaccineren, 5G-masten of het coronavirus[2].

Tegelijkertijd worden wij als maatschappij steeds afhankelijker van techreuzen. D66 heeft dit al vaker aangekaart, zoals in onze techvisies, het Actieplan Digitale Advertenties en in het pamflet 'Digitale Revolutie' over de datamacht van techreuzen[3]. Tijdens deze coronacrisis komt extra naar voren hoe zeer ook overheden afhankelijk zijn van private diensten als informatie- en communicatiekanalen. Bij deze grote impact van techbedrijven in de samenleving hoort volgens D66 ook een verantwoordelijkheid (zorgplicht) om desinformatie, discriminatie en polarisatie tegen te gaan. Het bestrijden van desinformatie is een taak van burgers en bedrijven, maar vanwege het ongrijpbare en manipulatieve karakter ervan moet de politiek, zowel Europees als nationaal, hier ook een actieve rol spelen.

Dat is een complexe opgave. Niet alleen is de schaal waarop desinformatie wordt verspreid toegenomen en de technologie erachter steeds slimmer, ook de uitingsvorm is de laatste jaren steeds geloofwaardiger en sluwer geworden[4].

Desinformatie is bedrieglijke informatie die gebruikt wordt om te beïnvloeden, te misleiden of schade toe te brengen aan mensen en organisaties. Desinformatie kent uiteenlopende verschijningsvormen, zoals onjuiste berichten over COVID-19, 5G-masten of vaccinatie; buitenlandse statelijke actoren die verkiezingen proberen te beïnvloeden door het zwartmaken van politici en partijen; politieke advertenties waarvan de afkomst niet bekend is; schimmige bitcoin-advertenties die ongevraagd de gezichten van BN'ers gebruiken om geld af te troggelen; complottheorieën over een kinderbloed drinkende elite (Qanon, deep state); en sinds een aantal jaar deepfakes, een nieuwe vorm van desinformatie die al onze zintuigen kan misleiden.

Wat zijn Deepfakes?[21]

Deepfakes zijn video's, berichten en audio die geautomatiseerd met het gebruik van kunstmatige intelligentie worden gemanipuleerd. Bij een Deepfake video wordt bijvoorbeeld iemands gezicht geplaatst op een andere video. Door stemmanipulatie kan het lijken alsof iemand woorden uitspreekt die hij of zij nooit heeft gezegd. Meer weten over deepfake-technologie? Lees dan: 'Deepfake Technology' The Infocalypse van Jarno Duursma.

Voorstellen

D66 wil dat het kabinet een aanvalsplan desinformatie opstelt. Wij doen een aantal voorstellen.

- Maak algoritmes openbaar en inzichtelijk voor de gebruiker. Stel een nationale algoritme-waakhond in en laat toezichthouders kritisch kijken naar de maatschappelijke impact.
- Niet-gepersonaliseerde advertenties moeten de standaard worden.
- Maak de samenleving digitaal vaardiger, door middel van lessen op school en

bewustwordingscampagnes.

- Stel een wettelijke zorgplicht in die techreuzen dwingt om duidelijke wetsovertredingen, zoals berichten die haatzaaien, oplichten of misleiden, binnen 24 uur te verwijderen.
- Dwing af dat deepfake-content beter herkenbaar wordt en investeer in deepfake-detectie.
- Leg politieke advertenties aan banden.
- Maak techbedrijven minder machtig door ze bij wetsovertredingen op te knippen of door fusies te verbieden.

2. Een pervers verdienmodel met twee poten.

Hoewel de nationale boekhandels AKO en Bruna onlangs besloten het complottheorie-tijdschrift ‘Gezond Verstand’ te verkopen, zijn Amerikaanse techreuzen en social mediaplatforms de voornaamste verspreiders van desinformatie. Ze zeggen desinformatie onwenselijk te vinden maar stellen niet direct verantwoordelijk te zijn voor de inhoud en ondernemen er relatief weinig actie tegen. Jarenlang deden ze niets en de afgelopen jaren is hun aanpak het best samen te vatten als selectieve symptoombestrijding. Facebook en Twitter verwijderden bijvoorbeeld anti-vaxx advertenties of suggestieve berichten over de Oekraïense activiteiten van Hunter Biden, maar deden niets aan de structurele oorzaak.

De reden voor deze passieve opstelling ligt in hun -op advertentieverkoop gebaseerde- verdienmodel. Dit verdienmodel speelt helaas een centrale rol in alle vormen van desinformatie en bestaat uit twee poten.

Allereerst hebben techreuzen als hoogste doel om mensen zo veel mogelijk tijd te laten doorbrengen op hun platforms, ongeacht de kwaliteit van de content. Zodat zoveel advertenties kunnen worden verkocht en zo veel mogelijk persoonsgegevens kunnen worden verzameld. Als mensen op de website blijven hangen, krijgen zij meer advertenties te zien, dus er wordt meer data verzameld en dus ontstaan nauwkeurigere gebruikersprofielen (*profiling*).

De tweede poot is het optimaal laten aansluiten van advertenties op de gebruikersvoorkeuren, gebaseerd op gebruikersprofielen: *targeting*. Hoe beter de advertenties aansluiten op de interesse van de ontvanger, hoe groter de kans dat die erop klikt. Het gevolg van dit tweeledige verdienmodel -meer advertenties en relevantere advertenties- is dat sensationele content voorrang krijgt van de algoritmes, want daar klikken mensen sneller op en die zorgt voor langere verblijftijd en dus voor meer dataverzameling[5].

Onder maatschappelijke druk halen de techreuzen soms dus een bericht of kanaal uit de lucht maar dit is vooral windowdressing om politici en adverteerders tevreden te houden. Het echte probleem zijn de algoritmes die hun heimelijke werk doen. Hier wordt al jaren niets aan gedaan,want dat is de kern van het verdienmodel. Hier speelt een belangrijk psychologisch gegeven: het is onzichtbaar dus lijkt het er niet te zijn. Terwijl het juist de machtige motor is die miljoenen mensen stuurt en miljarden euro's oplevert.

3. Politieke beïnvloeding, polarisatie, discriminatie, intimidatie.

Hoewel desinformatie grotendeels een online verschijning is, raken de effecten onze samenleving fysiek en materieel. Bij verkiezingen en op het vlak van de volksgezondheid kan desinformatie mensen op het verkeerde been zetten. Misleidende berichten over gevoelige onderwerpen wakkeren verdeeldheid en polarisatie aan. Ook kan desinformatie persoonlijke schade aanrichten, in de vorm van discriminatie, reputatieschade en identiteitsfraude, zoals nu gebeurt met deepfakes. De mensen die beweren dat desinformatie een verzonnen probleem is, of degenen die stellen dat het verspreiden van ontwrichtende informatie onderdeel is van de vrijheid van meningsuiting (wat klopt, maar wat zeker niet de enige invalshoek is) gaan voorbij aan de aantoonbare schade. Bovendien is geen enkel recht absoluut en kent ook de vrijheid van meningsuiting wettelijke beperkingen. Haatzaaien, laster, smaad, belediging, misleiding en bedrog zijn niet onbeperkt toegestaan. Ook bestaan er auteursrechtelijke beperkingen.

Desinformatie heeft bij verkiezingen in het Verenigd Koninkrijk (V.K.) en de Verenigde Staten (V.S.) een grote impact gehad, en ook bij de volgende verkiezingen in de V.S. heeft het hoofd van het *National Counterintelligence and Security Center* bekend gemaakt dat China, Rusland en Iran zich actief mengen in de verkiezingen, via het verspreiden van desinformatie met als doel mensen te ontmoedigen om te stemmen en om onrust te creëren[6]. In het V.K. werd na een twee jaar durend onderzoek door het Britse Parlement bekend dat er sprake was van advertenties waarbij niet duidelijk is wie de advertentie plaatste, ook wel *'dark ads'* genaamd. Deze advertenties bereikte meer dan 20 miljoen mensen en speelde in op angsten van mensen. Vluchtelingen zouden onderweg naar het V.K. zijn. Ook was er sprake van actieve misleiding;Turkije zou binnenkort onderdeel zijn van de Europese Unie (EU).

Desinformatie kan grote maatschappelijke onrust creëren. Dit kan bewust gebeuren, door statelijke actoren, versterkt door de algoritmes die de social mediaplatformen hanteren. Zoals eerder gesteld willen techreuzen dat gebruikers zo lang mogelijk tijd doorbrengen op hun platform. Als iemand een video kijkt over een controversieel onderwerp, dan wordt de gebruiker al snel andere controversiële video's aanbevolen. En middels de aanbevelingen vallen gebruikers steeds dieper in een *rabbit hole*, waar eenrichtingsverkeer van soortgelijke meningen vaak verweven zijn met complottheorieën en desinformatie. Zo belandt iemand op zoek naar een video over PCR-testen binnen twee kliks op complotvideos over de aanslag op 9/11 in New York of Qanon. Polarisatie en wantrouwen wordt op deze manier aangewakkerd.

Ook grootschalige discriminatie is een mogelijk gevolg, zo bleek afgelopen zomer, ten tijde van het protest op de Dam van Black Lives Matter. Unilever stopte toen met advertenties op Facebook nadat dochter Ben & Jerry's al eerder besloot geen advertenties meer te plaatsen op Facebook en Instagram [7]. Het ijsjesmerk sloot zich met die boycot aan bij de campagne *Stop Hate for Profit,* die Amerikaanse burgerrechtenorganisaties waren begonnen. Facebook krijg al langer kritiek van burgerrechtenbewegingen en activisten omdat ze te weinig doen om racistische en haatdragende uitingen op het platform te bestrijden. De burgerrechtenorganisaties namen Facebook onder meer kwalijk dat berichten die oproepen tot geweld tegen Black Lives Matter-demonstranten niet werden verwijderd. Ook het feit dat het bedrijf de uiterst rechtse website Breitbart als vertrouwde nieuwsbron aanmerkt, vinden ze een probleem. Na een advertentieboycot van tientallen bedrijven kondigde Facebook uiteindelijk aan zich actiever in te zetten om haatzaaien tegen te gaan in advertenties[8].

4. Deepfakes maken het nog geavanceerder.

Hoewel beïnvloeding, polarisatie en inmenging via online content al op grote schaal voorkomen, vormen deepfakes als geavanceerde uitingsvorm een extra risico. Deepfakes worden bijvoorbeeld offensief ingezet bij het afpersen en seksueel intimideren van vrouwen. Deepfakes worden gebruikt om de gezichten van vrouwen te verwerken in pornofilms, om deze fragmenten vervolgens online of in een gemeenschap te verspreiden. Dit komt nu al grootschalig voor met de gezichten van vrouwelijke beroemdheden, zoals gebeurde in een nepvideo van NOS-nieuwslezer Dionne Stax. Maar ook zien we dat deepfakes gebruikt worden om mensenrechtenactivisten in India de mond te snoeren en grote reputatieschade toe te brengen. Het wordt steeds goedkoper en gemakkelijker om deepfakes te produceren. Dit zal dus betekenen dat een deepfake van iemand - zoals een ex-partner met maar weinig foto of filmmateriaal - steeds laagdrempeliger kan worden gemaakt.

Zeker op het vlak van verkiezingen vormen deepfakes een risico. Straks kunnen deepfakes kiezers doen geloven dat een politicus iets heeft beweert wat de politicus nooit heeft gezegd. Daarnaast bestaat het risico van het 'leugenaars-dividend': mocht een politicus (of ander publiek figuur) geconfronteerd worden met zijn of haar eigen uitspraken, dan kunnen ze de uitspraak verwijten aan een deepfake of andere manipulatie. Op deze manier wordt bewijsmateriaal snel minder betrouwbaar geacht. De Amerikaanse president Trump die elke onwelgevallig bericht afdoet als *fake news* is het symbool voor deze praktijk.

Computerwetenschapper Hanry Farid van de Universiteit van Berkeley omschrijft waarom - door drie ingrediënten - deepfakes nu een extra grote dreiging vormen[9]: Allereerst wordt het manipuleren van video, geluid en teksten met deepfakes in grote mate geautomatiseerd met kunstmatige intelligentie. In het analoge tijdperk was het bewerken van foto's een tijdsintensief en specialistisch werk dat plaatsvond in donkere kamers. Met de introductie van photoshop werd het bewerken van

foto's al toegankelijker voor iedereen in het bezit van een computer, echter was er nog steeds veel technische kennis nodig om realistisch foto's te bewerken. Met de snelle ontwikkeling van kunstmatige intelligentie is de lat om zelf video of geluid te manipuleren verlaagd. Het maken van deepfakes wordt steeds goedkoper, zodat straks iedereen zijn eigen materiaal kan bewerken.

Ten tweede maakt sociale media verspreiding van informatie heel makkelijk en snel. Iedereen kan ieder moment berichten plaatsen of video's uploaden. Ophef en woede zorgen vaak voor een hogere betrokkenheid dan genuanceerde berichten die geen sterke emoties oproepen. De algoritmes van techreuzen zijn daarom bewust ingericht om dit soort berichten vaker aan mensen te tonen. De relevantie van de meeste berichten op sociale media wordt tegenwoordig niet gemeten in dagen, maar in uren. Het rectificeren van desinformatie is daardoor lastig. Vaak komen rectificaties te laat en bereiken maar een beperkt publiek[10].

Het derde ingrediënt zijn de gebruikers van sociale media zelf. Gebruikers zijn steeds meer bereid om nieuws tot zich te nemen dat aansluit bij hun wereldbeeld. Algoritmes creëren echokamers van soortgelijke meningen, zodat informatie die mensen te zien krijgen steeds meer gaat aansluiten bij eerder geuite interesses. Op die manier blijven gebruikers langer hangen. En desinformatie die inspeelt op vermoedens of vooringenomenheid, zal makkelijker verspreid worden omdat dit het bestaande wereldbeeld van gebruikers bevestigd. Zo komen mensen in een rabbit hole terecht, een vicieuze cirkel waarin gebruikers hun wereldbeeld zien bevestigd worden doordat ze steeds meer van dezelfde informatie voorgeschoteld krijgen. En zo worden mensen ook steeds vatbaarder voor (des)informatie die hierop aansluit.

5. Hoe is het mogelijk? Polarisatie als perverse prikkel!

De belofte van sociale media was ooit om mensen dichter bij elkaar te brengen. Maar sociale media halen ook iets slechts in ons naar boven. Techplatformen gebruiken onzichtbare algoritmes die ons emotionele brein misleiden, die het sterk beïnvloedbare individu informatie voorschotelen met als doel hem of haar langer en intensiever op hun platform te houden. Dit technologische en psychologische hoogstandje levert veel kijkplezier en vele miljarden op maar er wordt geen rekening gehouden met de nadelige maatschappelijke gevolgen van de berichten, noch verantwoordelijkheid genomen voor de haat, discriminatie en racisme die ze teweeg brengen.

Sterker nog: het lijkt soms alsof techreuzen mensen actief stimuleren in verharding van hun standpunten of zelfs tot radicalisering aanzetten[11]. Uit intern onderzoek van Facebook kwam naar voren dat 64% van de gebruikers die zich aansloot bij een extremistische Facebook groep, dit deed op advies van het Facebook algoritme[12].

Op YouTube, in bezit van Google, selecteren algoritmes niet de video's die het meest relevant zijn voor een gebruiker, of het meest interessant. Het algoritme kiest voor video's die gebruikers zo veel mogelijk tijd laten doorbrengen op YouTube. Het YouTube algoritme is zo succesvol, dat 70% van de tijd doorgebracht op YouTube komt van het kijken naar aanbevolen video's[13]. Algoritmes hebben –via *machine learning*- geleerd dat controversie, hevige emoties, sappig nieuws, complotten en nieuwigheid ons langer op platformen houden. De gevolgen hiervan worden niet meegewogen. Bij Facebook werd intern gerapporteerd dat "ons algoritme de voorkeur van het menselijk brein naar polarisatie uitbuit", en dat als hier niets tegen wordt ondernomen, Facebook gebruikers zou blijven voeden met "meer en meer polariserende content zodat de aandacht en tijdsbesteding van gebruikers toeneemt."[14]

Techreuzen doen er dus alles aan om zoveel mogelijk informatie te verzamelen over mensen. De algoritmes bepalen vervolgens wat we zien. Ze zijn zo sluw ingericht dat mensen vrijwel uitsluitend informatie zien die aansluit op eerder geuite interesses en het geeft sensationele content voorrang. Daar klikken mensen namelijk sneller op, zo wijst herhaaldelijk onderzoek uit. Het gaat de techreuzen om zoveel mogelijk *clicks, shares en likes*, zodat mensen meer advertenties te zien krijgen, en zo wordt kwaliteit en nuance verdrongen door ophef en boosheid. Waarom in hemelsnaam?

Begin 2000 klapte de dotcombubbel omdat techbedrijven geen verdienmodel hadden. Gepersonaliseerde advertenties konden later dit probleem oplossen: door de enorme hoeveelheid data die social media verzamelen van mensen, kunnen advertenties steeds beter worden toegespitst op specifieke bezoeker.

Dat techbedrijven tegenwoordig genoeg verdienen, betwijfelt niemand nog. Apple, Google en Facebook behoren tot de rijkste bedrijven ter wereld. Om deze positie te behouden is het voor hen beter om niet structureel op te treden tegen desinformatie. Immers: Desinformatie, smeuïg nieuws, polariserende video's houden de aandacht vast en zorgen voor voldoende mogelijkheid om advertenties te tonen[15].

De perverse prikkel om polarisatie toe te laten is een geval waarbij een commercieel belang (verdienen) en een maatschappelijk belang (verbinden) lijnrecht tegenover elkaar staan. Omdat gebruikers zeer verschillend naar dit fenomeen kijken (Leugens! Censuur!), omdat het ongrijpbare en grenzeloze karakter ingrijpen complex maakt en omdat de lobby van techreuzen sterk is, heeft dit lang kunnen voortduren. Maar het is tijd om deze praktijk een halt toe te roepen.

6. Een aanvalsplan tegen desinformatie.

De afgelopen jaren beloofden de techreuzen na ieder schandaal – op het gebied van privacyschendingen, schadelijke desinformatie, politieke inmenging en discriminatie - telkens beterschap. Maar steeds *to little, too late*: lang nadat de schade al was aangericht en het geld al binnen was[16]. We hebben nu lang genoeg gezien welke negatieve -individuele en maatschappelijke- impact desinformatie kan hebben. Daarom is het tijd dat politici -nationaal en Europees- gericht ingrijpen. Het feit dat nieuwe technologieën deze problemen creëren betekent niet dat technologie ook de enige oplossing vormt. Om toekomstige dreigingen aan te pakken moeten er zowel technische als politieke oplossingen komen. Om de kern aan te pakken moeten de twee poten van het verdienmodel onderuit gehaald worden. Zowel de algoritmes om je op het platform te houden als de gepersonaliseerde advertenties moeten ingeperkt worden.

De onzichtbare hand aanpakken

Allereerst is het nodig de werking van algoritmes bloot te leggen en inzichtelijk te maken. In de strijd tegen desinformatie is het verwijderen van informatie zelf niet voldoende. Het is ook zaak te reguleren hoe informatie ons bereikt. Met inzichtelijke algoritmes wordt duidelijk waarom een algoritme bepaalde beslissingen maakt, op basis van welke gegevens. Op deze manier kan verantwoording worden afgelegd waarom 64% van de mensen die lid wordt van een extremistische Facebook pagina, dit doet op basis van een algoritmische aanbeveling[17], of waarom sommige gebruikers vaker desinformatie aanbevolen krijgen dan anderen. Zulke uitwassen kunnen dan ook sneller verholpen worden. Het argument dat de algoritmes bedrijfsgeheim zijn gaat voorbij aan de maatschappelijke schade. Een economisch argument moet in verhouding staan tot de sociale gevolgen. Dat is nu niet meer het geval.

Om te weten wat er gebeurt moeten mensen zien dat er een algoritme aan het werk is en waarom het bepaalde informatie voorschotelt. Dit kan bijvoorbeeld gaan om

een leeftijdscategorie, geslacht of specifieke interesse. Dit zou voor alle informatie moeten gelden en zou in *plain text* kunnen worden aangegeven. Bewustwording onder gebruikers is echter niet voldoende.

Grote platformen weten zelf misschien niet eens waarom het algoritme mensen bepaalde informatie laat zien of een pagina adviseert. Daarom moet de werking van het algoritme door derden controleerbaar zijn. De politiek kan dit zelf doen maar beter is om onafhankelijke toezichthouders op te roepen een gezamenlijk onderzoek doen naar de gevolgen van algoritmes. Dit moet zowel op het gebied van privacy, non-discriminatie, economische rechtvaardigheid als maatschappelijke gevolgen. Onder leiding van D66 heeft de Kamer het kabinet opgeroepen tot het instellen van een algoritme-waakhond, maar zolang die er niet is zullen de Autoriteit Persoonsgegevens, het College voor de Rechten van de Mens, de Autoriteit Consument en Markt en de Rekenkamer een gezamenlijk onderzoek moeten uitvoeren naar de brede maatschappelijke gevolgen van algoritmes en hoe social mediagebruikers te beschermen.

Het perverse verdienmodel aanpassen

Ten tweede moeten niet-gepersonaliseerde advertenties de standaard worden. Dat betekent dat advertenties gebaseerd worden op de context van de website, in plaats van op de persoonsgegevens van de bezoeker. Zulke advertenties leveren vergelijkbare inkomsten op, maar brengen een halt toe aan de datahonger van techreuzen[18]. Wie toch wel graag gepersonaliseerde advertenties wil zien, kan dit zelf aangeven. Maar de standaardinstelling moet niet-gepersonaliseerd zijn en degene die dit wil veranderen moet moeite doen.

Niet-gepersonaliseerde advertenties worden steeds vaker de norm, door het toenemende gebruik van adblockers en browsers als Firefox die standaard third-party tracking cookies blokkeert. Het Europees Parlement heeft recent uitgesproken dat er meer grip moet komen op gepersonaliseerde advertenties[19]. Nederland moet het voortouw nemen om dit EU-breed vast te leggen.

Digitale vaardigheden en bewustwording

Om desinformatie de wind uit de zeilen te nemen, is het belangrijk dat mensen digitaal vaardig zijn en leren om verstandig met online informatie om te gaan. Zo worden mensen minder vatbaar voor desinformatie. Wie bekend is met het bestaan van deepfakes, zal hier beter tegen gewapend zijn. Voor kinderen is het daarom belangrijk om op onderwijsinstellingen les te krijgen in digitale vaardigheden. Naast enorme kansen kent het internetrisico's op het gebied van privacy en veiligheid. Door kinderen op school digitale vaardigheden te leren kunnen zij de digitale wereld beter begrijpen en er op een zorgvuldige wijze mee omgaan. Op aandringen van D66 en CDA heeft het kabinet eerder al publiekscampagnes opgezet om mensen te informeren over het gevaar van desinformatie. Zulke campagnes blijven belangrijk om de bewustwording van desinformatie te vergroten.

Zorgplicht om desinformatie te bestrijden

Platformen zijn goed in het verwijderen van inhoud die tegen hun verdienmodel indruisen, zoals inhoud dat onder een licentie valt. Ook pornografie of kindermisbruik worden snel verwijderd, omdat adverteerders dit niet graag naast hun advertenties zien. Maar als het om desinformatie gaat dan hebben platformen te vaak en te lang geen verantwoordelijkheid genomen. Een zorgplicht die techreuzen dwingt om duidelijke wetsovertredingen, zoals haatzaaien, oplichting of misleiding, te verwijderen binnen 24 uur kan hier verandering brengen. Zo wordt de snelle verspreiding van desinformatie een halt toegeroepen. Gezien het gevaar van censuur en inperking van de vrijheid van meningsuiting moet dit zeer zorgvuldig[20]. Bovendien bestaat er al wetgeving rondom haatzaaien, smaad, laster, belediging, bedrog, misleiding en auteursrechtelijk beschermd materiaal. Deze wetgeving moet worden toegepast. Techbedrijven moeten hiervoor verantwoordelijk worden gehouden.

Organisaties zoals #stopprofitfromhate kaarten al jaren aan hoe grote techreuzen profiteren van informatie met een haatdragende of misleidende boodschap. Techreuzen vervullen een steeds grotere rol in het sociale contact tussen mensen, en is een belangrijk informatiekanaal geworden van de overheid. Vanwege de grote informatieve functie, dragen techreuzen verantwoordelijkheid en een zorgplicht om

discriminatie en haat te bestrijden op hun platformen. Een aantal stappen kunnen hierbij helpen:

Allereerst moet er een apart moderatietraject komen voor mensen die gericht worden aangevallen op basis van identiteitskenmerken zoals ras, religie of seksuele oriëntatie. Hier kunnen experts op het gebied van discriminatie, haatzaaien en ander strafbaar gedrag de klachten behandelen. Daarnaast zijn interne mechanismes van groot belang die automatisch informatie kan signaleren in gesloten groepen die overeenkomen met extreme ideologieën. Deze informatie en de geassocieerde groep kan dan door een menselijke moderator beoordeeld worden. Als derde moeten gebruikers een melding krijgen als zij interacties hebben gehad met berichten of groepen die zijn verwijderd vanwege desinformatie, discriminatie of haatzaaien. Een soortgelijk meldingssysteem bestaat nu al voor gebruikers die interactie hebben gehad met coronahoaxes. Tenslotte is het belangrijk dat techreuzen het aantal meldingen van desinformatie, racisme en discriminatie per kwartaal openbaar maken, en toelichten hoe er met de meldingen is omgegaan.

In Duitsland kunnen gebruikers sinds kort wetsovertredingen melden op techplatformen. Techreuzen zoals Twitter, Facebook als Google hebben hiervoor speciale meldprocedures gemaakt. Op deze manier kan er gemeld worden of er sprake is van belediging, haatzaaien of opruiing. Na de melding is het aan de techreuzen om de inhoud te beoordelen. Volgens een Duitse wet moeten overduidelijke wetsovertredingen binnen 24 uur verwijderd zijn. Bij berichten waar een wetsovertreding minder duidelijk is, geldt een periode van 7 dagen. Als er herhaaldelijk niets met meldingen wordt gedaan, of als een meldprocedure ontbreekt, dan kunnen de techreuzen een boete ontvangen tot 50.000 euro. D66 wil dat ook Nederland dergelijke wetgeving serieus overweegt.

Deepfakes verwijderen

Deepfakes brengen nieuwe risico's mee voor individu en maatschappij. Sommige social mediaplatformen zijn daarom zelf begonnen met het nemen van maatregelen. Zo heeft Facebook een Deepfake detection challenge gehouden, een open competitie om algoritmes te ontwikkelen die deepfakes kunnen identificeren. Het lukt echter pas om twee derde van de deepfakes te identificeren. Platformen zullen dus meer moeten investeren in deepfake detectie tools en factcheckers om aan hun zorgplicht te kunnen voldoen en misleiding tegen te gaan.

Politieke advertenties aan banden leggen

Het is belangrijk dat we weten wie politieke advertenties plaatst. Dit zal dus openbaar en herleidbaar moeten zijn. Ook moet het transparant en openbaar worden op welke manier en op basis van welke kenmerken politieke advertenties worden gericht aan gebruikers. Nu is het nog enkel achteraf te controleren welke doelgroep de politiek advertentie heeft bereikt.

Techbedrijven minder machtig maken

Parallel aan het advertentie-verdienmodel is er een tweede probleem: de techreuzen zijn te machtig. Een handjevol Amerikaanse en Chinese bedrijven domineren de markt. Nieuwe bedrijven worden al opgekocht of kapot geconcurreerd door de huidige techreuzen. De enorme datavoorsprong van de techreuzen maakt concurreren zeer lastig voor nieuwe opkomende platformen. Innovatie wordt zo de kop ingedrukt. Er is daardoor geen ruimte noch zuurstof voor nieuwe platformen die betere bedrijfsmodellen en beleid hebben. In zowel de V.S. als in de EU lijkt er daarom steeds meer steun te komen voor het opbreken van de techreuzen om hun macht te beperken. De Tweede Kamer heeft dit jaar de D66 initiatiefnota over digitale mededinging aangenomen, inclusief voorstellen die het mogelijk maken om fusies te verbieden en bedrijven in bepaalde gevallen op te knippen. Het kabinet moet hier binnen de EU meer werk van maken.

Eindnoten

1 Bron: Jarno Duursma

2 De afgelopen jaren vonden herhaaldelijk incidenten plaats: het Cambridge Analytica schandaal, de politieke advertenties op Facebook tijdens de Amerikaanse presidentsverkiezingen 2016, genderdiscriminatie waarbij bepaalde (leidinggevende) vacatures alleen aan mannen getoond werden, misleidende Bitcoin-advertenties met bekende Nederlanders, onjuiste anti-vaxx berichtgeving, de advertentie-boycot van veel bedrijven vanwege een gebrek aan actie tegen racistische content op social media platforms, misbruik van vrouwen in deepfake-video's zoals onder meer Dionne Stax overkwam op de website Pornhub en de brede verspreiding van complottheorieën over Corona en Qanon.

3 Beide plannen zijn te lezen via: https://oud.d66.nl/content/uploads/sites/2/2019/06/ActieplanDigitalAdvertising.pdf en https://d66.nl/digitale-revolutie/

4 Dit is vergelijkbaar met zogeheten phishing mails. In het begin waren dit eenvoudig herkenbare en te vermijden berichten vol typefouten, tegenwoordig zijn het sms-berichten en mails die haast niet van het echt zijn te onderscheiden. Met als gevolg dat steeds meer mensen erin trappen.

5 D66 wil gepersonaliseerde advertenties terugdringen, FD, 25 juni 2019

6 U.S. counterspy chief warns Russia, China, Iran trying to meddle in 2020 election, Reuters, 7 Augustus 2020

7 https://www.parool.nl/wereld/unilever-stopt-met-adverteren-op-facebook-en-twitter-in-vs~bd6967a6/

8 Facebook past beleid aan na advertentieboycots, NRC, 26 juni 2020

9 Hany Farid on Deep Fakes, Doctored Photos and Disinformation, the Lawfare Podcast

10 The spread of true and false news online, Soroush Vosoughi, Science; Vol 359, Issue 6380,9 Maart 2018

11 https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html

12 Facebook Executives Shut Down Efforts to Make the Site Less Divisive, Wall Street Journal, 26 Mei 2020

13 YouTube's AI is the puppet master over most of what you watch, Thomas Hornigold, 17 Oktober 2019

14 Facebook Executives Shut Down Efforts to Make the Site Less Divisive, Wall Street Journal, 26 Mei 2020

15 Enerzijds is het voor adverteerders belangrijk dat de advertenties zo nauw mogelijk aansluiten bij iemands voorkeuren. Wie van zeilen houdt krijgt daarom veel zeiladvertenties te zien. Anderzijds worden algoritmes gebruikt om gebruikers zo veel mogelijk tijd te laten besteden social media platformen. Algoritmes bepalen welke berichten wel of niet op onze timeline verschijnen.

16 Het Facebook motto was dan ook jarenlang 'move fast and break things'. Dat van Google Don't be evil.

17 Facebook Executives Shut Down Efforts to Make the Site Less Divisive, Wall Street Journal, 26 Mei 2020

18 https://tweakers.net/nieuws/153486/advertenties-met-trackingcookies-leveren-vier-procent-meer-op-dan-gewone-ads.html

19 https://www.europarl.europa.eu/doceo/document/TA-9-2020-0158_EN.html

20 Het is niet de bedoeling dat overijverige moderatie leidt tot censuur, noch dat het nalaten van wetgeving volledig wordt geprivatiseerd richting de techreuzen.

21 'Deepfake Technology' The Infocalypse, Jarno Duursma, te lezen op: https://www.jarnoduursma.nl/app/uploads/2019/09/Jarno-Duursma-_-Deepfake-Technologie-The-Infocalypse.pdf