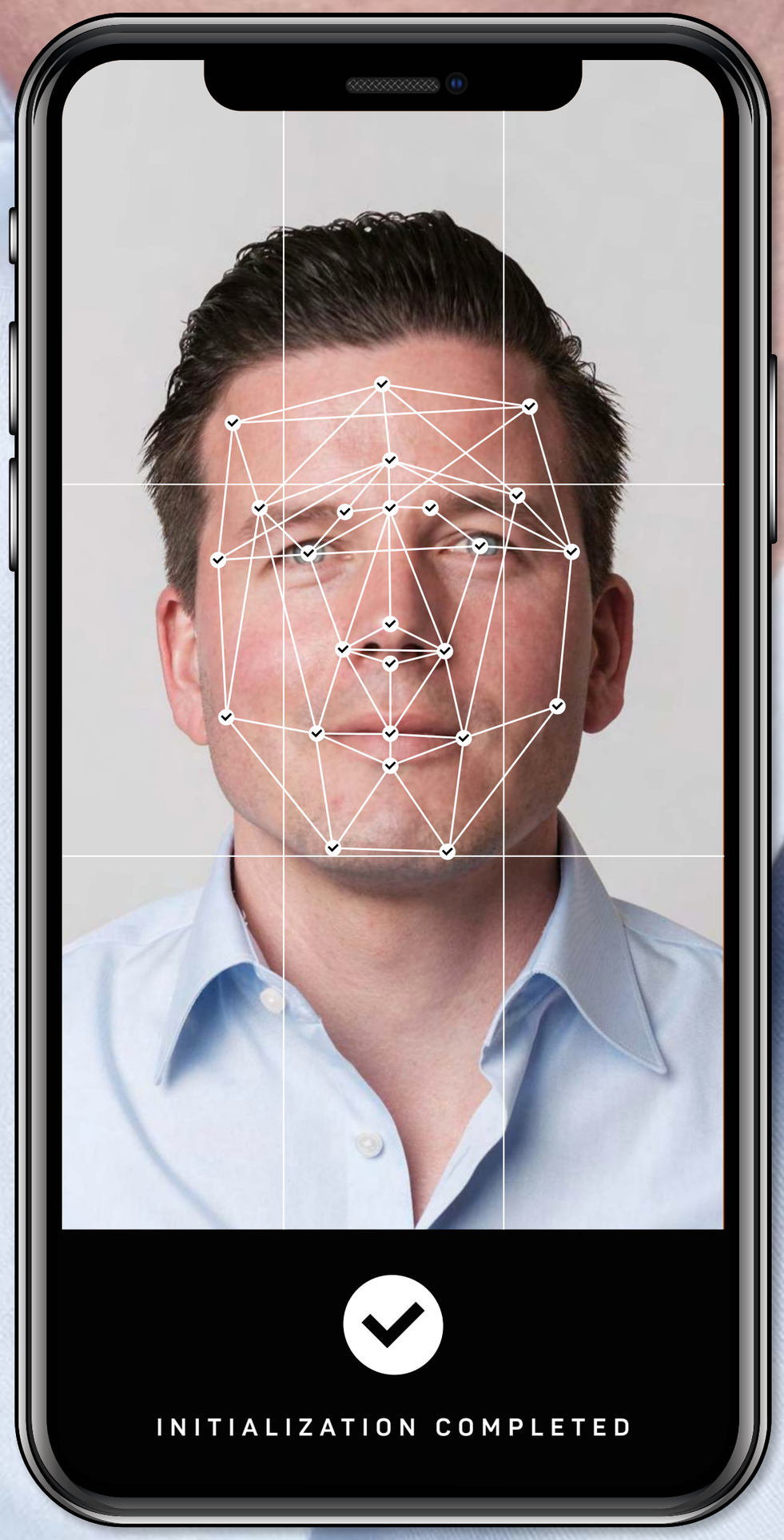
JARNO DUURSMA



DEEPFAKE TECHNOLOGIE

THE INFOCALYPSE

studio OVERMORGEN

DEEPFAKE TECHNOLOGIE

THE INFOCALYPSE

Jarno Duursma

Trendwatchter | Auteur | TEDx spreker

studio OVERMORGEN

INHOUDSOPGAVE

SAMENVATTING

Zoals geldt voor elke technologie kunnen ook door artificiële intelligentie (AI) gedreven generatieve technieken worden misbruikt. Het verschijnsel *deepfake*, ofwel de inzet van *generative AI software* voor het creëren van nepinformatie, omvat dergelijk misbruik. *Deepfake technologie: The Infocalypse* laat zien hoe door AI gegenereerde teksten, beelden, video's en audio ons kunnen misleiden. Wat daarvan de mogelijke bedreigingen zijn en de eventuele oplossingen.

Als deepfake-technologie digitale informatie kan creëren en manipuleren, laat die informatie zich gebruiken voor criminele doeleinden, waaronder chantage, beïnvloeden van verkiezingen, toebrengen van reputatieschade en om de tuin leiden van biometrische herkenning. In potentie kan deepfake-technologie op eenvoudige wijze een tsunami aan nepnieuws opleveren en daarmee het vertrouwen in de journalistiek ondermijnen of een apathie bij de consument opwekken voor nieuwsfeiten. De eenvoudige vraag of informatie echt of nep is, zal in de toekomst prangend worden. Want wanneer wij in onze hyperverbonden wereld onze ogen en oren online niet meer kunnen vertrouwen ontstaat een *infocalypse*: een extreme devaluatie van de betrouwbaarheid van informatie, met alle gevolgen van dien.

1 | INLEIDING

De term deepfake combineert de Engelse woorden *deep* en *fake*. Fake betekent nep, en slaat op nepinformatie die moderne *generatieve AI*-software kan creëren. Deze systemen creëren nieuwe digitale content zoals afbeeldingen, video's, teksten, menselijke stemmen en andere audio opnames. De term *deep* vindt zijn herkomst in de zogenaamde *deep learning*-netwerken, die te maken hebben met de huidige kwalitatieve groeispurt van deze generatieve AI-software.

Een veelbelovende ontwikkeling die de afgelopen jaren in dit domein is gestart, betreft de combinatie van twee netwerken in zogenaamde *generative adversarial networks*, of GAN's. Door gebruik te maken van trainingsmateriaal (tekst, audio, foto, video) probeert een GAN nieuwe content te maken. In dat proces gaan twee netwerken met elkaar in competitie. Het ene netwerk, de generator, probeert nieuwe output te maken die past binnen het trainingsmateriaal, en het tweede netwerk, de discriminator, keurt die nieuwe output goed of fout. Daarmee kunnen de twee systemen elkaar tot grote hoogten stuwen en zo hoge kwaliteit nep content produceren.

Er ontstaat nieuwe output, die lijkt op het gebruikte trainingsmateriaal, maar dat dus niet is. Dit soort systemen genereren nieuwe digitale content die ons bekend voorkomt, maar in feite volledig nieuwgeboren is.
Een wereld van nieuwe afgeleiden, nieuwe invalshoeken, nieuwe ideeën opent zich, met dank aan generatieve AI-software.

Over bovenstaande technologische trend gaat het rapport *Machines met verbeeldingskracht: een kunstmatige realiteit*, waar het rapport dat u op dit moment leest een tekstuele bewerking van is.

In Deepfake technologie: The Infocalypse gaat het vooral over generatieve AI-software die wordt ingezet om onze zintuigen, en dan met name onze ogen en oren, te misleiden door nepinformatie te creëren en verspreiden.

Recent was er bijvoorbeeld een app op basis van generatieve AI-technologie die foto's van (beroemde) mensen kon veranderen in naaktfoto's (Deepnude). Er zijn nu technieken die na het verwijderen van voorwerpen uit foto's, zoals een kei, een paal of een vlag, de foto bewerken zodat die er weer alledaags uitziet. Bij teksten kunnen generatieve AI-systemen, zoals GPT2 van OpenAI, in principe op basis van een voorgestelde krantenkop met een druk op de knop een nepbericht produceren. Het maken van nepnieuws wordt op deze wijze erg gemakkelijk.

Bij video's laat de techniek toe om gezichten te verwisselen (face swap), wellicht grappig op een verjaardagsfeest of in een chatgroep, maar misbruik ligt in het verschiet. Een generatief AI-systeem dat is getraind met video's van David Beckham, kan nieuwe video genereren waarin zijn lippen perfect synchroon tekst uitspreken in een taal die hij absoluut niet beheerst. De stap naar fake is dan al genomen. Er zijn systemen die zijn

getraind met video's van pratende mensen, daardoor het verband kunnen leggen tussen lipbewegingen en spraak, en deze stem vervolgens synthetisch kunnen namaken.
Er zijn al generatieve AI-systemen die het verband kunnen leggen tussen stem en uiterlijke kenmerken en met deze kennis een afbeelding kunnen genereren van iemands gezicht op basis van een stemopname.
U kunt straks wellicht uw stem klonen, waarna uw computer met die stem boeken aan de kinderen voorleest. Een gekloonde stem biedt ook mogelijkheden voor mensen die door ziekte hun spraak zullen verliezen.
De kwaliteit wordt alsmaar beter, en misbruik met frauduleuze stemopnamen ligt voor de hand. Een politicus lijkt dan omstreden uitspraken te doen, of een potentiële investeerder ontvangt een voicemail waarin een directeur hem lijkt te beledigen.

GAN-techniek kan kunstmatige gezichten maken die niet van echte foto's kunnen worden onderscheiden, en zelfs op basis van eenvoudige lijntekeningen realistische foto's genereren. Ook is er een GAN-systeem dat uw foto's omzet zodat die lijken op schilderijen van Vincent van Gogh of Paul Cézanne.

Combinatie van GAN-systemen kan zelfs nog een stap verder. Het StackGan-systeem, bestaande uit twee GAN-systemen, kan een aangeleverde beschrijving (tekst, bijvoorbeeld "Deze vogel is blauw met wit en heeft een korte snavel") begrijpen en daarvan een redelijk gedetailleerde en geloofwaardige afbeelding van genereren.

Machines kunnen dus innovatieve variaties maken op bestaande data, leidend tot nieuwe invalshoeken, nieuwe ideeën en afgeleiden daarvan. Die technologische ontwikkeling, van machines met verbeeldingskracht, kan leiden tot vele doorbraken op medisch, wetenschappelijk en technisch gebied, maar zoals bij iedere technologische ontwikkeling kent ook deze trend een negatieve keerzijde. Daarover gaat Deepfake technologie: The Infocalypse.

Bij deepfake-technologie [2] beïnvloedt generatieve AI software onze waarneming van de wereld op een negatieve manier.
Er zijn tal van manieren waarop deepfake-technologie kan toeslaan om meningen te manipuleren, mensen te chanteren of reputatieschade toe te brengen. We betreden daarmee een tijdperk waarin we online onze ogen en oren niet meer kunnen vertrouwen. [3]

Ik wil hier deze deepfake-technologie ontleden en duiden, omdat ik het mijn plicht vind om vanuit mijn rol als trendwatcher, onderzoeker en duider van alles wat zich afspeelt in de digitale frontlinie, u daarover te berichten. [4|5|6|7]

Ik wens u veel leesplezier. Heeft u vragen, opmerkingen of suggesties? Neem dan gerust contact met mij op. Ook ben ik vanzelfsprekend beschikbaar als spreker over dit onderwerp.

Het rapport Machines met verbeeldingskracht: een kunstmatige realiteit, waaruit deze publicatie afkomstig is, valt te beschouwen als een opvolger van mijn boek uit 2017, De digitale butler – Kansen en bedreigingen van kunstmatige intelligentie. [8] Daarin beschrijf ik hoe AI-systemen steeds meer menselijke vaardigheden van ons overnemen, zoals kijken, luisteren, spreken en lezen. Machines met verbeeldingskracht: een kunstmatige realiteit voegt daar een belangrijke menselijke vaardigheid aan toe: het kunnen toepassen van verbeeldingskracht. Het rapport is onafhankelijk van aard en heeft als doel om te informeren over generatieve AI-software, deepfake-technologie en synthetische media.

Credit: Nick Otto.
https://www.jarnoduursma.nl/wp-content/uploads/2019/03/Jarno-Duursma-Profiel0-def-profile.jpg

Over Jarno Duursma

Jarno Duursma is trendwatcher, futurist en TEDx-spreker. Hij is auteur van vier boeken over digitale technologie, onder andere over kunstmatige intelligentie en blockchaintechnologie. Jarno is vaak te zien en horen in de landelijke media en schrijft opinieartikelen voor onder andere *FD, NRC* en *de Volkskrant*. Hij is eigenaar van Studio Overmorgen en was jarenlang organisator van TechEvent SMC050.

2 | DEEPFAKE EN SYNTHETISCHE MEDIA

Beeldmanipulatie is niet voorbehouden aan ons digitale tijdperk. Er zijn vele voorbeelden in de wereldgeschiedenis waarbij achteraf bleek dat beelden waren gemanipuleerd. Zo liet bijvoorbeeld de Amerikaanse president Lincoln een gravure maken waarin zijn hoofd zich bevond op het lichaam van John C. Calhoun, een vicepresident van de VS in de eerste helft van de negentiende eeuw. Naar verluidt was dat om de uitstraling van president Lincoln meer 'presidentieel' te laten lijken. [9]

Een recenter voorbeeld betreft het gerenommeerde maandblad *National Geographic* dat in februari 1982 op de cover een afbeelding van de piramiden van Gizeh plaatste. Tot schrik van de fotograaf was de foto bewerkt, zodat de piramiden van Gizeh dichter bij elkaar stonden, dat paste mooier. Na de ontdekking van die beeldfraude moest het tijdschrift diep door het stof, de geloofwaardigheid was aangetast en het herstel van die zelf veroorzaakte reputatieschade kostte tijd.

An authentic photo of smoke burning from buildiungs in Beirut suburbs during Israeli air raid.

Manipulated version of photographer

Edited image from source: Week 9: Fake News Images – Writing for the Media. http://courses.dc.edu/mediawriting/week9/

In 1990 kwam Adobe met het programma Photoshop op de markt en sindsdien is 'photoshoppen' een werkwoord. De Libanese fotograaf Adnan Hajj maakte bijvoorbeeld in 2006 dankbaar gebruik van dat programma. Hij manipuleerde toen een foto [10] die hij had gemaakt na een Israëlische luchtmachtaanval op de Libanese hoofdstad Beiroet. Zowel de rook afkomstig uit een gebouw als ook het stedelijke landschap was gemanipuleerd om de situatie erger te doen laten lijken. Persbureau Reuters stopte na deze ontdekking onmiddellijk de samenwerking met de fotograaf.

En nu vrijwel iedereen in het bezit is van een smartphone met ingebouwde goede fotocamera inclusief alle beschikbare fotofilters is het niet zo gek om te stellen dat de kans groter is dat u online een gemanipuleerde foto tegenkomt dan het origineel.

Bewerkte en gemanipuleerde hyper-realistische video's waren tot voor kort voorbehouden aan de Hollywood-studio's met veel expertise en goedgevulde beurzen, maar inmiddels komen ze wereldwijd voor en zijn ze voor iedereen binnen handbereik. De afgelopen maanden verschenen in dat licht steeds vaker krantenartikelen en nieuwsberichten over deepfake-technologie. Veelal wordt er daarbij verwezen naar *face swapping*-technologie, waarbij het gezicht van de ene persoon wordt verwisseld met dat van een ander. Deze gemakkelijkste variant van deepfake-technologie is inmiddels beschikbaar als app op de smartphone. [11]

Edited image from source: Chinese Deepfake App ZAO Sparks Mass Downloads and Major Concerns. https://radiichina.com/china-deepfake-app-zao/

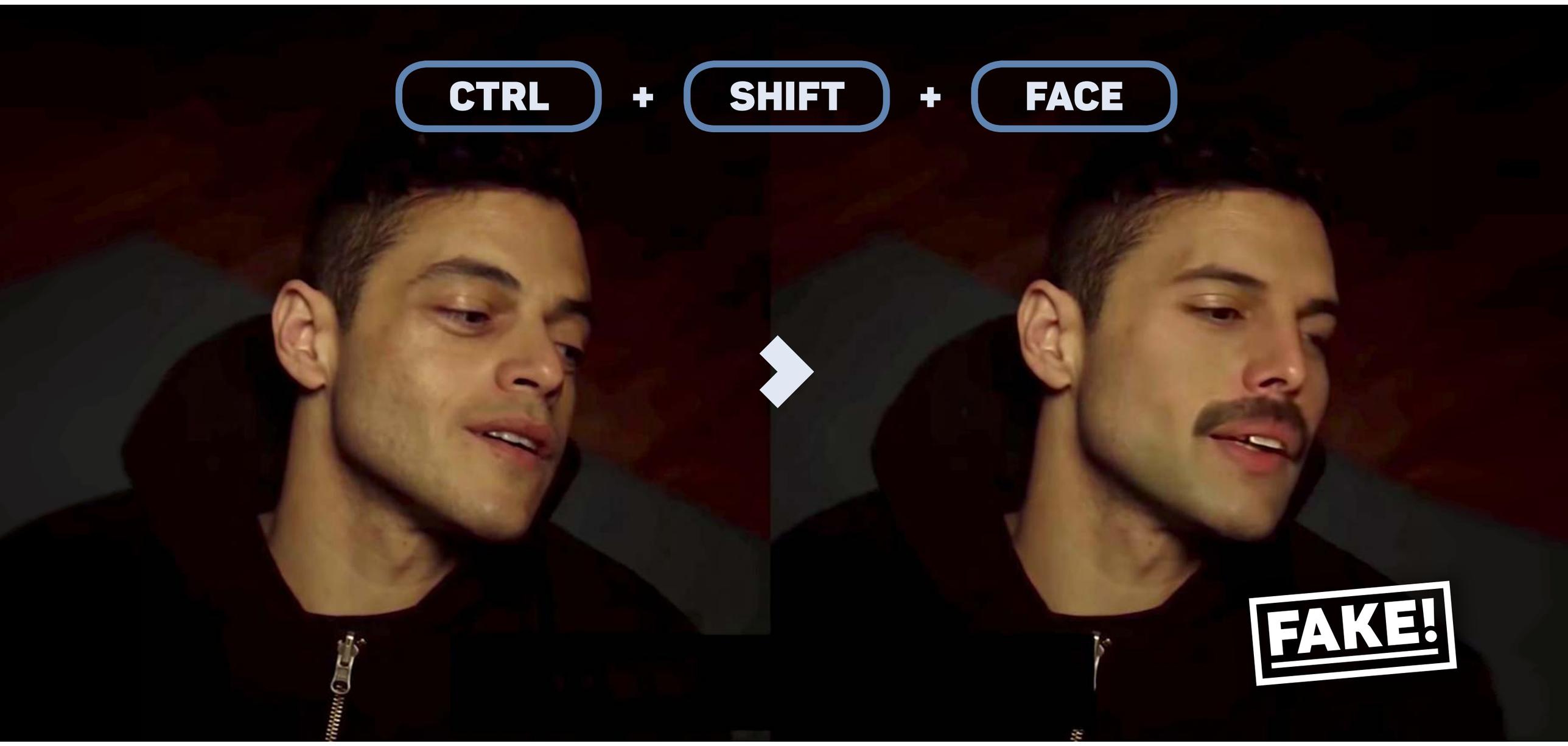
De meer serieuze en gerenommeerde nieuwsbronnen leggen vaak de focus op de mogelijke negatieve gevolgen van de nieuwe deepfake-technologie, bijvoorbeeld voor de journalistiek of bij toepassing voor criminele doeleinden. Veel andere nieuwsbronnen concentreren zich vooral op humoristische filmpjes vol met spot, parodie en satire. Op deze wijze raken steeds meer mensen bekend met deze technologische trend. In het kort is *deepfake*: de inzet van *generative AI technology* voor het creëren van nepinformatie. Deepfake-technologie houdt hier in: generatieve AI-software die *synthetische media* kan creëren zoals bijvoorbeeld gezichten, stemmen, teksten, beelden, bewegende mensen en geluiden. Deepfake gaat uit van negatieve intenties bij het gebruik van deze generative *AI technology*.

Wat voor impact heeft het nu wij steeds minder erop kunnen vertrouwen dat wat we online zien nóg minder een weerspiegeling is van de werkelijkheid? Sommigen hebben het al over een *infocalypse* [12], een vernietiging van de betrouwbaarheid van informatie omdat foto's, video's, geluidsopnamen, menselijke stemmen, geschreven teksten en geschreven recensies die we voortdurend tegenkomen als we onze *devices* raadplegen, allemaal nep kunnen zijn. Hoe gaan we ermee om dat generatieve AI-software steeds beter wordt en dat we steeds minder onze ogen en oren kunnen vertrouwen?

Bij de bespreking van de mogelijke negatieve gevolgen van generatieve AI-software in de vorm van deepfakes, zal ik helder aangeven wat er op dit moment reeds mogelijk is en speculeer ik voorzichtig over de mogelijkheden van de toekomst.

3 | SOORTEN DEEPFAKES-VIDEO'S

Wanneer we het hebben over deepfake-video's, wordt zoals gezegd meestal *face swapping*-technologie bedoeld. Er zijn echter meer varianten.



Edited image from source: Freddie Mercury DeepFake Rami Malek [VFX Breakdown] - YouTube.
https://www.youtube.com/watch?v=iwvF9orOnWI

Face swap / Facial replacement technology

Zoals aangegeven is *face swapping* het bekendste voorbeeld van deepfake-technologie: het gezicht van een persoon wordt verwisseld met dat van een ander. De bewegingen van het gezicht worden daarbij een-op-een overgenomen. Kunstmatig intelligente generatieve software is inmiddels zover gevorderd dat dit relatief eenvoudig is. De eerste varianten van deepfake-video's betroffen de gezichten van beroemdheden die op de lichamen van porno-actrices waren geplakt [13]. Dat zij het slachtoffer zijn, is vanuit technisch oogpunt logisch: van beroemdheden en politici is er immers redelijk veel foto- en videomateriaal aanwezig online. Dat dient dan als het benodigde trainingsmateriaal waarmee een generatief AI-systeem het gezicht kan leren na te bootsen.

READ MY LIPS

Voor het zaaien van verwarring, hoeft niet eens het gehele gezicht gemanipuleerd te worden. Het vervangen van de mond volstaat. Het met moderne digitale technieken kunstmatig genereren van de mond creëert vaak een geloofwaardige nepvideo.
Een voorbeeld daarvan leveren de hier weergegeven afbeeldingen uit een video van Obama. [14] Die video is mede zo realistisch omdat het kunstmatig intelligente systeem slechts de mond hoefde te vervangen.
De stem van Obama was overigens niet synthetisch gecreëerd, maar werd ingesproken door Jordan Peele, een Amerikaanse filmmaker in komedie- en horrorgenres.

Edited image from source: Fake Obama created using AI video tool - BBC News. https://www.bbc.com/news/av/technology-40598465/fake-obama-created-using-ai-tool-to-make-phoney-speeches

Ook in een inmiddels van internet verwijderde video van Kim Kardashian, waarbij zij zogenaamd toegaf haar volgers te manipuleren in ruil voor geld, maakte gebruik van deze *lipsync* deepfake-technologie. [15] Mede daardoor is het resultaat verbluffend goed.

Edited image from source: Bill Posters (@bill_posters_uk) • Instagram-foto's en -video's https://www.instagram.com/p/ByKg-uKIP4C/

Digital Puppetry/ Do-as-I-Do technology

Digital Puppetry is een variant van deepfake-video's waarbij een volledig kunstmatig hoofd of lijf wordt gegenereerd. Deze digitale 'marionet' *(puppet)* gegenereerd door een generatief AI-softwaresysteem kan daarbij worden aangestuurd door een externe bron. Die externe bron beweegt dan bijvoorbeeld het hoofd naar rechts en de digitale marionet doet hetzelfde. Ook gezichtsuitdrukkingen, bewegingen van lippen (om gesproken zinnen te simuleren) en zelfs bewegingen van het hele lijf kunnen worden nagebootst. Zoals het gezicht of het lichaam zich beweegt, zo beweegt de *digital puppet*. Deze volledig synthetische mensen zijn dus een digitaal beeld van wat anderen doen.

Misschien heeft u ooit de video voorbij zien komen van "Everybody Dance Now" [16], waarbij de synthetische versie van een wetenschapper werd aangestuurd door een professioneel danser. Op deze wijze leek zij zich geloofwaardig soepel voort te bewegen, zonder dat ze ooit dansles had genomen. Haar synthetische digitale beeld werd aangestuurd door een externe bron, een professionele danser.

Source video **Detected pose** **Source to target result**

PERSONALIZED AVATAR CREATION

Onderzoekers van de Universiteit van Heidelberg is het gelukt om *fullbody deepfakes* te maken. [17] Ze hebben hun systeem geleerd om het volledige lichaam van een persoon te plakken over videobeelden van een bestaand persoon. Stelt u zich voor: u wilt een video genereren waarin u overtuigend tennis speelt, maar u bent gewoon heel slecht in deze sport. Deze *personalized avatar creation*-technologie, hoewel zeker nog niet perfect, maakt het mogelijk om uw lichaam te plakken over dat van een bestaand persoon, bijvoorbeeld Rafael Nadal. Interessant aan dit onderzoek is dat het kunstmatig intelligente systeem geen volledige driedimensionale weergave als 'bronbestand' van uw volledige lichaam hoeft te hebben. Het kan als het ware 'raden' hoe u er van de zijkant uitziet en u zo laten meebewegen met het bronobject in de video. Zo kunt u toch een video maken waarin u als tennisprofessional te zien bent.

Cheapfake / ShallowFake

Voor de bredere context van dit rapport is het goed om nog een categorie video's te benoemen. Dat betreft de zogenaamde *shallowfake*-video [18], ook wel met *cheapfake* aangeduid. Dat is een video die weliswaar bewerkt of gemanipuleerd is, maar waarbij geen kunstmatig intelligent systeem heeft geholpen om het resultaat te creëren. Wanneer u die beschouwt door de koker van generatieve AI-technologie, dan hoort de video officieel niet in dit rapport thuis.

Edited image from source:
CREDO Slams Facebook for Not Pulling Pelosi Video - Broadcasting & Cable.
https://www.broadcastingcable.com/news/credo-slams-facebook-for-not-pulling-pelosi-video

Vaak zijn dit soort *shallowfake*-video's handmatig vertraagd, geknipt of gefilterd. Het bekendste voorbeeld uit de recente geschiedenis is die van spreker Nancy Pelosi, waarbij de videobeelden waren vertraagd zodat het leek alsof zij dronken was. [19] De video had vele miljoenen *views* [20]. Pelosi heeft Facebook nog verzocht de video te verwijderen, maar dat werd geweigerd door het socialemediaplatform [21]. Op een later moment gaf Facebook-oprichter Mark Zuckerberg [22] echter wel toe op dit vlak een inschattingsfout te hebben gemaakt. Of deze opmerking te maken heeft gehad met het feit dat Zuckerberg zelf in de tussentijd ook slachtoffer werd van een deepfake-video, is onbekend. [23] Hoewel bij deze *shallowfake*-video geen kunstmatig intelligent systeem is gebruikt dat nieuwe beelden creëert, kan die video wel degelijk een schadelijk gevolg hebben. Met simpele foto- of videosoftware kunnen beelden worden vertraagd, ingekort of verknipt en daarmee de waarheid geweld aandoen. Doordat video- en fotobewerkingsprogramma's gemakkelijker te bedienen zijn, komen gemanipuleerde beelden in het algemeen dus ook vaker voor.

Edited image from source:
Bill Posters Instagram: "'Imagine this...' (2019) Mark Zuckerberg. https://www.instagram.com/p/ByaVigGFP2U/

A Perfect Storm

Waarom horen we juist de afgelopen tijd zoveel over deze deepfake-video's? Hoe kan het dat deze ontwikkeling zo snel is gegaan? Waarom neemt de hoeveelheid deepfake-video's zo enorm toe? Het antwoord is relatief eenvoudig: alle seinen staan op groen om deze ontwikkeling bovenmatig te versnellen. De video's zijn eenvoudig te maken, gemakkelijk te distribueren en er is meer dan voldoende publiek. Trendwatcher Sander Duivestein is hierover heel duidelijk: "Het maken van een nepvideo wordt met deepfakes net zo gemakkelijk als het vertellen van een leugen" [24] Zoals er ideale weersomstandigheden kunnen zijn voor het creëren van een perfecte storm, zo is dat ook met de ontwikkeling van deze technologie.

Eenvoudige productie

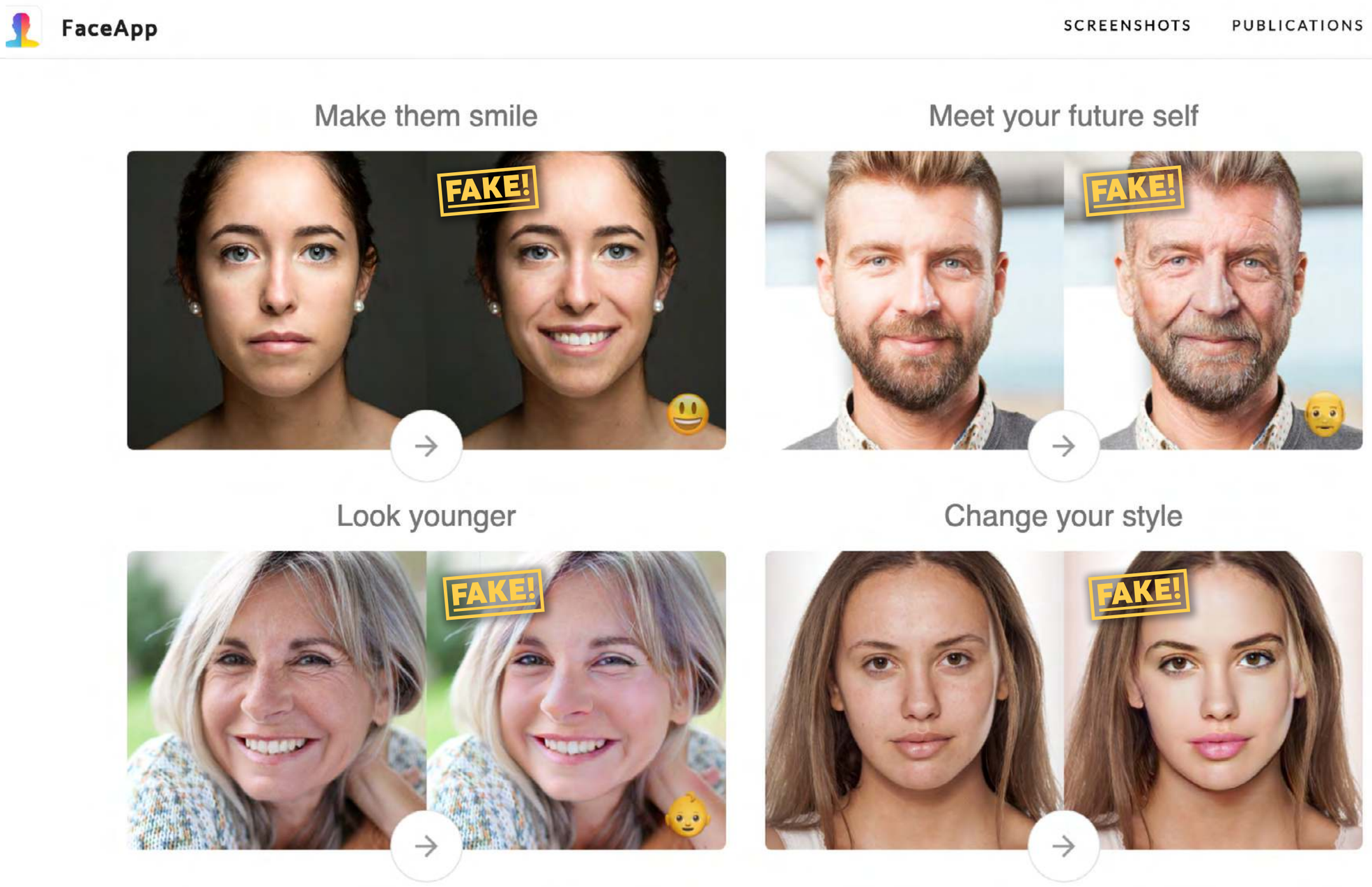
Deepfake-video's zijn, in vergelijking meteen paar jaar geleden, relatief eenvoudig om te maken en dat gemak gaat in de toekomst alleen maar toenemen. [25] Op het internet kunt u moeiteloos een handleiding vinden voor het maken van een deepfake-video en u heeft daarvoor tegenwoordig al geen supercomputer meer nodig. Een gamingcomputer van € 1000 volstaat, maar u kunt er waarschijnlijk zelfs in de *cloud* computerruimte voor huren. De bijbehorende handleidingen zijn meestal gedetailleerd en u heeft daarom nauwelijks programmeerkennis nodig. En wanneer u er niet uitkomt, zijn er vele fora waar u antwoorden op 'veel gestelde vragen' over deepfake kunt teruglezen of ze zelf kunt stellen.

Het enige wat u verder nog nodig heeft, is een aantal foto's of video's van een beroemdheid, politicus, journalist, ex-vriendin of buurman en u kunt aan de slag. Vanzelfsprekend geldt dat hoe meer foto's u van een persoon kunt verzamelen, des te hoger de kwaliteit van de video zal zijn. Deze haalbare kwaliteit zal in de komende maanden en jaren alleen maar toenemen.

Het is van belang om hierbij twee zaken te onderscheiden. Aan de ene kant is het namelijk eenvoudig om een deepfake-video te maken, en aan de andere kant is het moeilijk. [26] De gemakkelijkste variant van een deepfake-video is bijvoorbeeld *face swapping*-technologie met uzelf in de hoofdrol. U houdt uw eigen gezicht voor een camera en de software leert hoe uw gezicht eruitziet. Vervolgens kunt u binnen een app uw gezicht plakken op het gezicht van bijvoorbeeld beroemdheden. U kunt dit beschouwen als een geavanceerde vorm van de Snapchat-filters. Formeel heeft u dan een deepfake-video gemaakt, maar die kan vaak niet veel schade doen. U bent zelf de bron, u heeft zelf de regie en de reikwijdte van dit soort applicaties komt vaak niet verder dan beroemdheden of grappige verkleedpartijen. Aan de andere kant wordt het al veel moeilijker wanneer u een derde persoon iets wilt laten doen of zeggen wat hij of zij niet gedaan of gezegd heeft. Om een dergelijke video geloofwaardig te maken; daarvoor is nog wel wat tijd en moeite voor nodig.

Voetnoot: ook video's van relatief slechte kwaliteit kunnen overigens een grote impact hebben. Denk daarbij aan wraakporno gericht op een ex-vriendin.

" Het maken van een nepvideo wordt met deepfakes net zo gemakkelijk als het vertellen van een leugen. "



Credit: FaceApp - Free Neural Face Transformation Filters.
https://www.faceapp.com/

EEN ANDER GEZICHT

Het internet biedt inmiddels ook steeds vaker kant-en-klare apps voor bijvoorbeeld *face swapping*, zoals FaceApp. [27] Dat is een app waarmee u uw selfie kunt aanpassen. U kunt uzelf ouder laten lijken, of juist jonger. U kunt uzelf een gepolijst fotomodellengezicht geven óf een foto van uzelf met een neutrale blik laten veranderen in een glimlachend zelfportret. Ook kunt u make-up of tatoeages laten toevoegen. Er is overigens meerdere keren wat controverse geweest rondom deze app. Zo bleek dat die het eigendom was van een Russisch bedrijf. Amerikaanse politici vroegen zich openbaar af welke data werden verzameld door Rusland en of mensen zich er wel bewust van waren dat dit gebeurde. [28] Deze gevoeligheid komt vanzelfsprekend voort uit de beïnvloeding door Rusland van de Amerikaanse presidentsverkiezingen van 2016. [29] FaceApp verzekerde echter dat de data werden verwerkt in de *cloud* in de VS, bij Google en Amazon.

In 2017 was er ook al controverse rondom de app omdat het de huidskleur van donkere mensen lichter kleurde om ze “fysiek meer aantrekkelijk te maken”. Eenzelfde soort kritiek heeft FaceApp al eerder gekregen toen filters iemand een bepaalde etniciteit konden geven, wat volgens critici racistische stereotypering in de hand werkte. [30]

Distributie

De distributie van deepfake-video's is erg eenvoudig via sociale media, smartphone-apps, fora, chat-apps, noem maar op. Binnen een paar uur kan een video miljoenen keren bekeken worden. In onze hyperverbonden samenleving verspreidt nieuws zich razendsnel en in grote volumes. Bovendien is het voor socialemediaplatformen als Facebook en YouTube helaas niet altijd gemakkelijk een deepfake-video te herkennen.

Publiek

Een andere factor die bijdraagt aan de perfecte storm van deepfake-video's is het publieke klimaat. De echokamer van sociale media zorgt voor een filterbubbel waarin mensen hun eigen wereldbeeld en bijbehorende vooroordelen graag bevestigd willen zien. Nieuws is steeds meer emotie geworden, zo lijkt. Wij als consumenten van audiovisuele media klikken immers graag op (en delen graag) sappig, negatief, prikkelend, vooroordeel bevestigend nieuws met de buitenwereld. Na één swipe of druk op de knop verspreiden de video's zich, dwars door alle platformen heen. Informatie die het (terecht of ten onrechte) bestaande wereldbeeld bevestigt, wordt sneller gedeeld, zeker wanneer die nieuw en negatief van aard is. [31]

Ook zien we dat de traditionele media continu worden aangevallen door de Amerikaanse president [32] *("fake media")*, maar ook bepaalde politieke partijen in Nederland stellen de betrouwbaarheid van deze media steeds ter discussie. Mede als gevolg daarvan accepteren veel consumenten soms liever 'nieuwsfeiten' van minder bekende bronnen, dan van de gebruikelijke.

" More than anything else, the dynamics that define the web — frictionless sharing and the monetization of attention — mean that deepfakes will always find an audience [33]. "
- James Vincent, techjournalist The Verge

4 | DE KRACHT VAN DEEPFAKE

Wanneer iedereen ter wereld met gemak een video kan maken waarin andere mensen ongevraagd en tegen hun zin de hoofdrol vertolken, heeft dat nogal wat gevolgen. Deze technologische ontwikkeling doet een dringend appel op bijvoorbeeld overheid, politie en nieuwsorganisaties. Zij krijgen te maken met een steeds hogere kwaliteit digitaal gecreëerde manipulatieve of provocatieve video's. Dat heeft allerlei concrete nadelige gevolgen. Wat het probleem versterkt en in de toekomst wellicht onbeheersbaar maakt, zijn de snelheid en de schaalbaarheid van deze technologie.

Snelheid

Allereerst is er het probleem van de snelheid. Video's kunnen bijzonder snel gecreëerd worden en snel worden verspreid: ze kunnen binnen een uur al miljoenen keren bekeken zijn. Hoe gaat de overheid, politie of nieuwsorganisatie deze video's op tijd vinden? Hoeveel tijd is er vervolgens om een video te ontkrachten en informatie te rectificeren? Is daar de juiste technologie voor aanwezig? Zijn medewerkers bekwaam genoeg om feit van fictie te onderscheiden?

Schaalbaarheid

Behalve de snelle productie en snelle verspreiding van deepfake-video's is de schaalbaarheid ook een potentieel probleem. Wanneer de benodigde software wereldwijd even gemakkelijk te gebruiken is als bij wijze van spreken het internetbankieren, zal ook de kwantiteit van deepfake-video's gaan toenemen. Voor overheid, politie en nieuwsorganisaties is dat een enorme kluif. Zij zullen met hun beperkte middelen en kennis waarschijnlijk keuzes moeten maken. Daarbij kunnen ze natuurlijk geholpen worden door technologie, maar het is de vraag of dat voldoende zal zijn.

Het probleem voor bovengenoemde organisaties wordt trouwens nog iets complexer: veel deepfake-video's zullen uiteindelijk nooit het mainstream publiek bereiken, maar altijd in de niches van bepaalde groeperingen blijven hangen. Denk daarbij aan kringen van religieus extremisme of anti-establishment-ideologie. Video's die in die kringen circuleren, zijn bedoeld om het eigen wereldbeeld te bevestigen, maar zullen via traditionele organisaties nooit ontkracht kunnen worden. Zij bevinden zich in een onzichtbare laag van het internet, binnen een afzonderlijke internetbel. De video's hebben daardoor geen grootschalig, maar wel een langdurig effect. Ze kunnen er bijvoorbeeld voor zorgen dat onjuiste overtuigingen in stand blijven of worden versterkt, wat radicaal gedrag in de hand werkt. Daarover gaat ook het interview met trendwatcher Sander Duivestein.

INTERVIEW: SANDER DUIVESTEIN

TRENDWATCHER EN ONDERZOEKER BIJ HET VERKENNINGSINSTITUUT NIEUWE TECHNOLOGIE VAN SOGETI.

Hoe kijk jij vanuit jouw rol naar GAN- en deepfake-technologie?
Ik volg deepfake- en GAN-technologie al een tijdje en ik vind het een interessant fenomeen. In de basis ben ik optimistisch, maar vanzelfsprekend zie ik ook de negatieve mogelijkheden om deze technologie te gebruiken voor misleiding en bedrog. Ik maak mij zorgen dat binnen bepaalde extremistische doelgroepen deepfake-video's of deepfake-stemopnamen kunnen aanzetten tot daadwerkelijke gewelddadige acties in de fysieke wereld. Dat er daadwerkelijk een gek tussen zit die op basis van een nepvideo actie gaat ondernemen.
Je zag de kracht van dit soort manipulatie al bijvoorbeeld bij pizzagate. Tegenstanders van de presidentiële campagne van Clinton in 2016 verspreidden toen de samenzweringstheorie dat de democraten zich bezig hielden met kindermisbruik.
En dat een restaurant in Washington het middelpunt daarvan was. Een mafkees reisde vervolgens naar het restaurant om de samenzwering te onderzoeken en schoot aldaar met een geweer. Bizar. Maar vergis je niet: er zijn al heel veel fora op het internet waar het wemelt van de complottheorieën. Wanneer je als kwaadwillende vanuit je luie stoel dus allerlei nepvideo's en nepopnamen kunt maken om zo extra kracht te zetten in je manipulatie, zul je zien dat dit werkt. Het zal gretig aftrek vinden als versteviging van verschillende complottheorieën.

Maar je bent ook positief over de technologie; leg eens uit.
Klopt; ik zie ook wel heel duidelijk de positieve mogelijkheden van GAN- en deepfake-technologie. Wat die laatste betreft; mensen worden bijvoorbeeld nog meer genoodzaakt om kritisch te zijn richting wat ze zien. En journalisten moeten meer hun best gaan doen. Een goede ontwikkeling.

Maar ik zie deepfake-technologie maar als een klein onderdeel van een groter geheel. Wanneer je kijkt naar de onderliggende GAN technologie, zal dit volgens mij een explosie van creativiteit veroorzaken binnen nu en een paar jaar. Machines komen met nieuwe opties. Met deze GAN-technologie kan bijvoorbeeld iedereen zijn eigen Hollywoodfilm maken. Je kunt je eigen 3Dpersoonlijke avatar ontwerpen en laten meespelen in video's. Je fysieke digitale tweeling kun je laten meespelen in een film waarvan je zelf de regisseur bent.
Omdat GAN-technologie onze creativiteit kan versnellen, kijk ik heel positief tegen dit fenomeen aan. Ik zie bijvoorbeeld dat er gemakkelijker nieuwe medicijnen bedacht kunnen worden. De software denkt namelijk met ons mee. Of dat je volledig nieuwe werelden kunt creëren in digitale simulaties. Denk aan het project van NVIDIA, waar een volledige stad wordt gecreëerd in 3D door een machine. Zo'n wereld zou vervolgens als lesmateriaal kunnen dienen voor de zelfrijdende auto.

Hoe ontvouwt GAN-technologie zich in de toekomst?

De scheidslijn tussen echt en nep wordt steeds dunner. Ik stel me zo voor dat je over een paar jaar een selfie kunt maken waarbij je de omliggende wereld in het geheel kunt aanpassen. Dat je bijvoorbeeld zogenaamd op een idyllisch eiland bent in de nabijheid van allerlei Hollywoodsterren. Ik geloof zelfs dat je digitale weergave op de lange termijn belangrijker wordt dan je fysieke representatie in de echte wereld. Het is voor heel veel mensen aanlokkelijk om bezig te zijn met een wereld die ze naar eigen wens kunnen aanpassen. Dat ze precies controle hebben over hoe ze in de digitale wereld worden gezien. De virtuele wereld, die wij bekijken via ons beeldscherm of smartphone, krijgt steeds meer aandacht. Deze wereld kan ook steeds beter worden gemanipuleerd. En zeker wanneer dit steeds gemakkelijker wordt om te doen, met allerlei apps, zal de populariteit ervan stijgen.

Ik stel mijzelf voordat je in de nabije toekomst een app op je telefoon hebt waarmee je een video maakt van jezelf en dat een GAN-systeem jou vervolgens aankleedt met allerlei soorten kleding. De software wordt dan een creatieve machine die met allerlei opties komt, ook opties waar jezelf misschien nooit aan had gedacht.
Het is natuurlijk lastig om gedetailleerd te zien hoe het er over tien jaar uitziet, maar dat GAN-technologie en nieuwe golf van creativiteit gaat brengen, dat is voor mij helder.

5 | BEDREIGINGEN

Welke problemen kunnen er ontstaan nu deepfake-video's relatief gemakkelijk en snel kunnen worden gecreëerd en verspreid? Hieronder een korte, niet volledige opsomming.

Credit: This tool could help detect doctored videos of world leaders - CNN.
https://edition.cnn.com/2019/06/12/tech/deepfake-2020-detection/index.html

Onrust en polarisatie

Stel dat er een deepfake-video opduikt van een belangrijke Nederlandse politicus die omgekocht lijkt te worden. Of zo'n video waarin een FBI-medewerker vertelt dat die graag iemand uit de Trump-familie wil oppakken voor vermeende banden met Rusland en dat hij daarvoor zelf bewijs aan het genereren is. Of Russische nepvideo's met daarin een in scène gezet opstootje tussen Amerikaanse politici, een anti-Amerika-demonstratie in Saoedi-Arabië of Amerikaanse militairen die een koran verbranden. Er zijn tal van voorbeelden te bedenken waarbij geënsceneerde video's of geluidsopnamen gegarandeerd leiden tot geopolitieke onrust of sociale polarisatie in de samenleving.

Het kan een serieuze strategie zijn van het ene land om in een ander, vijandig land verdeeldheid te zaaien. Door te polariseren ontstaat steeds minder saamhorigheid en daardoor verzwakking omdat de besluitvorming minder effectief wordt. Het fundament van een democratische staat is immers een gedeelde perceptie van de werkelijkheid en een bijbehorende overeenstemming over feitelijkheden. Wanneer dat ontbreekt, kunnen nationale problemen ontstaan die zich naderhand zeer lastig laten oplossen.

Ook kunnen regimes deepfake-technologie gaan inzetten voor hun eigen propaganda, zowel om politieke tegenstanders zwart te maken als om de eigen politici en leiders op een positieve manier af te schilderen. Nepvideo's laten dan bijvoorbeeld zien hoe die leiders aanwezig waren in penibele situaties en zich gedroegen als ware helden. Internationaal gezien lijken overigens vooral Iran, China, Noord-Korea, de VS en Rusland erg actief te zijn in de ontwikkeling van deze deepfake-videotechnologie.

Chantage

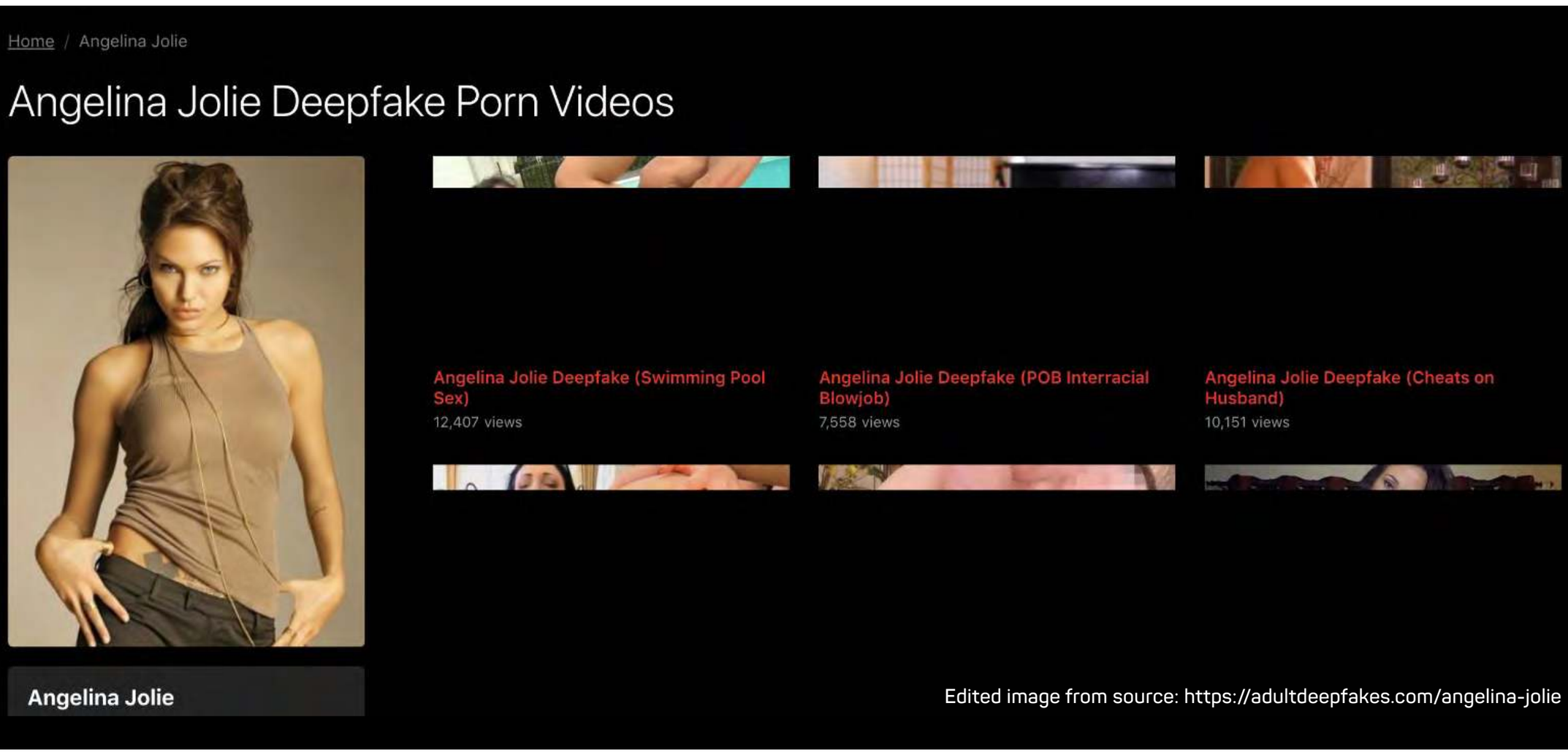
Het is niet ondenkbaar dat politici, journalisten, buitenlandse militairen, directeuren van grote bedrijven [34], klokkenluiders en financieel verantwoordelijken in de toekomst te maken krijgen met chantage met deepfake-video's. Op een dag kunnen ze dan onderstaande e-mail in hun inbox aantreffen:

"Hallo, dit is een link naar een online video waarin jij de hoofdrol speelt. Je vindt het vast onprettig wanneer jouw seksuele escapades te zien zijn voor de buitenwereld. Wat zouden je familie en je vrienden ervan vinden? Dit is vast niet goed voor je reputatie. Ik denk dat het goed is dat je (vul hier de wens van de crimineel in) dus ik ga er vanuit dat je aan onze wensen tegemoet komt."

Zelfs wanneer de deepfake-video een matige kwaliteit heeft, wil de hoofdrolspeler vanzelfsprekend niet dat die wordt verspreid. Alleen al de suggestie van onethisch, crimineel, afwijkend seksueel gedrag kan flink wat reputatieschade en -schande tot gevolg hebben. Om je vervolgens van alle blaam te zuiveren, kost vervolgens enorm veel tijd en energie, want suggesties zijn hardnekkig en kunnen iemand jarenlang achtervolgen. Buitenstaanders denken immers, dat waar rook is, ook vuur moet zijn. Met deepfake-technologie hebben kwaadwilligen een erg krachtig chantagemiddel in handen.

Reputatieschade

Een van de meest vanzelfsprekende effecten van deepfake-technologie is het toebrengen van reputatieschade. Activistische milieu-groeperingen zouden daarmee directeuren van biotechnologiebedrijven in een kwaad daglicht kunnen stellen. Commerciële bedrijven zouden met de technologie een concurrent ten val kunnen brengen. Op de avond vóór een beursgang kan een filmpje van een financieel directeur opduiken, waarin hij zogenaamd toegeeft dat er veel minder liquide middelen op de balans staan dan de officiële papieren vermelden. Aan de vooravond van een verkiezing kan een filmpje opduiken waarin een politicus seksistische, racistische of agressieve taal uitslaat. Als er al eerherstel komt, is dat te laat om de eventuele electorale schade te repareren.



Edited image from source: https://adultdeepfakes.com/angelina-jolie

Zeker mensen voor wie reputatie erg belangrijk is, zullen op korte termijn de gevaren en oplossingen van deze technologie onder ogen moeten zien. Overigens manifesteerde zich in het begin van de ontwikkeling van deepfake-technologie op het Reddit-forum al een zeer venijnige vorm van reputatieschade: deepfakeporno. Met die toepassing werd het gezicht van vrouwen geplakt op dat van een porno actrice. Reddit greep in en verwijderde dit *'non-consensual pornography'*-onderdeel [35], maar dat betekent niet dat de toepassing niet meer bestaat. Wanneer het gebruik van dergelijke deepfake-technologie zich verder verspreidt, zullen meer en meer vrouwen slachtoffer worden. Er zijn al websites waarop vrouwelijke beroemdheden in een pornografische setting worden gemanipuleerd, zoals de afbeelding hierboven [36] laat zien. In de recente Nederlandse geschiedenis werd ex-NOS-nieuwslezeres Dionne Stax slachtoffer van een gemanipuleerde pornografische video [37].

Een ander voorbeeld betreft de Indiase journaliste Rana Ayyub. Nadat ze campagne voerde voor een verkrachtingsslachtoffer, verscheen ze als hoofdrolspeelster in een pornografische deepfake-nepvideo. Die werd vele tienduizenden keer gedeeld en dat had een forse impact, zowel op haar professionele functioneren als op haar als mens. De video was bedoeld om haar reputatie-schade [38] | [39] toe te brengen vanwege haar persoonlijke overtuigingen.

Original photo

Deepnude version

Edited image from source: DeepNude, the Software That 'Undresses' Women, Is Up for Grabs for Starting Price of $30,000 - Sputnik International
https://sputniknews.com/science/201907201076304335-deepnude-online-auction-starting-price/

Deepnude

In de zomer van 2019 kwam de software Deepnude [40] online, aangeprezen als een *AI Powered X-Ray App* waarmee het mogelijk was om foto's van geklede vrouwen te veranderen naar naaktfoto's [41]. Het systeem kon op basis van vele trainingsvoorbeelden als het ware 'raden' hoe een lichaam, meestal van een vrouwelijke beroemdheid, er uitzag en dat beeld vervolgens creëren. Met die software kon relatief gemakkelijk elke foto van een vrouw transformeren naar een blootfoto. De software riep wereldwijd veel protest en weerstand op. Inmiddels is de software offline gehaald, maar duikt deze in verschillende vormen [42] online weer op.

Liars' dividend/ Apathie

Het is helder waar zich de risico's van de deepfake-technologie ongeveer bevinden. En omdat deze technologie steeds gemakkelijker in gebruik wordt, zal het algemene publiek steeds meer gewend raken aan gemanipuleerde beelden, teksten en stemmen. Aan de ene kant is dat goed: wanneer u een geluidsopname ontvangt van een bekende die u vraagt om geld over te maken, is het goed als u beseft dat die stem nep kan zijn. De gewenning aan deepfake-technologie kan echter ook een keerzijde hebben. Die keerzijde kan zijn dat we, door de golf van nepvideo's, nepartikelen en nep geluidsopnamen, kwaadwillenden ongewild een troef in handen geven. Vermeende daders kunnen dan bewijsmateriaal, verzameld door journalisten, burgers of onderzoeksbureaus, gewoonweg afdoen als deepfake-video. Het fenomeen dat iedere belastende informatie kan worden geframed als een synthetisch gecreëerde nepvideo of -geluidsopname staat onder meer bekend als het *liars'* dividend.

Credit: Donald Trump and Vladimir Putin: Helsinki Summit and the Stakes for NATO. https://theglobepost.com/2018/07/15/trump-putin-summit-nato/https://sputniknews.com/science/201907201076304335-deepnude-online-auction-starting-price/

Ik kan daar een voorbeeld van schetsen. Veronderstel dat er een echte geluidsopname opduikt van de privé-ontmoeting in 2018 in de Finse hoofdstad Helsinki tussen de Amerikaanse president Trump en zijn Russische collega Poetin.
Deze ontmoeting [43] vond plaats achter gesloten deuren, zonder assistenten of notulisten. Als uit deze geluidsopname blijkt dat de Amerikaanse president chantabel is, dan kan hij nu anno 2019 deze opname afdoen als deepfake-technologie. Niets aan de hand.

De tweede keerzijde van een mogelijke golf van nepvideo's, nepartikelen en nepgeluidopnamen is dat we als samenleving apathisch worden voor nieuws. Dat er zo veel mogelijke leugens zijn, dat het publiek zijn schouders ophaalt voor elke video of geluidsopname die iets onthult, ook als die op waargebeurde feiten berust.

Zodra er apathie optreedt jegens alle content, verliest de journalistiek haar belangrijke rol als luis in de pels van het bedrijfsleven en de overheid. Het meest krachtige journalistieke wapen, het blootstellen aan het daglicht, verliest dan zijn kracht. Dat zou de democratie [44] en het collectieve morele kompas ernstig kunnen bedreigen.
Over deepfake, generative AI software en criminaliteit spreek ik met Mark Wiebes Innovatiemanager bij Nationale Politie.

INTERVIEW: MARK WIEBES

INNOVATIEMANAGER BIJ NATIONALE POLITIE. OP PERSOONLIJKE TITEL

Hoe kijk jij vanuit jouw rol aan tegen deepfake-technologie?
Ik ben innovatiemanager bij de politie en vanuit die hoedanigheid volg ik technologische ontwikkelingen. Dus ook deepfake-technologie, in al haar verschijningsvormen. Voor de politie is het natuurlijk erg belangrijk om echt van nep te kunnen onderscheiden. Dat is niet iets van 2019, dat is al veel langer zo.

Kun je daar een voorbeeld van noemen? We vragen ons bijvoorbeeld bij de politie natuurlijk heel vaak af wanneer we op een plaats delict komen: wat zien we hier? Wat is echt en wat is nep?
Een aspect dat hierin de laatste tijd concreet mijn aandacht heeft, is bijvoorbeeld het onderzoek naar DNA op een plaats delict. Wat je je zou kunnen voorstellen is dat er, om verwarring te zaaien, DNA van meerdere mensen op die locatie is verspreid. De forensisch rechercheurs moeten dus kunnen vaststellen welk DNA daar op welke manier is gekomen. En dus ook of iets er moedwillig door anderen is geplaatst. Voor vingerafdrukken zou dat ook kunnen gelden. Als iemand een 'stempel' zou kunnen maken van de vingerafdruk van een ander, moeten wij de nepvingerafdruk wel van een 'echte' afdruk kunnen onderscheiden.

Waar komt deepfake-technologie om de hoek kijken?
Het is voor ons belangrijk om deze technologische ontwikkeling van 'fake-technologie' te volgen. Je kunt je voorstellen dat er in de toekomst actief gefraudeerd wordt met een nep-stem van een bestuurder, financieel verantwoordelijke of directeur. Nu gebeurt dat, in de zogenaamde CEO-fraude, soms met mailtjes, maar als dat met een voicemailberichten zou gaan, is dat misschien nóg overtuigender. Wij hebben goeie forensisch rechercheurs die gespecialiseerd zijn in audio-analyse, dus ik verwacht dat ze nu wel in staat zijn om echt van nep te kunnen onderscheiden. Maar deze technologie ontwikkelt zich heel snel en het wordt steeds moeilijker om realiteit en illusie van elkaar te onderscheiden. Daar moeten we dus beter in worden.

En bij iets als wraakporno, bijvoorbeeld, waarbij ex-partners videobeelden verspreiden uit wraak, heeft de politie daar ook te maken met deze technologie?
Daar zou het kunnen gaan om een delict als 'afpersing' bijvoorbeeld. Eigenlijk doet het er dan niet zoveel toe of die beelden echt of nep zijn.

En als nu iemand uit zakelijk belang een concurrent in kwaad daglicht wil zetten door een video te publiceren waarin die concurrent verwerpelijke uitspraken doet?
Als zo'n video nep is, en gebruikt wordt voor dat doel, zou het wellicht smaad of laster kunnen zijn. Als de opnamen echt zijn, maar bijvoorbeeld uit zijn verband zijn gerukt of zo, dan wordt dat al lastiger, denk ik.

Zijn er nu ook zaken die nog niet zijn voorgekomen, maar wel toekomstscenario's zouden kunnen zijn? Stel: er verschijnt in de media een beeld van een dreigende verstoring van de openbare orde, en het lijkt alsof er een groep mensen op weg is om iets uit te halen?
Dat kan ik mij wel voorstellen, ja. Dan is het voor de politie natuurlijk zaak om geen 'trekpop' te zijn, om niet 'aan een touwtje' te zitten. Wij moeten dan vaststellen hoe reëel zo'n bericht is. Niet met toeters en bellen reageren als het niet nodig is. Dat is, trouwens, voor ons niet vreemd om te doen. Er zijn veel valse meldingen. Bij elke melding die wij krijgen moeten we nu ook al inschatten of we er wél of niet op moeten reageren.

Wij maken daarbij natuurlijk ook wel eens fouten en wij zullen in de toekomst vast ook op dit vlak fouten blijven maken. Maar, zoals gezegd, dat sommige mensen ons proberen te foppen, zijn we wel gewend. Wij hebben wel manieren om te kunnen verifiëren of een melding echt of nep is. Dus, hoewel ik denk dat wij in de toekomst verrast gaan worden door uitingen van deze nieuwe deepfake-technologie, hebben wij in de afgelopen decennia wel wat ervaring opgedaan met ongeveer vergelijkbare verschijnselen.

Positief scenario

Gelukkig valt er ook een positief scenario te bedenken bij alle van de hierboven geschetste ontwikkelingen. Daarbij nemen we als uitgangspunt dat we als samenleving steeds meer te maken krijgen met gemanipuleerde content. Het logische gevolg is dat we daaraan steeds meer gewend raken, zodat het geen indruk meer maakt. Dergelijke content wordt dan even gewoon als portretten die door een Snapchat-filter zijn verfraaid. In die context kan de mogelijkheid om met deepfake-technologie video's, stemmen en artikelen te produceren, een uitgelezen kans betekenen voor het stimuleren van goede journalistiek en waarheidsbevinding door nieuwsconsumenten.

Het eerste positieve gevolg zou inhouden dat wereldwijd de nieuwsconsument dankzij bovengenoemde ontwikkeling wordt gestimuleerd om nog serieuzer op zoek te gaan naar de werkelijke feiten. De nieuwsconsument kan zich dan uitgedaagd voelen om, binnen de wirwar van alle post *reality*-verslaglegging [45], op zoek te gaan naar hoe de vork écht in de steel zit. De nieuwsconsument kan het zich dan eigen maken om aan bronnenonderzoek, verificatie van gegevens en hoor- en wederhoor te doen.

Edited image from source: Why it's getting harder to spot a deepfake video https://edition.cnn.com/videos/business/2019/06/11/deepfake-videos-2020-election.cnnhttps://sputniknews.com/science/201907201076304335-deepnude-online-auction-starting-price/

In dat eerste positieve scenario is juist de druk die deepfake-technologie uitoefent op de scheidslijn tussen illusie en realiteit, een sterke drijfveer voor de consument om te onderzoeken naar wat waar is en wat niet. Die druk zorgt ervoor dat nieuwsconsumenten beter ingelezen, beter geïnformeerd en alerter worden, zonder te vervallen in cynisme.

In het meest ideale scenario noodzaakt deepfake-technologie de nieuwsconsument dus tot een actieve en gezond kritische rol. Dat zou een ideaal scenario zijn, maar het veronderstelt wel dat de gemiddelde nieuwsconsument een extra stap moet kunnen en willen zetten en die zich tevens bewust is van zijn of haar natuurlijke neiging om nieuws dat het eigen wereldbeeld bevestigt, te omarmen. In verscheidene interviews met professionals die voor dit rapport zijn gehouden ontstaat het beeld dat de gemiddelde nieuwsconsument bovenstaande stap onvoldoende heeft gezet. De verwachtingen hieromtrent zijn dan ook niet heel hoog.

Het tweede positieve gevolg kan zijn dat gerenommeerde nieuwsbronnen groeien in hun autoriteit. Die nieuwsbronnen zullen dan voor steeds meer mensen het eerste aangewezen loket zijn voor verificatie van nieuwsberichten. Partijen als het *NOS-journaal*, *The Washington Post* en *The New York Times* kunnen daarmee hun imago als betrouwbare partij enorm versterken. In dat ideale geval brengt deepfake-technologie het tegenovergestelde op gang van wat tot dusver als het meest waarschijnlijke risico van de technologie is geschetst. In mijn interview met Joost Schellevis, techjournalist bij de NOS, benoemt hij dit positieve gevolg en uit hij zijn vertrouwen erin dat de NOS in de toekomst meer in deze rol zal groeien.

INTERVIEW: JOOST SCHELLEVIS

TECHJOURNALIST BIJ DE NOS

Zijn jullie bij de NOS ook bezig met het verschijnsel van deepfake?
Wij hebben het op onze radar en denken daar ook echt wel over na. Op dit moment zien wij op de redactie echter nog niet veel verschijnselen van deepfake. En ik moet ook zeggen dat ik de kwaliteit ervan vaak ook nog niet heel overtuigend vind. Wanneer het onderscheid met de realiteit volledig weg is, dan wordt het natuurlijk gevaarlijk. Nu kan ik het verschil vaak nog wel met het blote oog zien. Ik vind deepfake-technologie voor nu dus ook nog niet zo'n groot probleem.

Wat zijn de risico's wanneer echt en nep niet meer van elkaar te onderscheiden zijn?
Op dat moment zijn er zeker wel risico's. Wij als redactie kunnen natuurlijk zelf in een filmpje trappen dat achteraf nep blijkt te zijn. Maar dat risico bestaat nu ook al. Op Twitter bijvoorbeeld zie je ook veel beeldmateriaal dat uit z'n context is getrokken, en bijvoorbeeld van een totaal andere gebeurtenis afkomstig is.
Het tweede risico is dat buitenstaanders een filmpje kunnen maken waarin ze presentatoren van het Achtuurjournaal dingen laten zeggen die zij niet daadwerkelijk gezegd hebben. Zo kan bijvoorbeeld hun reputatie en die van de NOS worden geschaad.

Hoe ga je daar mee om? En in de toekomst?
Ik denk dat het in het algemeen voor ons gewoon heel belangrijk is om onze journalistieke taak iedere dag nog beter uit te voeren. Wij hebben traditionele onderzoeksmethoden die al heel goed werken. Wij kijken bijvoorbeeld naar de bron van een verhaal, naar getuigen van vlees en bloed en er is forensische technologie om te ontdekken of een filmpje echt is of nep. Ik denk dat veel problemen die we in de toekomst met deepfake-technologie gaan tegenkomen, grotendeels getackeld gaan worden omdat bestaande methoden afdoende blijken te zijn.

Stel bijvoorbeeld dat er een filmpje opduikt waar een Nederlandse minister enorme hoeveelheden cocaïne gebruikt. Over een paar jaar kan deze politicus zeggen: dat was ik niet, dat was deepfake-technologie.
Dat zou inderdaad een risico kunnen zijn. Dat bewijslast die we willen gebruiken vanuit onze journalistieke rol gemakkelijker te ontkennen valt. Maar ook dan proberen we onze traditionele onderzoeksmethoden toe te passen. We gaan dan op zoek of er ook daadwerkelijk mensen bij zijn geweest die die gebeurtenis kunnen onderschrijven. Maar wellicht wordt het voor ons inderdaad wel wat lastiger wanneer het gaat om beelden van bijvoorbeeld een beveiligingscamera. Ook op dit gebied herkennen we de uitdagingen.

In mijn rapport beschrijf ik hoe de steeds grotere hoeveelheid nepnieuws ook apathie of desinteresse zou kunnen veroorzaken bij het algemene publiek waardoor de journalistiek haar kracht verliest en kwaadwillenden relatief vrij spel hebben. Hoe kijk je daar naar?
Ik vind het lastig om te speculeren over de toekomst, sommige websites verspreiden nu

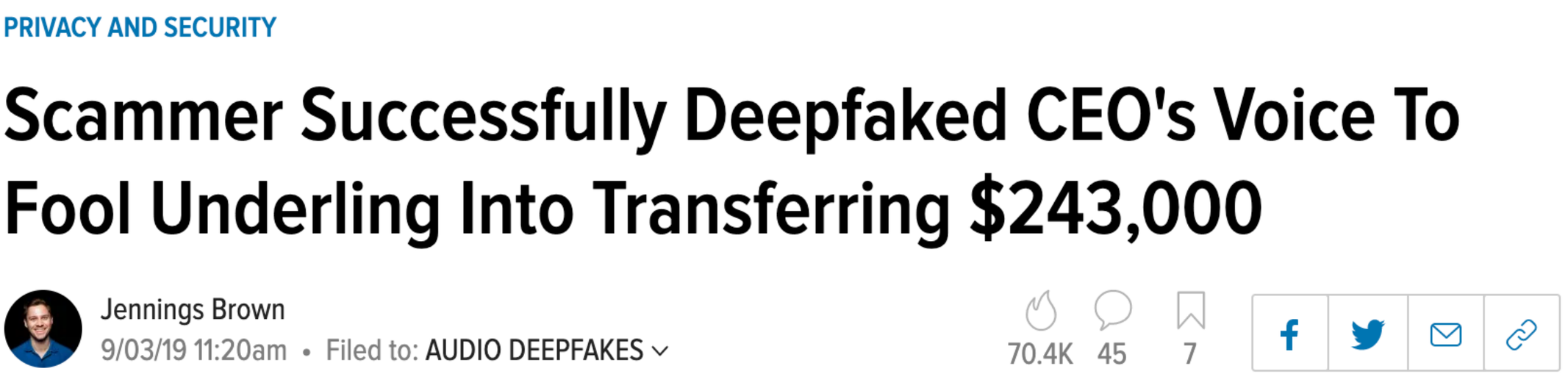
ook al nepnieuws. Het kan inderdaad zijn dat mensen hun interne kompas wat verliezen door de grote hoeveelheid nepnieuws die op hen afkomt. Dat moeten we nog even afwachten. Ik denk overigens dat het met de mediawijsheid van mensen in Nederland niet slecht gesteld is. Ik denk dat we ons meer zorgen moeten maken over de landen waar een minder sterke mediatraditie is. De bevolking daar is wellicht meer gevoeliger voor manipulatief nepnieuws.

Stel dat er een partij is, bijvoorbeeld Rusland, die geopolitieke spanningen wil veroorzaken door het nieuws te verspreiden dat een Amerikaanse militair een koran in het toilet heeft gegooid of dat er protesten zijn bij de Amerikaanse ambassade in Islamabad.
Ik snap wat je bedoelt, maar ik durf niet te zeggen of dat soort dingen gaan gebeuren. Er is nu ook al veel op het internet te vinden aan geruchten en nepnieuws. Het feit dat iets zou kunnen, betekent nog niet dat het ook gaat gebeuren. Het scenario dat je schetst lijkt me wat speculatief. En je moet ook niet vergeten: de meest simpele dingen worden over het algemeen het meest gebruikt. Jouw voorbeeld lijkt me wat te complex. Met deze simpele dingen bedoel ik bijvoorbeeld om een bestaande foto in een hele andere context te plaatsen.
Het zou overigens wel zo kunnen zijn dat er door de ontwikkeling van deze technologie het gemakkelijker wordt om op grote schaal realistische Twitter-bots in te zetten. Met echt lijkende profielen en foto's en betere teksten. Dat zou inderdaad nog wel kunnen gebeuren.

Ligt hier voor jullie bij de NOS dan ook een kans?
Dat denk ik wel. Kijk: Twitter is een prachtig medium om de allereerste nieuwsfeiten van een evenement naar je toe te halen. Er is niks sneller in de wereld op dat gebied. Maar al vrij snel begint de speculatie. En daar wordt het dus onbetrouwbaar. Ik denk dat wij ons kunnen onderscheiden omdat we betrouwbaar nieuws leveren. Wij controleren alles wat we plaatsen en willen niet speculeren. Wij beloven betrouwbaarheid. Ik denk namelijk niet dat je van de gemiddelde nieuwsconsument kunt verwachten dat die 100% mediawijs is. Dus die rol pakken wij heel serieus op.

6 | VOICE CLONING TECHNOLOGY

In alle genoemde deepfake-voorbeelden ligt het accent vooral op video. Dat komt omdat video een krachtig medium is en omdat de vooruitgang op dit gebied al ver gevorderd is. De manipulatie van audio gaat echter ook steeds beter, zoals bijvoorbeeld het kunnen klonen van iemands stem. De menselijke stem is belangrijk voor communicatie maar ook identificatie; op de radio herkennen we immers veel mensen aan hun stem.



PRIVACY AND SECURITY

Scammer Successfully Deepfaked CEO's Voice To Fool Underling Into Transferring $243,000

Jennings Brown
9/03/19 11:20am • Filed to: AUDIO DEEPFAKES

70.4K 45 7

Credit: Scammer Successfully Deepfaked CEO's Voice To Fool Underling Into Transferring $243,000
https://gizmodo.com/scammer-successfully-deepfaked-ceos-voice-to-fool-under-1837835066

Techblog *Gizmodo* beschreef in september 2019 bijvoorbeeld dat de CEO van een energiebedrijf in het Verenigd Koninkrijk dacht dat hij aan de telefoon sprak met de CEO van zijn moederbedrijf in Duitsland. Deze beller, met een Duits accent, liet hem € 220.000 overmaken naar een Hongaarse rekening. De stem van de Duitse CEO was echter gecreëerd met een kunstmatig intelligent systeem [46]. Deze vorm van fraude zal in de toekomst nog veel vaker voorkomen. De technologie is nog niet perfect, maar sommige financieel verantwoordelijken zullen toch op basis van een voicemail ten onrechte geld overmaken. *Voice cloning*-fraude is immers een nog erg onbekend verschijnsel, net zoals phishing dat was tijdens de opkomst van e-mail. Argeloos klikten destijds vele mensen op links in phishing-mails omdat ze niet wisten dat die bestonden.

Net zoals bij vele andere toepassingen van digitale technologie zal de kwaliteit van *voice cloning* snel toenemen. Het zal steeds gemakkelijker worden om uw eigen stem of die van iemand anders na te bootsen en die stem van alles te laten zeggen. Om de vergelijking met phishing-mails aan te houden: anno 2019 klikken nog steeds heel veel mensen op foute hyperlinks. In de toekomst zullen mensen ook in frauduleuze stemopnamen blijven trappen.

Politici, beroemdheden, klokkenluiders, journalisten en medewerkers van justitie en politie zijn vanzelfsprekend kwetsbaar voor chantage of reputatieschade wanneer deze *voice cloning*-software volmaakt is. Hun stem kan dan worden nagebootst en zo kan hen woorden in de mond worden gelegd. Dat kan willens en wetens worden gebruikt om iemand reputatieschade toe te brengen of te chanteren met reputatiebeschadiging.

Het toebrengen van reputatieschade wordt met behulp van *voice cloning*-software dus veel eenvoudiger. Een gedachte: Het gebeurt bijvoorbeeld wel eens dat iemand per ongeluk zijn of haar telefoon in een jaszak of broekzak laat bellen. De ontvanger kan dan onbedoeld meeluisteren met de stemmen en geluiden die de microfoon van de verzender registreert. Kwaadwillenden zouden nepversies van dergelijke telefonische geluidsopnamen kunnen genereren, die verzenden vanaf een anoniem nummer en met die opname de suggestie wekken alsof de ontvanger meeluistert met iets wat niet voor zijn of haar oren is bestemd. Een journalist ontvangt bijvoorbeeld een 'broekzaktelefoontje' met een onthulling van een politicus.
Een mogelijke potentiële investeerder in een bedrijf ontvangt een 'broekzak-voicemail' van de bestaande directeur waarin die neerbuigend spreekt over deze potentiële investeerder. Een fake ingesproken bericht kan via de smartphone worden verstuurd naar een collega, manager of bekende.

Er zijn vele toekomstscenario's denkbaar waarin nepopnamen van iemands stem worden misbruikt, bijvoorbeeld voor het ontfutselen van persoonlijke informatie voor een latere cyberhack of een rechtstreekse fraude-aanval. *"Hey hallo met Jarno, ik bel even met een andere telefoon omdat mijn iPhone kapot is en nu sta ik voor de deur van ons nieuwe kantoor, wat is ook alweer de entree-code? Die staat namelijk ook in mijn iPhone. App 'm maar even, dankjewel!"*

Credit: Amazon B0792KRW2J Echo Dot (3rd Gen) Voice Assistant With Alexa - Charcoal at The Good Guys.
https://www.thegoodguys.com.au/amazon-echo-dot-3rd-gen-voice-assistant-with-alexa---charcoal-b0792krw2j

Een gekloonde stem zou ook iemands digitale assistent een opdracht kunnen geven. "Alexa, maak $ 350 over aan het volgende rekeningnummer", "Alexa, stuur een bericht naar Harold waarin je hem vraagt de reis naar China te annuleren", "Alexa, verwijder alle afspraken uit mijn agenda tot 1 september." Aangezien wij steeds meer met stembesturing onze apparaten en software bedienen, is bovenstaand scenario gevaarlijk én niet onwaarschijnlijk.

INTERVIEW: LODEWIJK VAN ZWIETEN

OPENBAAR AANKLAGER CYBERCRIMINALITEIT, OPENBAAR MINISTERIE. OP PERSOONLIJKE TITEL.

Hoe kijk je aan tegen deepfake-technologie vanuit jouw rol bij het OM?
Ik volg deepfake-technologie al een tijdje en ik moet zeggen dat het mij best wel zorgen baart. De voorbeelden die ik ervan zie zijn van goede kwaliteit en ze worden steeds beter. Deepfake-video's zijn bijvoorbeeld al niet meer van echt te onderscheiden.
Vanuit mijn werk zie ik natuurlijk dat heel veel technologie niet wordt gemaakt met de intentie om er slechte dingen mee te doen, maar dat dat uiteindelijk wel gebeurt. Dat vrees ik met deepfake ook; veel technologie wordt uiteindelijk gebruikt met een criminele doelstelling. En wanneer iedereen in staat is om de werkelijkheid online te manipuleren, dan vind ik dat wel een zorgelijke ontwikkeling.

Met name omdat onderscheid tussen echt en nep, tussen goed en kwaad, steeds moeilijker te maken is. Wanneer je computer besmet wordt met een computervirus, dan weet je: ik moet antivirussoftware updaten of installeren. Bij deepfake-software is dat anders; de slechte intentie is vaak veel moeilijker te zien. Nepvideo's spelen bijvoorbeeld in op emoties als woede en angst of bevestigen bestaande vooroordelen. Deze technologie is dus veel slinkser, veel sluwer. Het moment van: 'nu moet ik alert zijn', is veel minder helder. En als ik nu al zie hoe snel mensen soms achterop de bagagedrager springen van manipulatief nepnieuws, dan maak ik me wel zorgen om de toekomst.

Wat is het verschil met bestaande manipulatie van nieuws?
Natuurlijk wordt nieuws al veel langer geframed, maar nu is er iets anders. Bij framing verander je als het ware het camerastandpunt jegens een nieuwsfeit. Bij deepfake-technologie bestaat dat nieuwsfeit helemaal niet in de echte wereld of wordt nieuwe informatie toegevoegd aan een bestaand nieuwsfeit. Het is op zichzelf al nep. En waar het raadplegen van verschillende nieuwsbronnen in het verleden een goed medicijn was tegen framing, vervalt deze optie wanneer de gebeurtenissen volledig nep zijn.

In welke vorm verwacht je deze technologie tegen te komen?
Ik verwacht in mijn werk deze technologie tegen te komen in verschillende verschijningsvormen. Bijvoorbeeld als voice cloning. Dat de stem van de bedrijfseigenaar of financieel verantwoordelijke wordt gecreëerd, nagemaakt op zo'n manier dat die niet meer van echt te onderscheiden is en dus kan worden gebruikt bij financiële fraude. Een CFO krijgt een voicemail van een CEO met een gesproken betalingsopdracht.

Ook verwacht ik dat deze voice cloning-technologie in de toekomst zo goed gaat worden dat deze software telefoongesprekken met mensen kan voeren en hen zo kan misleiden of geld afhandig kan maken. Ik geloof dat afpersing met nepseksvideo's ook een

probleem gaat worden. Daarbij worden dan mensen gedwongen tot bepaalde financiële transacties of het weggeven van bedrijfsgeheimen omdat ze worden afgeperst met een nepseksvideo.
Ik kan mij ook voorstellen dat deze software in de toekomst zo intelligent is dat die kan corresponderen via e-mail met potentiële slachtoffers. Of dat het ingezet wordt voor smaad; dat kwaadwillenden een nepvideo maken van iemand die een strafbaar feit begaat.

Overigens zijn deze zaken niet altijd zo gemakkelijk te veroordelen als dat ik graag zou willen. Een voice clone maken van iemands stem is op zichzelf bijvoorbeeld geen strafbaar feit, maar wanneer deze opname als instrument wordt gebruikt (of wordt gepoogd) om anderen geld afhandig te maken, dan weer wel. En zo is bijvoorbeeld het creëren van een nepseksvideo van iemand op je computer thuis op zichzelf niet strafbaar (met uitzondering van kinderporno), maar weer wel als je het gaat of probeert te verspreiden.

Zou zo'n deepfake-opname ook kunnen dienen als vals bewijsmateriaal?
Nee, dat denk ik niet. Niet zo snel althans. Ik ben niet bang dat een fakevideo gaat leiden tot een veroordeling of zo. Wanneer we strafrechtelijk onderzoek doen, doen we natuurlijk tevens uitgebreid forensisch technisch onderzoek. En het zal echt niet zo'n vaart lopen dat we iets niet als nep kunnen classificeren. Daar hebben we de expertise en technische oplossingen wel voor. En als we de echtheid niet kunnen vaststellen, moeten we het misschien buiten beschouwing laten. Het zal echter soms wel wat tijd van ons vragen. Je moet niet vergeten: een video is vrijwel nooit het enige bewijs voor een veroordeling. We verzamelen daarvoor vaak nog veel meer data. Wel ben ik ongerust over de maatschappelijke onrust die bepaalde video- of audio-opnamen kunnen veroorzaken. In hele brede zin, maar ook heel specifiek.

Wat bedoel je daarmee?
Stel dat er bijvoorbeeld een app komt waarmee je heel gemakkelijk een pornografische deepfake-video kunt maken: dan kun je erop wachten dat deze wordt gebruikt op een middelbare school. Een reputatie van een meisje kan dan razendsnel kapot worden gemaakt. En dan doet het er trouwens helemaal niet toe dat deze video nep is. Het feit dat deze video bestaat, brengt al schade toe. Dat zou je ook in breder maatschappelijk perspectief kunnen zien. Soms is het feit dat een bepaalde video bestaat en deze maatschappelijke onrust veroorzaakt, al genoeg. Dan doet het er niet meer toe dat een video nep is.

Wat kunnen we als burger dan doen?
Ik zou op zich graag willen dat mensen wat meer oplettend zouden worden (gemaakt) ten opzichte van wat ze zien of horen. Maar dat heeft ook met onze geschiedenis te maken: we verwachten namelijk in beginsel dat alles wat we horen en zien via tv, telefoon of het internet waarheidsgetrouw is. We leren 't langzaam: we snappen in 2019 inmiddels wel dat niet alles wat we op het internet lezen of zien de waarheid is. Maar: mensen overschatten zichzelf ('ik kijk daar wel doorheen') en zijn gewoonweg niet getraind op het herkennen van dit soort zeer gewiekste manipulatieve video's, artikelen of stemopnamen. Het zit echt in de genen van mensen om te geloven wat ze zien of horen. Het zou al heel veel schelen als Apple een waarschuwing zou geven als je gesprekspartner in een Facetime-gesprek de beeldmanipulatietechniek die daarin zit, gebruikt.

7 | OMGAAN MET DEEPFAKE

In algemene zin zijn de problemen van deepfake-video's helder en inmiddels zijn er ook oplossingen voorgesteld. Het meest voor de hand liggend is die te zoeken in de hoek van de technologie zelf: dat kan zowel bij het registreren van feitelijke gebeurtenissen als ook in de detectie en filtering van nep-content. Maar ook wetgeving en voldoende educatie moeten gepaste weerstand bieden tegen deepfakes.

Technologische oplossingen

Hoewel er op dit moment nog geen perfecte technologische oplossing is voor de problemen die deepfake-content ons bezorgt, wordt er hard aan gewerkt. We zullen een aantal technologische invalshoeken [47] nagaan waarmee het deepfake-probleem wordt getackeld.

Ten eerste zijn er technologische oplossingen aan het begin van het proces, waar wordt geprobeerd feitelijke gebeurtenissen onweerlegbaar vast te leggen als daadwerkelijke feiten [48]. Ook het kunnen ontmaskeren of juist verifiëren van bepaald materiaal met behulp van technologie komt daarbij aan bod. Ten tweede blijkt technologie een belangrijk hulpmiddel aan het einde van de keten, voor het tegengaan van verspreiding van deepfake-content.

Controlled capture

De opkomst van deepfake-technologie zorgt voor urgentie bij wetenschappers en bedrijven om oplossingen te creëren. Een manier waarop nepnieuws, chantage en reputatieschade kunnen worden tegengegaan, is bijvoorbeeld om gebeurtenissen vast te leggen met *controlled capture*. Daarbij worden bijvoorbeeld tijd, locatie en daadwerkelijke gebeurtenis versleuteld in apps met blockchaintechnologie.

Credit: Truepic | Technology. https://truepic.com/technology/

De blockchain [49] is een gedeelde database, verdeeld over duizenden computers wereldwijd, waarbij achteraf geen aanpassingen aan de geregistreerde gegevens meer mogelijk zijn. Alles dat wordt geregistreerd in de blockchaindatabase ligt onwrikbaar vast. Middels encryptie- en blockchaintechnologie kunt u daardoor met een hoge mate van zekerheid vaststellen dat bepaalde gebeurtenissen hebben plaatsgevonden op een bepaalde tijd en locatie.

Er zijn inmiddels verscheidene applicaties die deze *controlled capture* aanbieden, zoals Truepic [50]. U kunt daarmee met de camera van uw smartphone onweerlegbaar bepaalde gebeurtenissen vastleggen, waarbij zeer specifiek bijvoorbeeld tijd, smartphone-identiteit en locatie worden geregistreerd. Verder bieden sommige applicaties aan om beelden dusdanig te watermerken dat het onmogelijk is om de beelden te vervalsen met generatieve AI-software. Sommige *controlled capture*-software detecteert bijvoorbeeld als opnamen afkomstig zijn van externe beeldschermen en verifieert die beelden dan vervolgens niet. Dat soort software zal steeds vaker door zowel beroeps- als burgerjournalisten worden gebruikt om nieuws te registreren, zodat op ieder later moment kan worden vastgesteld dat de gebeurtenissen hebben plaatsgevonden met een aan zekerheid grenzende waarschijnlijkheid. Dat is belangrijk, in een wereld waarin echt en nep steeds meer in elkaar vervloeien.

Life Logging

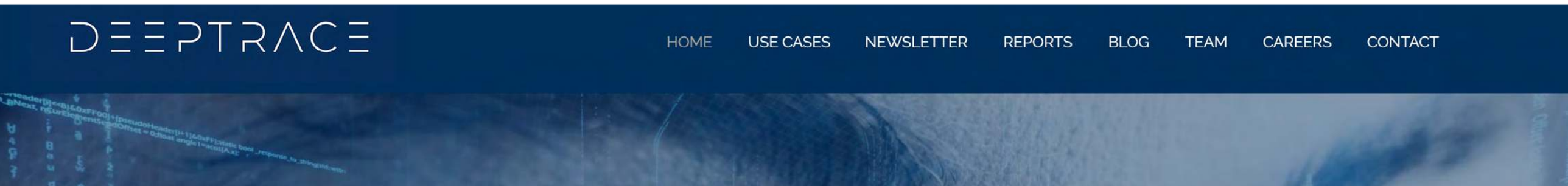
Life logging is een manier van handelen waarbij software het dagelijkse leven nauwgezet registreert. Een vorm van *controlled capture* dat op sommige aspecten geautomatiseerd gaat vastleggen wat iemand aan het doen is. Belangrijke politici, beroemdheden, journalisten en zakenlieden zullen in de toekomst wellicht intensief daarmee hun dagelijks leven vastleggen. Wanneer ze dat doen, beschikken ze immers altijd over een ijzersterk alibi wanneer er bijvoorbeeld belastende videobeelden [51]

opduiken. Door minutieus vast te leggen wat ze doen en dat te verankeren in blockchaintechnologie hebben ze het bewijs bij de hand wanneer anderen middels deepfake-materiaal hun reputatie willen beschadigen of hen willen chanteren.

Een belangrijke voetnoot daarbij is echter dat het op de lange termijn denkbaar is dat de overheid op een zeker moment dergelijke databases wil inzien, aan de hand van een gerechtelijk bevel. En hoe weet de cliënt zeker dat de commerciële bedrijven die dit soort diensten aanbieden, uiteindelijk niet de gegevens gaan verkopen aan adverteerders of andere data-handelaars? Dat is de steeds weer terugkerende keerzijde van registratie: niemand weet hoe gegevens later kunnen worden gebruikt.

Detectie en filtering

Een van de belangrijkste mogelijkheden om de negatieve gevolgen van deepfake-technologie tegen te gaan, is het gebruik van kunstmatige intelligentie voor detectie en filteren. In de toekomst moet het gemakkelijk zijn video's te testen om te zien of ze zijn gemanipuleerd. Een internetbrowser krijgt in de toekomst hopelijk geïntegreerde software die gemanipuleerde video's markeert of blokkeert. Inmiddels zijn verscheidene bedrijven bezig met dit soort detectie- en filtersoftware, zoals Deeptrace [52] en Deepfact [53].



Credit: Deeptrace | The antivirus for deepfakes.
https://www.deeptracelabs.com/

In onze zoektocht naar de waarheid zullen we steeds meer afhankelijk worden van kunstmatig intelligente systemen die de beoordeling doen, zeker wanneer de menselijke zintuigen, het oog en het oor, de gemanipuleerde video- en audio-opnamen niet meer van echt kunnen onderscheiden. Dan moeten we ons wenden tot kunstmatig intelligente systemen om ons te helpen.

Uiteindelijk is het de bedoeling dat die software in realtime nepvideo's als zodanig kan detecteren. Dat is bijvoorbeeld zeer belangrijk bij *breaking news*. Kunstmatig intelligente systemen kunnen dat bijvoorbeeld doen door het extreem subtiele verschil op te merken tussen lipbewegingen en gesproken audio, het herkennen van een te homogene kleur in de huid of het ontbreken van de subtiele weergave van een hartslag. Ook een onregelmatige of bijzondere schaduw op het gezicht of onregelmatige bewegingen van mensen en voorwerpen kunnen een indicatie zijn van een nepvideo. De kracht van kunstmatig intelligente systemen is dat ze veel gedetailleerder kunnen zoeken naar oneffenheden. Daar waar het menselijk oog ontoereikend is, ligt ook de meerwaarde van dit soort technologie.

Credit: Where the 2020 Democratic Candidates Stand on Reparations — Inside the 2020 Presidential Debate Around Reparations.
https://www.elle.com/culture/career-politics/a27170939/2020-democratic-reparations-issues/

Soft biometric

Sommige politici of beroemdheden zullen beter dan gemiddeld worden beschermd [54] door bepaalde software, omdat van hen veel meer video- en audiomateriaal beschikbaar is. Een detectiemachine leert daarmee de zeer specifieke persoonlijke trekjes te herkennen, zoals het optrekken van een wenkbrauw, het leggen van accent op bepaalde klemtonen en het draaien van het hoofd. Deze *soft biometric*-kenmerken helpen bij het onderscheid maken tussen een echte en een nepvideo. Die politici en beroemdheden zullen daardoor beter beschermd zijn tegen deepfake-video's dan de doorsnee burger.

Nepdetector in de webbrowser

Uiteindelijk zullen we als consument waarschijnlijk een plug-in in onze browser krijgen die gelijk kan detecteren of een video nep is of niet. Deze software zal niet waterdicht zijn, maar toch werken verschillende partijen eraan om de 'bulk' van nepvideo's te kunnen filteren. De grote socialemediaplatformen zijn ook bezig om detectiesoftware te ontwikkelen die hun platform verschoond houden van deepfakes.

Socialemediabedrijven

Socialemediabedrijven hebben de verplichting haatdragende content van hun platform te verwijderen. Facebook startte in september 2019 mede vanwege deze reden zelfs met een 'Deepfake Detection Challenge' [55]. Het is overigens de vraag of die socialemediabedrijven bijzonder krachtig gaan optreden tegen de verspreiding van deepfake-content. Zij verdienen immers hun geld met de tijd die gebruikers doorbrengen op het platform en het klikken op advertenties door die gebruikers. Door de viraliteit van fake-content, vormt die voor de socialemediaplatformen ook een bron van inkomsten.

Credit: Tackling the 'Deep Fake,' House Grasps for Solution to Doctored Videos.
https://www.courthousenews.com/tackling-the-deep-fake-house-grasps-for-solution-to-doctored-videos/

Er zijn wetenschappers die pleiten voor een 'deepfake-vertragingsmodule' [56] wanneer video's worden geüpload bij de grote socialemediaplatformen. Dat geeft die platformen namelijk een kans om content zorgvuldig te scannen en virale verspreiding waar nodig tegen te gaan. Een andere oplossing zou zijn wanneer de grote socialemediaplatformen de meer 'betrouwbare' en 'gewaardeerde' nieuwsbronnen op hun platform meer bereik geven en profielen en pagina's die vaker nepvideo's [57] hebben getoond, geen bereik meer geven. In de Amerikaanse politiek [58 | 59] gaan er stemmen op om Facebook, Twitter en YouTube te verplichten software voor detectie en filteren van deepfake-content te ontwikkelen en te implementeren. Dat is geen gekke gedachte, gezien hun grote bereik. Het maken, blijvend actualiseren en onderhouden van dat soort detectie- en filter-software kost echter veel tijd, geld en aandacht. De software moet constant geüpdatet worden. Dat verschaft de socialemedia-reuzen indirect een veel machtiger positie ten opzichte van kleinere concurrenten. Alleen de reuzen hebben immers de mogelijkheid dat soort software te ontwikkelen, implementeren en te onderhouden. Beginnende concurrerende startups wordt het daardoor nog moeilijker gemaakt om de grote namen uit te dagen.

Hany Farid, een computerwetenschapper van de Universiteit van Dartmouth die zich specialiseert in het onderzoeken van gemanipuleerde foto's en video's, maant de grote techreuzen dikwijls om haast te maken met hun ontwikkelingen: "Als een bioloog zegt: 'Hier is een echt cool virus; laten we eens kijken wat er gebeurt als het publiek dat in handen krijgt', dan zou dat niet acceptabel zijn. En toch is het wat Silicon Valley de hele tijd doet. Het is een indicatie van een zeer onvolwassen industrie. We moeten eerst de mogelijke risico's begrijpen en daarom de manier waarop we technologie als deze [60] inzetten, vertragen".

Ter geruststelling, veel technologische oplossingen zijn al in ontwikkeling en het is de verwachting dat een gedeelte van alle synthetisch gecreëerde media in de toekomst wordt herkend door kunstmatig intelligente software. Maar het blijft belangrijk te beseffen dat we niet volledig kunnen vertrouwen op technologie.
Het blijft een kat-en-muisspel tussen makers en speurders [61]. Voor journalisten blijft vanzelfsprekend het traditionele bronnenonderzoek daarom van groot belang. En de inzet van de eigen waarneming [62].

Eigen waarneming

Stel: u bent journalist, pr-professional of communicatieadviseur en u moet kunnen bepalen of een video echt is of niet.
Natuurlijk komt er software op de markt die op een veel specifiekere manier kan kijken naar videobeelden. En vaak ziet software oneffenheden die het menselijk oog niet kan waarnemen.
Heeft u geen intelligente software tot uw beschikking en moet u het doen met uw eigen oren, ogen en gezond verstand, dan heb ik hier een aantal tips [63|64|65|66].

- Kijk of u de video in een videobewerkingsprogramma kunt laden en bepaalde frame- onderdelen specifieker kunt bekijken, op zoek naar onnatuurlijke vormen of andere toegevoegde elementen.
- Controleer of er oudere versies van de videobeelden online staan.
- Vanzelfsprekend zijn ook alle redactionele controles van toepassing: een e-mail of telefoontje naar de bron doet vaak al wonderen.
- Ook kunt u kijken naar oneffenheden in de video: zo vloeien mensen en achtergronden nog wel eens in elkaar over en zijn er onnatuurlijke bewegingen, afwijkende kleurvlekken of surrealistische objecten te zien. Kijk goed naar de details zoals handen, tanden, haar en oren. Deze zijn voor een computer het meest lastig om na te maken. Ze zijn onnatuurlijk gevormd of vloeien teveel in elkaar over.
- Kijk naar vreemde schaduwen op gezichten of ledematen en of een gezicht een onnatuurlijke variatie heeft in kleuren. Let ook op of achtergronden die (on-)natuurlijk overkomen.

Voor nog meer tips over omgaan met gemanipuleerde video's in het algemeen verwijs ik u graag naar *The Washington Post's guide to manipulated video.*

Wetgeving

Zoals dat gaat met iedere nieuwe krachtige technologische ontwikkeling, klinkt al snel de roep om nieuwe wetgeving. Zo eenvoudig als dat klinkt, zo weerbarstig blijkt de praktijk. Allereerst is er al veel wetgeving rondom smaad, laster, reputatieschade, verspreiden van haat, identiteitsfraude, copyright, auteursrecht enzovoort. Het is daarom de vraag of er überhaupt behoefte is aan nieuwe wetgeving.

Het eerste wetsvoorstel in Amerika gericht op deepfakes, de Malicious Deep Fake Prohibition Act [67], werd in december 2018 ingediend en de Deepfakes Accountability Act [68] volgde in juni van 2019. In verschillende staten, waaronder Virginia [69], Californië, New York en Texas, is ook deepfake-wetgeving ingevoerd. Het is aannemelijk dat er in de Verenigde Staten meer regelgeving op komst is, waarschijnlijk in de vorm van socialemediaregulering. In de Verenigde Staten wordt enorme haast gemaakt met deze wetgeving [70], waarschijnlijk met het oog op de Amerikaanse presidentsverkiezingen van 2020.

In Nederland is er op dit moment nog geen officiële specifieke wetgeving voor deepfake-video's. Wel staat deze ontwikkeling in politiek Den Haag op de kaart blijkt uit een recente kamerbrief [71|72]. Ook wordt in deze kamerbrief gesproken over educatie: mensen bekend maken met deze technologie zodat zij bewuster omgaan met de informatie die zij tot zich nemen via het internet.

Educatie

En dat is een belangrijk onderdeel:
dat burgers kennisnemen van deze technologische deepfake-ontwikkeling. Dat er kennis ontstaat over de uiteenlopende mogelijkheden om digitale content te vervalsen. Dat deepfake niet alleen beperkt blijft tot de grappige gezichtsfilters op je smartphone, maar dat deze toepassing van deepfake technologie breed inzetbaar is. Dat het ook de mogelijkheid creëert om teksten, video's, menselijke stemmen en andere audio opnames te genereren.

In de afgelopen jaren is er bij scholen en bedrijven steeds meer aandacht voor mediawijsheid. Voor het kunnen ontleden, analyseren en duiden van alledaags nieuws. De grote hoeveelheid informatie die dagelijks tot ons komt moet immers goed beoordeeld en op waarde kunnen worden geschat. Aan het 'boek van mediawijsheid' moet op korte termijn het nieuwe 'deepfakes' hoofdstuk worden toegevoegd: informatie die echt lijkt te zijn, maar volledig nep is. Waarbij het zintuiglijk waarnemen met ogen en oren vaak niet afdoende meer is om de echtheid te kunnen bepalen.
Overheid, bedrijven en onderwijsinstellingen zullen de komende jaren een actieve rol moeten aannemen om gezamenlijk de samenleving bewust te maken van deze technologische trend. Een belangrijke uitdaging daarbij is de mogelijke neiging tot onverschilligheid ten aanzien van al het nieuws. Onverschilligheid vanuit de overtuiging dat alles wat je ziet, hoort of leest nep kan zijn.
En dat is misschien nog wel een grotere uitdaging dan het creëren van het bewustzijn van het fenomeen op zichzelf: dat burgers niet apathisch worden ten aanzien van al het nieuws.

Voor verdere verdieping aangaande de bedreigingen en oplossingen van deepfake technologie kan ik u het rapport *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security* [73] sterk aanbevelen.

8 | CONCLUSIE

Generatieve AI-software is een verzamelbegrip voor alle verschillende AI-technologieën die gemeen hebben dat ze nieuwe digitale content, ideeën, hypotheses en invalshoeken kunnen creëren. Wij gaan de komende jaren de zoete vruchten plukken van machines met een *vonk van verbeeldingskracht*. Diezelfde software laat zich echter ook inzetten voor negatieve doeleinden. Je kunt daarmee volledig nieuwe geloofwaardige video's creëren, in video's onderling gezichten verwisselen, iemands stem nabootsen, geloofwaardige nepteksten genereren en objecten wegpoetsen alsof ze nooit hebben bestaan. We naderen een tijdperk waarin we online onze ogen en oren niet meer kunnen vertrouwen.

Dat heeft gevolgen. Toename van risico's als sociale onrust, geopolitieke spanningen, chantage en reputatieschade zijn niet ondenkbaar. Ook ontstaat er afbreuk van bewijslast tegen overtreders van morele of juridische grenzen. Wanneer bewijslast kan worden afgedaan als deepfake-technologie, verdwijnt de kracht van journalistiek. En wanneer wij als samenleving ook nog eens onverschillig worden jegens nieuwsonthullingen is dat zelfs een bedreiging voor onze democratie. Hoewel er op lange termijn technologische oplossingen zullen zijn die een gedeelte van de deepfake-content zullen filteren, zal dat nooit afdoende zijn, de wedloop tussen echt en nep is al aan de gang. In dat licht bekeken creëert deepfake-technologie vanaf nu voor ons een *nieuwe realiteit*.

Overige bronnen

A digital breadcrumb trail for deepfakes – Axios

https://www.axios.com/deepfake-authentication-privacy-5fa05902-41eb-40a7-8850-5450bcad0475.html

A Spy Used a Deepfake Photo to Infiltrate LinkedIn Networks

https://futurism.com/the-byte/spy-deepfake-photo-infiltrate-linkedin-networks

AI = Artificial Imagination?

https://knowledge.insead.edu/node/11761/pdf

AI is making inroads in scientific discovery and innovation https://www.allerin.com/blog/when-the-invention-becomes-the-inventor-how-ai-is-making-inroads-in-scientific-discovery-and-innovation

An optimistic view of deepfakes – TechCrunch

https://techcrunch.com/2019/07/04/an-optimistic-view-of-deepfakes/

Congress' flawed proposals to regulate deepfakes.

https://slate.com/technology/2019/07/congress-deepfake-regulation-230-2020.html

Dali Atomicus: Phillipe Halsman & Salvador Dali's Photography

https://www.youtube.com/watch?v=pbi94KWIDwQ

De authenticiteit van nep. Siri Beerends – YouTube

https://www.youtube.com/watch?v=u5Ez80EyUqo

Deep Dream comes true – Merzazine – Medium

https://medium.com/merzazine/deep-dream-comes-true-eafb97df6cc5

deepart.io

https://deepart.io/

Deepfake debunking tool may protect presidential candidates. For now.

https://www.cnet.com/news/deepfake-debunking-tool-may-protect-2020-presidential-candidates-trump-warren-obama-hillary-clinton/

Deepfake videos: Inside the Pentagon's race against disinformation

https://edition.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/

Deepfakes – Center for Internet and Society

http://cyberlaw.stanford.edu/our-work/topics/deepfakes

Deepfakes and Synthetic Media: What should we fear? What can we do? - WITNESS Blog

https://blog.witness.org/2018/07/deepfakes/

Deepfakes and the New Disinformation War – International Prose – Inverse Times Newswire

https://inversetimes.hopto.org/news/story-442.html?

Deepfakes Are Bad but They Could Also Have Some Advantages

https://interestingengineering.com/deepfakes-are-bad-but-what-are-some-of-the-possible-advantages

Deepfakes Are Getting Better. But They're Still Easy to Spot – WIRED

https://www.wired.com/story/deepfakes-getting-better-theyre-easy-spot/

Deepfakes are here, now what? – The Internet Health Report 2019

https://internethealthreport.org/2019/deepfakes-are-here-now-what/

Deepfakes aren't a tech problem. They're a power problem – Oscar Schwartz –

https://www.theguardian.com/commentisfree/2019/jun/24/deepfakes-facebook-silicon-valley-responsibility

Deepfakes: An Unknown and Uncharted Legal Landscape

https://towardsdatascience.com/deepfakes-an-unknown-and-uncharted-legal-landscape-faec3b092eaf

DensePose: Dense Human Pose Estimation In The Wild

https://www.youtube.com/watch?time_continue=4&v=Dhkd_bAwwMc

Echt Nep.pdf SanderDuivestein.

https://drive.google.com/file/d/1AG2hP3_pTTLbToNgy7xBEOxuVoH8TGyh/view

Engineering deepfake.pdf

https://engineering.purdue.edu/~dgueraco/content/deepfake.pdf

Ganbreeder

https://ganbreeder.app/category/random

GANs: tooling van morgen

https://www.ictmagazine.nl/achter-het-nieuws/gans-tooling-van-morgen/

Generating Character Animations from Speech with AI – NVIDIA Developer News Center

https://news.developer.nvidia.com/generating-character-animations-from-speech-with-ai/

High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs –

https://www.youtube.com/watch?v=3AIpPlzM_qs

Hoe gaat u deepfake en nepnieuws tegen? - DataExpert

https://dataexpert.nl/nieuws-overzicht/2019/04/zien-is-niet-meer-altijd-geloven-hoe-gaat-u-deepfake-en-nepnieuws-tegen

How Artificial Intelligence Is Changing Science – Quanta Magazine

https://www.quantamagazine.org/how-artificial-intelligence-is-changing-science-20190311/

How Do You Spot a Deepfake? It Might Not Matter

https://nymag.com/intelligencer/2019/06/how-do-you-spot-a-deepfake-it-might-not-matter.html

How to spot deepfake videos – and why you should care

https://www.avanade.com/en/blogs/avanade-insights/artificial-intelligence/how-to-spot-deepfake-videos

How we teach computers to understand pictures – Fei Fei Li – YouTube

https://www.youtube.com/watch?v=40riCqvRoMs

Imagination Machines: A New Challenge for Artificial Intelligence

https://pdfs.semanticscholar.org/d3c6/4b2497fb02d496709c4fa8fff00f4581399c.pdf

Lies, Line Drawing, and (Deep) Fake News

https://digitalcommons.law.ou.edu/cgi/viewcontent.cgi?article=1343&context=olr

Microsoft Word - 2019.Perception Inception Report.V2.EMBARGOED TILL 21 MAY.docx

https://static1.squarespace.com/static/5ca2c7abc2ff614d3d0f74b5/t/5ce2634eb6e19700014
2e716/1558340442631/2019.Perception+Inception+Release+EMBARGOED+TILL+21+MAY.PDF

Neural Networks: pix2pix (Conditional GANs) for Facial Segmentation, face2sketch, and sketch2face! – YouTube

https://www.youtube.com/watch?v=vrvwfFej_r4

New AI Imaging Technique Reconstructs Photos with Realistic Results – NVIDIA

https://news.developer.nvidia.com/new-ai-imaging-technique-reconstructs-photos-with-realistic-results/

New deepfake algorithm allows you to text-edit the words of a speaker in a video

https://newatlas.com/edit-talking-head-text-deepfake/60160/

Nightmarish: Lawmakers brace for swarm of 2020 deepfakes – POLITICO

https://www.politico.com/story/2019/06/13/facebook-deep-fakes-2020-1527268

Object-driven Text-to-Image Synthesis via Adversarial Training

https://arxiv.org/abs/1902.10740

People get better at catching deepfakes with practice, research says – VentureBeat

https://venturebeat.com/2019/07/12/people-get-better-at-catching-deepfakes-with-practice-research-says/

Progressive Growing of GANs for Improved Quality, Stability, and Variation –

https://www.youtube.com/watch?v=XOxxPcy5Gr4&t=69s

Research at NVIDIA: AI Reconstructs Photos with Realistic Results – YouTube

https://www.youtube.com/watch?v=gg0F5JjKmhA

Research at NVIDIA: Generating and Editing High-Resolution Synthetic Images with GANs –YouTube

https://www.youtube.com/watch?v=G6o_7Pz35Sk

Research at NVIDIA: Medical Image Synthesis for Data Augmentation and Anonymization Using GANs – YouTube

https://www.youtube.com/watch?v=BMuFk2PjEuM

Six lessons from my deepfake research at Stanford – Medium

https://medium.com/jsk-class-of-2019/six-lessons-from-my-deepfake-research-at-stanford-1666594a8e50

Synthesizing Audio with Generative Adversarial Networks

https://towardsdatascience.com/synthesizing-audio-with-generative-adversarial-networks-8e0308184edd

Text-based Editing of Talking-head Video – YouTube

https://www.youtube.com/watch?v=0ybLCfVeFL4

The coming deepfakes threat to businesses – Axios

https://www.axios.com/the-coming-deepfakes-threat-to-businesses-308432e8-f1d8-465e-b628-07498a7c1e2a.html

The Cyberlaw Podcast-227.pdf

https://www.steptoe.com/images/content/1/7/v2/176543/TheCyberlawPodcast-227.pdf

The holy grail in artificial intelligence – YouTube

https://www.youtube.com/watch?v=dfiE7uBIcWk

The Newest AI-Enabled Weapon: 'Deep-Faking' Photos of the Earth - Defense One

https://www.defenseone.com/technology/2019/03/next-phase-ai-deep-faking-whole-world-and-china-ahead/155944/

The Rise Of Deepfake And Media Synthetization.pdf

https://www.pkflawyers.com/pdf/TheRiseOfDeepfakeAndMediaSynthetization.pdf

The State of Deepfakes: Reality Under Attack. A 2018 Report [pdf] – Hacker News

https://news.ycombinator.com/item?id=18806020

The Synthetic Generation: Sogeti VINT rapport

https://www.sogeti.com/globalassets/global/downloads/reports/digital-happiness/TheSyntheticGeneration.pdf

To Catch a Fake: Machine learning sniffs out its own machine-written propaganda – ZDNet

https://www.zdnet.com/article/to-catch-a-fake-machine-learning-sniffs-out-its-own-machine-written-propaganda/

Toward Multimodal Image-to-Image Translation BicycleGAN –

https://www.youtube.com/watch?v=JvGysD2EFhw&t=0s

Toward Multimodal Image-to-Image Translation

https://junyanz.github.io/BicycleGAN/

Trashy Muse put on the world's first virtual avatar fashion show – Dazed

https://www.dazeddigital.com/fashion/article/45206/1/trashy-muse-virtual-avatar-fashion-show-augmented-reality-lil-miquela-paris

Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

https://arxiv.org/abs/1703.10593

Vincent NVIDIA GTC 2017 Europe: Generative Adverarial Paintings – YouTube

https://www.youtube.com/watch?v=P1sBNac83ls

Wat is Synthetische Transformatie? Raimo van der Klein – YouTube

https://www.youtube.com/watch?v=tVS-eHydsMQ&t=1139s

Why Artificial Intelligence will enable new scientific discoveries

https://www.graphcore.ai/posts/why-artificial-intelligence-will-allow-us-to-make-new-scientific-discoveries

Will Scientific Research be able to avoid Artificial Intelligence pitfalls?

https://towardsdatascience.com/will-scientific-research-be-able-to-avoid-artificial-intelligence-pitfalls-b818e96c0cdd

WITNESS Media Lab – OSINT Digital Forensics – WITNESS Media Lab

https://lab.witness.org/projects/osint-digital-forensics/

Zien is geloven? Zo ga je mediawijs om met deep fake https://www.mediawijsheid.nl/deep-fake/

Eindnoten

1 **Creative Commons — Attribution-NonCommercial 4.0 International — CC BY-NC 4.0**
https://creativecommons.org/licenses/by-nc/4.0/

2 **Deepfake – Wikipedia**
https://en.wikipedia.org/wiki/Deepfake

3 **Deepfakes: we naderen een tijdperk waarin we onze oren en ogen niet langer kunnen vertrouwen – de Volkskrant**
https://www.volkskrant.nl/columns-opinie/deepfakes-we-naderen-een-tijdperk-waarin-we-onze-oren-en-ogen-niet-langer-kunnen-vertrouwen~be63ceb2/

4 **BNR Nieuwsradio – Interview Jarno Duursma Deepfakes**
https://www.bnr.nl/cookiewall?target=/podcast/beeldbepalers/10381775/nepnieuws-raakt-in-versnelling-met-deepfake-video-s

5 **Jouw gezicht in nepvideo – RTL Nieuws**
https://www.rtlnieuws.nl/editienl/laatste-videos-editienl/video/4827811/jouw-gezicht-nepvideo

6 **Deepfakes: we naderen een tijdperk waarin we onze oren en ogen niet langer kunnen vertrouwen – de Volkskrant**
https://www.volkskrant.nl/columns-opinie/deepfakes-we-naderen-een-tijdperk-waarin-we-onze-oren-en-ogen-niet-langer-kunnen-vertrouwen~be63ceb2/

7 **Het gevaar van deepfakes - Wat het daglicht niet verdragen kan - NPO Radio 1**
https://www.nporadio1.nl/wat-het-daglicht-niet-verdragen-kan/onderwerpen/511875-het-gevaar-van-deepfakes

8 **De digitale butler - Kansen en bedreigingen van kunstmatige intelligentie. Jarno Duursma. Zaltbommel: Haystack, 2017.**
https://www.managementboek.nl/boek/9789461262424/de-digitale-butler-kansen-en-bedreigingen-van-kunstmatige-intelligentie-jarno-duursma

9 **Iconic Abraham Lincoln portrait revealed to be two pictures stitched together.**
https://www.dailymail.co.uk/news/article-2107109/Iconic-Abraham-Lincoln-portrait-revealed-TWO-pictures-stitched-together.html

10 **Adnan Hajj photographs controversy – Wikipedia**
https://en.wikipedia.org/wiki/Adnan_Hajj_photographs_controversy

11 **A face-swapping app takes off in China, making AI-powered deepfakes for everyone**
https://www.nbcnews.com/tech/security/face-swapping-app-takes-china-making-ai-powered-deepfakes-everyone-n1049501

12 **He Predicted The 2016 Fake News Crisis. Now He's Worried About An Information Apocalypse.**
https://www.buzzfeednews.com/article/charliewarzel/the-terrifying-future-of-fake-news

13 **/r/Deepfakes was banned February 2018**
https://www.reddit.com/r/AgainstSubredditBans/comments/9g571u/rdeepfakes_was_banned_february_2018/

14 **You Won't Believe What Obama Says In This Video! – YouTube**
https://www.youtube.com/watch?v=cQ54GDm1eL0

15 **Kim Kardashian Deepfake Taken Off of YouTube Over Copyright Claim – Digital Trends**
https://www.digitaltrends.com/social-media/kim-kardashian-deepfake-removed-from-youtube/

16 **Everybody Dance Now – YouTube**
https://www.youtube.com/watch?v=PCBTZh41Ris

17 **These Full-Body Deepfakes are Like Nothing We've Ever Seen**
https://futurism.com/full-body-deepfakes

18 **Deepfakes are solvable—but don't forget that "shallowfakes" are already pervasive – MIT Technology Review**
https://www.technologyreview.com/s/613172/deepfakes-shallowfakes-human-rights/

19 **Doctored Pelosi video highlights the threat of deepfake tech – YouTube**
https://www.youtube.com/watch?v=EfREntgxmDs

20 **Fake Facebook video of Nancy Pelosi drunk shows its danger to truth**
https://eu.usatoday.com/story/opinion/2019/05/28/facebook-fake-video-nancy-pelosi-drunk-responsibility-column/1249830001/

21 **Facebook Refuses To Remove Fake "Drunk" Video Of Nancy Pelosi**
https://www.buzzfeednews.com/article/davidmack/facebook-nancy-pelosi-doctored-video

22 **Facebook CEO says delay in flagging fake Pelosi video was 'execution mistake' – Reuters**
https://www.reuters.com/article/us-facebook-deepfake/facebook-ceo-says-delay-in-flagging-fake-pelosi-video-was-execution-mistake-idUSKCN1TS023

23 **Bill Posters op Instagram: "Mark Zuckerberg reveals the truth about Facebook"**
https://www.instagram.com/p/ByaVigGFP2U/

24 **Deepfakes zijn een groeiend gevaar voor onze democratie**
https://fd.nl/opinie/1305951/deepfakes-zijn-een-groeiend-gevaar-voor-onze-democratie

25 **Another convincing deepfake app goes viral prompting immediate privacy backlash – The Verge**
https://www.theverge.com/2019/9/2/20844338/zao-deepfake-app-movie-tv-show-face-replace-privacy-policy-concerns

26 **Bright maakte zijn eigen deepfake: zo werkt dat – Bright**
https://www.bright.nl/nieuws/artikel/4757251/bright-maakte-zijn-eigen-deepfake-zo-werkt-dat

27 **FaceApp - Free Neural Face Transformation Filters**
https://www.faceapp.com/

28 **DNC warns 2020 campaigns not to use FaceApp 'developed by Russians' – CNN**
https://edition.cnn.com/2019/07/17/politics/dnc-warning-faceapp/index.html

29 **Commentary: Deepfakes, the future of video manipulation and election hacking – CNA**
https://www.channelnewsasia.com/news/commentary/deep-fakes-the-future-of-election-hacking-10925604

30 **FaceApp removes 'Ethnicity Filters' after racism storm – Daily Mail Online**
https://www.dailymail.co.uk/sciencetech/article-4777954/FaceApp-removes-Ethnicity-Filters-racism-storm.html

31 **False news stories are 70% more likely to be retweeted on Twitter than true ones – MarketWatch**
https://www.marketwatch.com/story/fake-news-spreads-more-quickly-on-twitter-than-real-

32 **news-2018-03-08**
https://twitter.com/realDonaldTrump/status/1171046015106990081?s=20

33 **Deepfake detection algorithms will never be enough – The Verge**
https://www.theverge.com/2019/6/27/18715235/deepfake-detection-ai-algorithms-accuracy-will-they-ever-work

34 **Enterprises need to plan for deep fake technology**
https://www.laurashouse.org/PDF/6.25.19-search-cio.pdf

35 **What was /r/deepfakes and how did it influence the recent Reddit rule changes?**
https://www.reddit.com/r/OutOfTheLoop/comments/7vydoi/what_was_rdeepfakes_and_how_did_it_influence_the/

36 **Celebrities – AdultDeepFakes.com**
https://adultdeepfakes.com/celebrities/

37 **Dionne Stax duikt op in pornovideo**

https://www.shownieuws.nl/rubrieken/sterren/2019/dionne-stax-duikt-op-pornovideo/

38 **I Was The Victim Of A Deepfake Porn Plot Intended To Silence Me – HuffPost UK**
https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316

39 **Fake-porn videos are being weaponized to harass and humiliate women: 'Everybody is a potential target' - The Washington Post**
https://www.washingtonpost.com/technology/2018/12/30/fake-porn-videos-are-being-weaponized-harass-humiliate-women-everybody-is-potential-target/

40 **AI Powered X-Ray App | DeepNude**
https://www.deepnude.com/

41 **This Horrifying App Undresses a Photo of Any Woman With a Single Click - VICE**
https://www.vice.com/en_us/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman

42 **Copies of AI deepfake app DeepNude are easily accessible online — and always will be - The Verge**
https://www.theverge.com/2019/7/3/20680708/deepnude-ai-deepfake-app-copies-easily-accessible-available-online

43 **2018 Russia–United States summit – Wikipedia**
https://en.wikipedia.org/wiki/2018_Russia%E2%80%93United_States_summit

44 **Deepfakes zijn een groeiend gevaar voor onze democratie**
https://fd.nl/opinie/1305951/deepfakes-zijn-een-groeiend-gevaar-voor-onze-democratie

45 **Zo makkelijk is het om zelf een deepfake (of deepnude) te maken – Marketingfacts**
https://www.marketingfacts.nl/berichten/zo-makkelijk-is-het-om-zelf-een-deepfake-of-deepnude-te-maken

46 **Scammer Successfully Deepfaked CEO's Voice To Fool Underling Into Transferring $243,000**
https://gizmodo.com/scammer-successfully-deepfaked-ceos-voice-to-fool-under-1837835066

47 **Here's how algorithms can protect us against deepfakes**
https://thenextweb.com/syndication/2019/07/13/heres-how-algorithms-can-protect-us-against-deepfakes/

48 **Deepfakes, Blockchains, and Factom – Factomize**
https://factomize.com/deepfakes-blockchains-and-factom/

49 **Blockchain – Wikipedia**
https://en.wikipedia.org/wiki/Blockchain

50 **Truepic – Technology**
https://truepic.com/technology/

51 **Deep fakes: how immutable blockchain-based life logs could combat them, and the implications for privacy**
https://www.privateinternetaccess.com/blog/2019/01/deep-fakes-how-immutable-blockchain-based-life-logs-could-combat-them-and-the-implications-for-privacy/

52 **Deeptrace | The antivirus for deepfakes**
https://www.deeptracelabs.com/

53 **Deepfact |**
https://3duniversum.com/product/deepfact/

54 **Protecting_World_Leaders_Against_Deep_Fakes**
http://openaccess.thecvf.com/content_CVPRW_2019/papers/Media%20Forensics/Agarwal_Protecting_World_Leaders_Against_Deep_Fakes_CVPRW_2019_paper.pdf

55 **The Deepfake Detection Challenge**
https://deepfakedetectionchallenge.ai/

56 **Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security by Robert Chesney, Danielle Keats Citron – SSRN**

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954

57 **Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security by Robert Chesney, Danielle Keats Citron – SSRN –**
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954

58 **House Intelligence chief presses social media companies on deepfake policies – Reuters**
https://www.reuters.com/article/us-usa-election-deepfakes/house-intelligence-chief-presses-social-media-companies-on-deepfake-policies-idUSKCN1UA2GC

59 **US Congress holds hearing on "deepfakes" and artificial intelligence – YouTube**
https://www.youtube.com/watch?v=lArPEDS0GTA

60 **"Everyone Potential Target": Artificial Intelligence Weaponises Fake Porn**
https://www.ndtv.com/world-news/everyone-potential-target-artificial-intelligence-weaponises-fake-porn-1970275

61 **Deepfake detection algorithms will never be enough – The Verge**
https://www.theverge.com/2019/6/27/18715235/deepfake-detection-ai-algorithms-accuracy-will-they-ever-work

62 **Prepare, don't panic: dealing with deepfakes and other synthetic media – YouTube**
https://www.youtube.com/watch?v=ZWKI1h5SYTI

63 **How to recognize fake AI-generated images - Kyle McDonald – Medium**
https://medium.com/@kcimc/how-to-recognize-fake-ai-generated-images-4d1f6f9a2842

64 **How to spot deepfake videos – and why you should care**
https://www.avanade.com/en/blogs/avanade-insights/artificial-intelligence/how-to-spot-deepfake-videos

65 **How The Wall Street Journal is preparing its journalists to detect deepfakes – Nieman Journalism Lab**
https://www.niemanlab.org/2018/11/how-the-wall-street-journal-is-preparing-its-journalists-to-detect-deepfakes/

66 **This Video May Not Be Real – The New York Times**
https://www.nytimes.com/2019/08/14/opinion/deepfakes-adele-disinformation.html?smid=nytcore-ios-share

67 **Malicious Deep Fake Prohibition Act of 2018 – Congress.gov – Library of Congress**
https://www.congress.gov/bill/115th-congress/senate-bill/3805/text?format=txt

68 **Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019 – Congress.gov – Library of Congress**
https://www.congress.gov/bill/116th-congress/house-bill/3230

69 **Virginia updates its revenge porn law to include deepfakes**
https://www.engadget.com/2019/07/02/virginia-deepfake-revenge-porn/

70 **Report: 2020 Candidates Are Going to Get Owned by Deepfake**
https://futurism.com/the-byte/2020-candidates-deepfakes

71 **Kabinet komt met campagne tegen nepnieuws in verkiezingstijd – NOS**
https://nos.nl/artikel/2263298-kabinet-komt-met-campagne-tegen-nepnieuws-in-verkiezingstijd.html

72 **Kamerbrief over dreiging desinformatie en beïnvloeding verkiezingen – Kamerstuk – Rijksoverheid.nl**
https://www.rijksoverheid.nl/documenten/kamerstukken/2018/12/13/kamerbrief-over-dreiging-desinformatie-en-beinvloeding-verkiezingen

73 **Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security by Robert Chesney, Danielle Keats Citron – SSRN**
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954