OPINIE ZATERDAG PAGINA 56 | ZATERDAG 21 NOVEMBER 2020

Pak deepfakes aan voor het te laat is

Misleiding

Nu de technologie om media te vervalsen nog in de kinderschoenen staat, moeten we snel bepalen hoe die mag worden gebruikt, waarschuwt Nina Schick, desinformatie-expert.

MARJOLEIN VAN TRIGT

NINA SCHICK

De succesvolste desinformatiecampagne van 2020 was niet afkomstig van Russische trollen, antivaxers of QAnon-aanhangers. Ze kwam uit het Witte Huis, stelde desinformatie-expert Nina Schick (33) in de loop van het jaar met groeiende ongerustheid vast. In maart begonnen president Trump en zijn aanhang met het verspreiden van het verhaal dat de verkiezingen doorgestoken kaart waren. Hoewel er tot nu toe geen bewijs voor is gevonden, wist Trump ruim 70 procent van de Republikeinen te overtuigen dat er grootschalige verkiezingsfraude is gepleegd.

We stevenen af op een 'infocalyps', zegt Schick, een samenleving waarin desinformatie en misinformatie welig kunnen tieren en bevolkingsgroepen uit elkaar drijven. Schick was politiek adviseur voor de Britse regering over de Brexit en de immigratiecrisis. Ook werkte ze voor de verkiezingscampagne van de Franse president Emmanuel Macron. Tegenwoordig is ze schrijver en onafhankelijk politiek commentator voor onder meer de BBC, CNN en Al Jazeera. In de aanloop naar de Amerikaanse verkiezingen verscheen haar boek Deepfakes and the Infocalypse - What You Urgently Need to Know.

Schick maakt onderscheid tussen desinformatie, die bewust is gemaakt om mensen te misleiden, en misinformatie, slechte informatie die goedbedoeld wordt verspreid. Leugens zijn zo oud als de mensheid, schrijft ze. Relatief nieuw is de technologie waarmee ze zich razendsnel verspreiden: sociale media en berichtendiensten zoals WhatsApp. Volgens haar is typisch voor de infocalyps dat het steeds moeilijker wordt om een consensus te bereiken over hoe de wereld moet worden gezien. Of het nu gaat om Black Lives Matter, abortus of corona, het voelt al snel alsof er partij moet worden gekozen.

Al nagelbijtend heeft ze vanuit Londen de Amerikaanse verkiezingen gevolgd, vertelt Schick via Skype. Wat er in de Verenigde Staten gebeurt op het gebied van desinformatie, zet de toon voor de rest van de wereld, schrijft ze in haar boek. In 2018 adviseerde ze op uitnodiging van Anders Fogh Rasmussen, voormalig secretaris-generaal van de Navo, een groep wereldleiders over inmenging in de verkiezingen en desinformatie. Ook Joe Biden was daarbij. 'We hadden op dat moment al veel kennis over de manier waarop met name de Russen online desinformatie verspreidden om chaos te veroorzaken', zegt Schick. 'We voorzagen niet hoe succesvol Trump verdeeldheid zou zaaien met zijn claims over ver-

kiezingsfraude. Mogelijk zijn we van Trump af, maar dat zaadje is geplant. En het zal blijven groeien, daar ben ik zeker van.'

De term infocalyps is bedacht door de Amerikaanse technologie-onderzoeker Aviv Ovadya. Een samenleving kan kapotgaan aan een vloedgolf van slechte informatie, waarschuwde hij in 2016. Naast al het nepnieuws doemt aan de horizon een nog veel groter gevaar op: synthetische media. Dat zijn afbeeldingen, stukken tekst, audio- of videofragmenten die zijn aangepast of zelfs helemaal gegenereerd door kunstmatige intelligentie (AI). Vaak worden synthetische media 'deepfakes' genoemd. Schick bewaart die term uitsluitend voor synthetische media die bewust worden ingezet om desinformatie te verspreiden.

'Niet alle gevolgen van de komst van synthetische media zijn slecht', zegt ze. 'De filmindustrie zal ervan profiteren, net als de mode- en game-industrie. Over tien jaar kan een tiener met een smartphone dezelfde soort visuele effecten creëren waarvoor een Hollywoodstudio nu nog een miljoenenbudget en een team van specialisten nodig heeft. We zullen allemaal consumenten en producenten worden van door AI gegenereerde media. Dat betekent helaas ook dat synthetische media een grote rol gaan spelen bij de verspreiding van mis- en desinformatie.'

Wanneer realiseerde u zich het gevaar van synthetische media?
'Zodra de eerste 'face swaps' op internet verschenen, drie jaar geleden. Iemand had AI getraind om de gezichten van beroemde actrices op een ander lichaam te plaatsen. Tot op dat moment was het heel moeilijk om geloofwaardige nepmensen te maken. De eerste kwaadaardige inzet van deepfakes was gericht tegen vrouwen. Hun gezichten werden op die van pornoactrices geplaatst - zonder hun goedkeuring uiteraard. 96 procent van de bestaande deepfakes is porno. Niet alleen van beroemdheden, maar ook van gewone vrouwen. Met wat foto's en video's, een opname van je stem, kan ik AI trainen om precies zoals jij te klinken en eruit te zien. Dat was niet mogelijk met de computereffecten die we voorheen tot onze beschikking hadden. Niet alleen is de techniek nu beter, ze wordt ook gratis en breed beschikbaar. Die combinatie is gevaarlijk.'

Tegenstanders van de Indiase journalist en moslima Rana Ayyub verspreidden een deepfake- pornovideo van haar via WhatsApp om haar monddood te maken. Bent u bang dat u door dit boek ook doelwit wordt van deepfakemakers?

'Nee. Ik kan mezelf verdedigen, ik heb een platform. Maar de onbekende vrouwen die slachtoffer worden van deepfakeporno hebben geen juridisch team klaarstaan. Ironisch genoeg maak je momenteel het meeste kans om een deepfakepornofilm te laten verwijderen door erop te wijzen dat het copyright van de filmstudio wordt geschonden.'

Zijn deepfakes vooral gevaarlijk voor vrouwen?
'Zoals zo vaak in het internettijdperk staan vrouwen in de frontlinie, maar naarmate deze technologie breder beschikbaar wordt, zal ze op veel meer manieren worden misbruikt. Er zijn al fraudegevallen met deepfaketechnologie bekend. Een Brits energiebedrijf raakte 250 duizend euro kwijt aan oplichters die een deepfake hadden gebruikt van de stem van de Duitse directeur.'

De makers van South Park lanceerden onlangs de eerste 'deepfakecomedy', met rollen voor onder meer Mark Zuckerberg en Julie Andrews. Is dat volgens uw definitie geen deepfake?
'Het is een van de beste voorbeelden van wat je op dit moment met de technologie kunt doen. Ik zou het inderdaad strikt gezien geen deepfake noemen, omdat de technologie wordt gebruikt voor satire. Een ander voorbeeld van goedaardig gebruik van synthetische media komt van het Londense bedrijf Synthesia. Met een paar muisklikken regelen zij dat je jouw videoboodschap in verschillende talen uitspreekt.'

U beschrijft hoe de leider van de Indiase Bharatiya Janata Party (BJP) soortgelijke technologie inzette voor een politieke campagne. In video's leek hij allerlei dialecten te spreken die hij in feite niet kende. Is dat niet gewoon kiezersbedrog?

'Je kunt synthetische media best inzetten om een grotere groep kiezers te bereiken, zolang je het maar duidelijk aangeeft en mensen niet voor de gek houdt. Het is belangrijk dat we hier een gesprek over aangaan nu de technologie nog in de kinderschoenen staat. Nu kunnen we nog bepalen hoe het mag worden gebruikt. Het duurt nog een paar jaar voordat het alomtegenwoordig zal zijn, maar dát het alomtegenwoordig zal zijn, staat vast.'

Als politiek adviseur zag Schick de invloed van desinformatie en misinformatie op de wereldpolitiek groeien. Haar eerste baan was bij een Britse denktank over een toen nog obscuur onderwerp: de Brits-Europese betrekkingen. Door het Brexitreferendum raakte haar carrière in een stroomversnelling. 'Rond die tijd raakte ik geïnteresseerd in de infocalyps. De informatiewapenwedloop vond plaats onder mijn neus.' In haar boek beschrijft ze hoe het Kremlin in 2013 de Internet Research Agency (IRA) opzette, een onderdeel van de geheime dienst, met als missie het infiltreren van het publieke debat in westerse landen via sociale media. Met behulp van trollen, nepnieuws en nep-Facebookgroepen wisten de Russen verdeeldheid te zaaien over onder meer de Krimoorlog, de immigratiecrisis, het Brexitreferendum en de Amerikaanse verkiezingen van 2016. Ook de presidentscampagne van Emmanuel Macron, waarvoor Schick in 2017 ging werken, kreeg te maken met een Russische hackoperatie, waarbij gestolen documenten van het campagneteam samen met nepdocumenten online werden gezet.

Haar Nepalese moeder groeide op zonder elektriciteit, sanitair of stromend water, zei Schick in de podcast Making Sense with Sam Harris. Zelf studeerde ze aan de Universiteit van Cambridge en University College London, spreekt zeven talen en adviseert wereldleiders - nogal een stap in één generatie. 'Pas nu ik wat ouder ben, realiseer ik me hoe ongewoon mijn achtergrond is.' Haar Duitse vader, een strafrechtadvocaat, reed in de jaren zeventig op de bonnefooi naar Nepal, dat kort daarvoor de grenzen had geopend. Hij besloot er te blijven om de Himalayaanse cultuur te documenteren. 'Zo ontmoette hij mijn moeder, die nooit naar school was gegaan, omdat dat verboden was voor meisjes. Door mijn vader leerde ze het Westen kennen.'

Schick en haar broer groeiden op in de grote internationale gemeenschap in Kathmandu. Tijdens hun jeugd was de Nepalese politiek tumultueus, met opstanden, burgeroorlog en een maoistisch bewind. 'Ik werd erg aangetrokken door het westerse politiek sys-

teem, dat daadwerkelijk leek te functioneren.'

Onder andere China en Iran kijken de kunst van het verspreiden van desinformatie inmiddels van Rusland af, met als doel het destabiliseren van andere landen. Maakt u zich zorgen over landen als Nepal?
'Zeker. We zijn de afgelopen tien jaar in het Westen echt slecht geweest in het aanpakken van de crisis van de mis- en desinformatie. Maar wij hebben institutionele waarborgen, zoals de vrije pers. In de Verenigde Staten bestonden er zorgen dat Trump zichzelf en zijn familie voor eeuwig op de troon zou hijsen met zijn desinformatiecampagne. Hij heeft een poging gedaan, maar het systeem werkt. Alle stemmen zijn geteld en alle media zeggen dat Biden heeft gewonnen, ook Fox News.

'Zou dat ook gebeuren in een land waar de democratische instituties niet zo sterk zijn? In delen van de wereld, zoals Nepal, leven veel mensen grotendeels zoals driehonderd jaar geleden, maar iedereen heeft een smartphone. Ze betreden het informatie-ecosysteem zonder beschermingsmiddelen, met nauwelijks digitale geletterdheid. In Myanmar, waar nepnieuws op Facebook werd gebruikt om een genocide tegen de Rohingya te ontketenen, zag je hoe rampzalig dat kan uitpakken. Iets soortgelijks gebeurt in Ethiopië. Politiek gemotiveerde onzinberichten op Facebook leidden daar tot de moord op zanger Hachalu Hundessa, met een geweldsspiraal tot gevolg. En dan hebben we het nu nog over 'cheapfakes': video's en afbeeldingen die uit hun context worden gehaald of bewerkt. Dat zijn nog geen deepfakes.'

Er is niet één manier om deepfakes aan te pakken, schrijft Schick. Het is belangrijk dat meer mensen begrijpen wat deepfakes zijn en hoe ze verschillen van synthetische media. Daarnaast moeten burgers zich kunnen verdedigen tegen deepfakes - met behulp van journalistiek, wetenschap en organisaties die desinformatie in kaart brengen, maar ook met detectiesoftware en technologie die laat zien waar en wanneer iets is gemaakt. Ten slotte hebben ze een 'psychologische verdedigingsmuur' nodig, waar desinformatie niet zomaar langskomt.

Ter voorbereiding van dit artikel sprak ik met Max Welling, hoogleraar Machine Learning aan de Universiteit van Amsterdam. Hij verwacht dat deepfakes op den duur zo goed worden dat ze niet meer met technologie te detecteren zijn. Denkt u dat ook?

'AI-experts zijn het daar nog niet over eens. Op dit moment zijn deepfakes goed genoeg om mensen voor de gek te houden. Kijk maar eens op thispersondoesnotexist.com. Met elke klik kun je een nieuwe afbeelding van een niet-bestaand mens genereren, helemaal opgebouwd door AI. Daarom wordt overal ter wereld detectiesoftware gebouwd. Maar de oplossing moet zeker niet alleen van technologie komen.'

Welling denkt niet dat het probleem onoplosbaar is. Volgen hem moeten we gaan leren om afbeeldingen, geluiden en video's net zo kritisch te benaderen als teksten. We verwachten van een tekst ook niet automatisch dat deze de waarheid bevat.
'Wanneer we niet meer kunnen geloven wat we zien, zullen mensen ook authentieke media niet meer geloven. Een van de problemen van deepfakes is the liar's dividend - als alles kan worden vervalst, kan alles worden ontkend. Dat geeft slechteriken veel speelruimte.

'Vlak nadat ik mijn manuscript had ingeleverd, beweerde een Republikeinse kandidaat voor het Huis van Afgevaardigden dat de video waarin we zien hoe een agent met zijn knie op de nek van George Floyd drukt een deepfake is. In de toekomst zullen we veel van dat soort geluiden horen. Dus ja, we moeten kritischer worden, maar in geen geval cynisch. Anders geloof je alleen nog dat wat past bij je overtuigingen.'