JARNO DUURSMA

MACHINES MET VERBEELDINGSKRACHT

EEN KUNSTMATIGE REALITEIT

studio OVERMORGEN

MACHINES MET VERBEELDINGSKRACHT

EEN KUNSTMATIGE REALITEIT

Jarno Duursma

Trendwatchter | Auteur | TEDx spreker

studio OVERMORGEN

INHOUDSOPGAVE

MACHINES MET VERBEELDINGS-KRACHT

EEN KUNSTMATIGE REALITEIT

SAMENVATTING

Kunstmatig intelligente systemen hebben de afgelopen decennia een sterke ontwikkeling doorgemaakt, door zowel alsmaar meer computerkracht als door steeds meer verfijnde algoritmen. In de afgelopen jaren neemt daarbij de kwaliteit van verschillende, door AI gedreven generatieve software toe. Dat zijn verschillende soorten kunstmatig intelligente systemen die zelf content kunnen genereren zoals videobeelden, stemmen, foto's en teksten. In dit rapport heeft de auteur de onderling verschillende systemen samengevoegd onder de noemer *generative AI-software*. Specifieke aandacht in het eerste gedeelte van het rapport krijgen de *generative adversarial networks* of GAN's, systemen van twee netwerken die na training met een verzameling data in onderlinge competitie nieuwe data genereren van een verrassend hoge kwaliteit. Die technologie maakt grote stappen vooruit *en geeft machines een vonk van verbeeldingskracht*.

Dit rapport, te beschouwen als een vervolg op het boek *De digitale butler – Kansen en bedreigingen van kunstmatige intelligentie*, dat de auteur in 2017 publiceerde, schetst de betekenis van deze nieuwe technologie en de bredere trend van 'door machines gegenereerde content'.

In het eerste deel van dit rapport toont de auteur een overzicht van de vele prachtige ontwikkelingen, mogelijke toepassingen van de technologie en trends die we kunnen verwachten op dat gebied. *Generative AI-software als ideeënmachine.*

Zoals elke technologie kunnen deze, door AI gedreven generatieve technieken, ook worden misbruikt. In het tweede deel van dit rapport beschrijft de auteur het verschijnsel deepfake, de inzet van *generative AI technology* voor het creëren van nepinformatie. Het overzicht laat zien hoe kunstmatig gecreëerde tekst, beeld, video en audio kunnen worden ingezet 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. Dezelfde technologie laat zich mogelijk op grote schaal inzetten om in voorkomende gevallen die vraag, "is het echt of is het nep", te beantwoorden. Daarmee ontstaat er een digitale wedloop achter de schermen van onze hyperverbonden wereld, waarbij generatieve AI-systemen ons kunnen misleiden en ons kunnen beschermen.
In negen interviews, die u verspreid over het rapport kunt vinden, laten professionals in verscheidene vakgebieden hun licht schijnen op de kansen en de bedreigingen die 'generatieve AI-technologie' oplevert.

DANKWOORD

Mijn dank gaat uit naar:

Jelmer Frank Wijnia (Van Wijnen),

Joost Schellevis (NOS),

Siri Beerends (SETUP),

Mark Wiebes (Politie),

Lodewijk van Zwieten (OM),

Henry Ajder (DeepTrace),

Frank Smilda (Politie)

voor hun inhoudelijke bijdrage.

En:

Berco Beute (Media2B),

Bennie Mols (wetenschapsjournalist),

Marcel van Gerven (Radboud Universiteit),

Thomas Derksen (gw20e.com),

Sander Duivestein (Sogeti),

Floor Duursma,

Gert Gritter,

Martijn Striker,

Erick Vermeulen

voor het meedenken, bijdragen, meelezen, becommentariëren, redigeren of opmaken.

INTRODUCTIE

SOFTWARE IS EATING HUMANS

In augustus 2011 deed Marc Andreessen van investeringsmaatschappij A16Z de beroemde uitspraak *"software is eating the world"*. Hij doelde daarmee op het feit dat steeds meer producten en diensten in de fysieke wereld worden vervangen door software. De boekenwinkel werd Amazon, cd's werden vervangen door Napster en later Spotify. Vrijwel alle producten veranderen in softwarediensten. De uitspraak van Andreessen is in 2019 nog steeds erg actueel. Sterker nog, door de kwalitatieve groeispurt in het domein van kunstmatige intelligentie zien we een nóg grotere versnelling ervan. En we zien bijvoorbeeld ook dat AI-machines goede voorspellingen kunnen doen en dat ze steeds beter snappen wie wij zijn, wat we doen en waarom we dat doen. We zien zelfs de sterke ontwikkeling van specifieke en geavanceerdere systemen die op basis van data kunnen raden hoe wij ons voelen vanbinnen [2].

Het rapport dat u voor zich heeft liggen, is te beschouwen als een opvolger van mijn boek uit 2017, De digitale butler – Kansen en bedreigingen van kunstmatige intelligentie.[3] Daarin beschrijf ik hoe AI-systemen steeds meer menselijke vaardigheden van ons overnemen zoals kijken, luisteren, spreken en lezen. Met dit rapport voeg ik daar een belangrijke menselijke vaardigheid aan toe: *het kunnen toepassen van verbeeldingskracht*. Software eet de wereld op. En *"Software is eating humans"*.

Door een kwalitatieve groeispurt binnen kunstmatige intelligentie, *machine learning*, generatieve software in het algemeen en GAN-technologie in het bijzonder kunnen machines innovatieve variaties maken op bestaande data. Een wereld vol met nieuwe invalshoeken, nieuwe ideeën en afgeleiden ontstaat hiermee. Machines met GAN-technologie kunnen in het meest ideale geval zorgen voor nieuwe, verfrissende ideeën en hypotheses en op deze wijze innovatieve processen enorm versnellen.

Het eerste deel van dit rapport, 'Machines met verbeeldingskracht', gaat over de kwalitatieve opkomst van dit soort generatieve software en GAN-systemen in het bijzonder. Dit soort software is een gereedschapskist om onze creatieve processen te assisteren en soms zelfs van ons over te nemen. In het eerste deel beschrijf ik ook een aantal mogelijke toekomstige scenario's waar deze technologie ingezet gaat worden. Want wat laten we in de toekomst aan machines over en wat blijven we zelf doen?

Maar zoals bij iedere technologische ontwikkeling heeft ook deze trend een negatieve keerzijde. Daarover gaat het tweede deel van het rapport, *Deepfake-technologie: The Infocalypse*. Bij deepfake-technologie [4] wordt generatieve software en de 'machine met verbeeldingskracht' gebruikt om onze waarneming van de wereld op een negatieve manier te beïnvloeden. Om daarmee meningen te manipuleren,

mensen te chanteren of reputatieschade toe te brengen. We zien namelijk dat het met behulp van generatieve AI-software steeds gemakkelijker wordt om iemands gezicht [5] of stem te vervalsen. Om een video- of audio opname te maken van iets wat een persoon niet gedaan of gezegd heeft. Om onderling gezichten te verwisselen in een video. We naderen daarmee een periode waarin we online onze ogen en oren niet meer kunnen vertrouwen. [6] Hoe gaan we daar mee om?

Dit rapport heeft ook als doel om deze deepfake-technologie te ontleden en te duiden. Ik vind het vanuit mijn rol als trendwatcher, onderzoeker en duider van alles wat zich afspeelt in de digitale frontlinie, namelijk mijn plicht om daarover te berichten. [7] [8] [9] [10]

Verdeeld over dit rapport vindt verdere verdieping plaats over de onderwerpen generatieve AI-software, GAN-technologie, deepfake-technologie en synthetische media (met AI gecreëerde of gemodificeerde media). Professionals vanuit verschillende beroepsgroepen laten hun deskundige licht schijnen over deze onderwerpen. Hoe beïnvloedt deepfake-technologie de journalistiek? Hoe gaat het Openbaar Ministerie om met deze technologische trend? Hoe verandert GAN-technologie de game-industrie? Wat betekent het voor de samenleving wanneer echt en nep niet meer van elkaar te onderscheiden zijn? Hoe helpt GAN-technologie als creatieve assistent bij het ontwerpen van stedenbouwkundige plannen? Is de politie voorbereid op voice cloning-technologie? Professionals vertellen over hun visie, hun ideeën, hun zorgen en richten tevens hun blik op de toekomst.

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.

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.

Gegevens Jarno Duursma
Website: jarnoduursma.nl
E-mail: info@jarnoduursma.nl
Twitter: twitter.com/jarnoduursma
LinkedIn: linkedin.com/in/jarnoduursma

Ook ben ik vanzelfsprekend beschikbaar als spreker over dit onderwerp.

MACHINES MET VERBEELDINGS-KRACHT

MACHINES MET VERBEELDINGSKRACHT

Een opvallende bewering: machines met verbeeldingskracht. Dat roept gelijk als vraag op: wat is dat eigenlijk, verbeeldingskracht? Ontelbaar veel filosofen, creatievelingen en wetenschappers hebben zich in onze geschiedenis al gebogen over deze vraag.

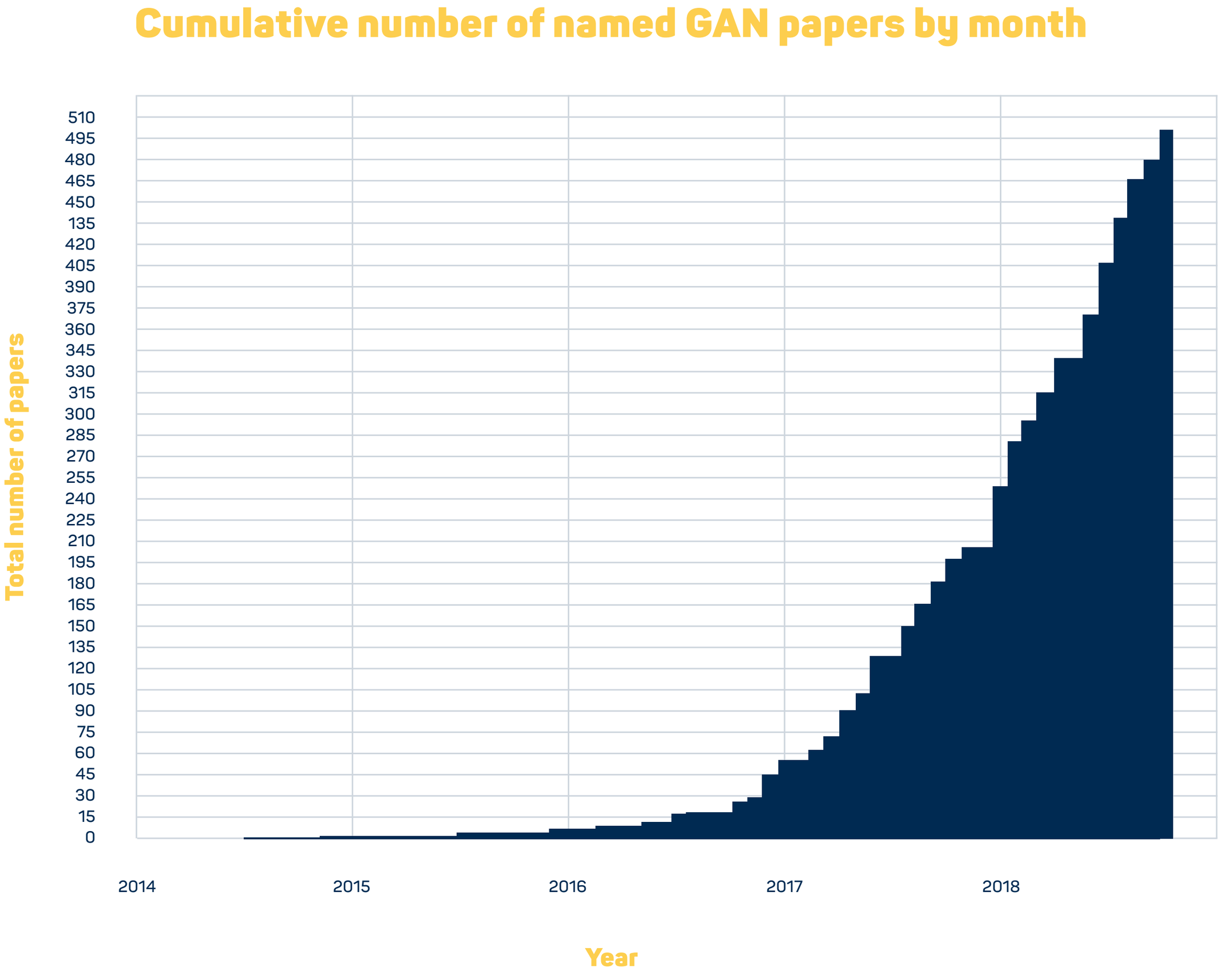
Ik heb het gevraagd aan meerdere mensen in mijn omgeving en daarbij merkte ik dat er een veelheid aan invalshoeken bestaat. “Verbeeldingskracht is iets dat te maken heeft met dat je je kunt voorstellen dat iets in de toekomst gebeurt.” “Een uitkomst van iets bedenken zonder dat je het hoeft uit te proberen.” “Je iets kunnen voorstellen wat er nog niet is.” “Het zit in hetzelfde domein als creativiteit, fantasie en mijmeren: daar waar je hersenen je soms naartoe brengen als je ergens aan denkt.”

Wikipedia heeft een prachtige omschrijving: “De menselijke verbeeldingskracht of fantasie is het vermogen om mentale beelden, ideeën en/of gevoelens op te roepen, zonder dat men deze zintuiglijk waarneemt. (...) De verbeelding is de basis voor inspiratie en nieuwe ideeën en speelt een belangrijke rol in het leervermogen van de mens. Verbeeldingskracht [11] kan op die manier dus worden gezien als de basis van innovatie en ontwikkeling.”

Ik ga niet proberen de alomvattende definitie van verbeeldingskracht te vinden. Dat laat ik liever aan anderen over. Wat voor mij echter wel helder is geworden, is dat met de opkomst van generatieve AI-technologie in het algemeen en de *generative adversarial networks* (GAN’s) in het bijzonder, de menselijke verbeeldingskracht bijzonder sterk kan worden aangevuld door computertechnologie. Kunstmatig intelligente systemen bedenken nieuwe invalshoeken en nieuwe afgeleiden van bestaande concepten en leiden daarbij soms zelfs tot ronduit nieuwe ideeën.

“Een uitkomst van iets bedenken zonder dat je het hoeft uit te proberen. Je iets kunnen voorstellen wat er nog niet is.”

Yann LeCun – de directeur van FAIR, het onderzoeksinstituut voor kunstmatige intelligentie van Facebook (Facebook Artificial Intelligence Research), en tevens hoogleraar aan de Universiteit van New York – noemt GAN-technologie bijvoorbeeld 'het meest interessante idee in de afgelopen tien jaar in *machine learning*'. [12] In een later interview noemt hij het 'het coolste idee in *machine learning* in de laatste tien of twintig jaar'. [13] Voor een serieus en gerenommeerd wetenschapper in het vakgebied van machinaal leren is dat nogal een uitspraak. Wanneer we kijken naar de cijfers, zien we ook dat het aantal rapporten over deze GAN-technologie in de afgelopen jaren enorm is gestegen.



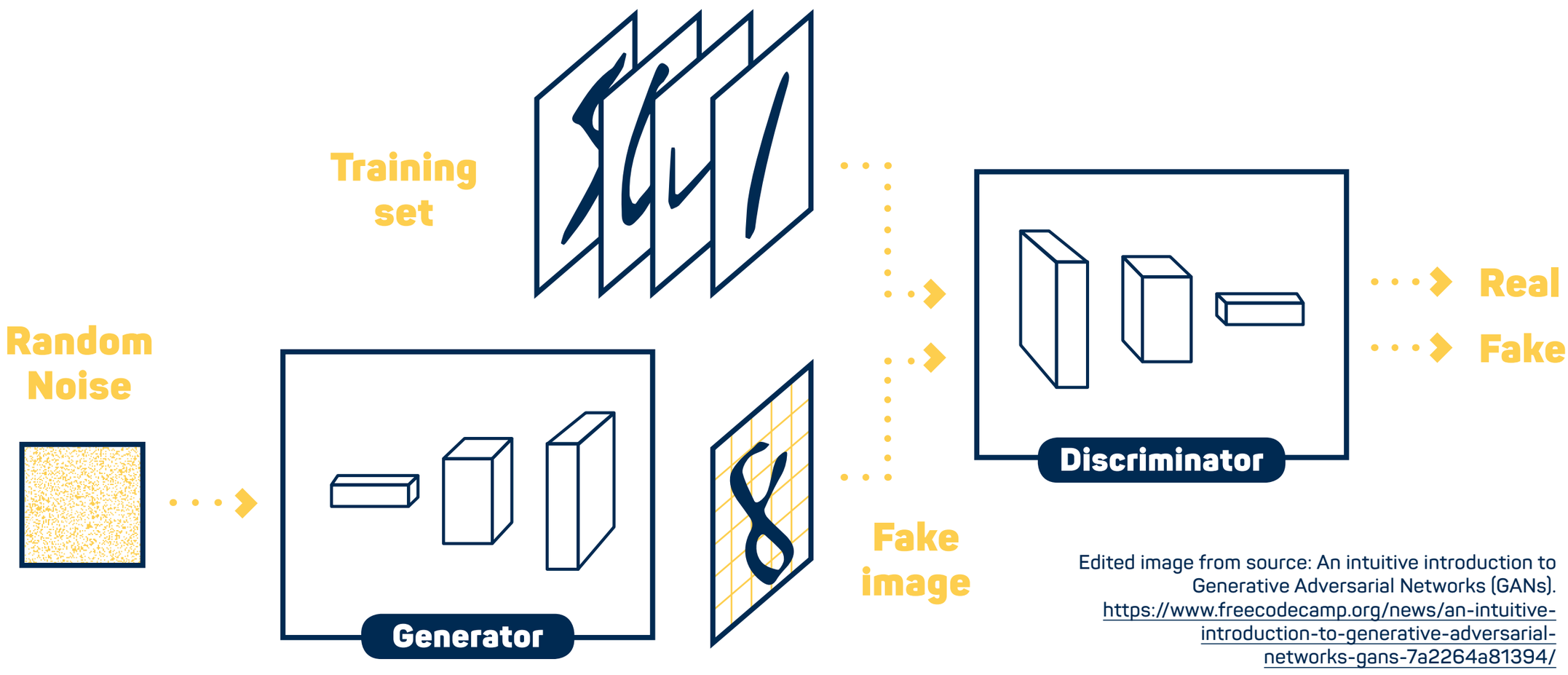
Edited image from source: Bruno Gavranović GitHub.
https://github.com/bgavran

1.1 | DE STRIJD IN GAN-NETWERKEN

GAN-technologie [14] verdient een beknopte uitleg, voor een goed begrip van de toepassingen en om de vele voorbeelden in dit rapport te kunnen duiden. Het is een belangrijke en veelbelovende ontwikkeling binnen de grotere trend van generatieve AI-software, dat meer als een verzamelnaam kan worden beschouwd van allerhande AI-softwaresystemen die uiteenlopende soorten content kunnen genereren. Kort door de bocht kan gesteld worden dat generatieve AI-software synthetische media (met AI gegenereerde, gecreëerde of gemodificeerde media) fabriceert, waarvan bijvoorbeeld deepfake media de negatieve uitkomst is. Maar allereerst kijken we dus naar GAN systemen.

Generative adversarial networks

De technologie van *generative adversarial networks* [15] is een paar jaar geleden ontwikkeld door een team onder leiding van Ian Goodfellow, die nu een onderzoeker is bij Apple. Wanneer we het begrip *generative adversarial networks* ontleden, zien we drie woorden die iets vertellen over de onderliggende technologie. *Generative* verklaart dat deze kunstmatig intelligente systemen content kunnen genereren. *Adversarial* beschrijft de strijd tussen twee verschillende neurale netwerken (waarmee tegelijkertijd ook het derde woord is verklaard, *networks*), om de kwaliteit van de output te verbeteren. Het woord *adversarial*, tegenstrevend, heeft al gauw een wat vijandige bijsmaak, en in deze context kunt u het beter beschouwen als een 'samenwerkingsverband' tussen deze twee netwerken. Hoe gaat dat? [16 | 17]



Edited image from source: An intuitive introduction to Generative Adversarial Networks (GANs). https://www.freecodecamp.org/news/an-intuitive-introduction-to-generative-adversarial-networks-gans-7a2264a81394/

Een GAN-systeem bestaat in feite uit twee concurrerende neurale netwerken die vanuit hun onderlinge samenwerking realistische output genereren. Zij strijden met elkaar om samen steeds betere resultaten te bereiken, bijvoorbeeld realistische foto's te genereren van niet bestaande hamburgers. Dit soort GAN systemen zijn bijzonder goed in het kunnen genereren van visuele output, vandaar ook het voorbeeld van te genereren foto's.

Het eerste neurale netwerk, het *discriminative network* ofwel de discriminator, wordt gevoed met een grote dataset met trainingsgegevens, in dit voorbeeld foto's van hamburgers. Het netwerk leert uit die verzameling foto's wat een hamburgerfoto definieert.
Het andere netwerk, het *generative network*, gaat tegelijkertijd aan het werk met de discriminator en gaat output produceren, het is de generator. Dat netwerk probeert output te creëren waarvan de discriminator dénkt dat die thuishoort in de oorspronkelijke dataset, maar die nieuw gecreëerd is door de generator. Het generatieve systeem 'liegt' dus tegen de discriminator door nieuwe input te maken die niet behoort tot de dataset, maar die wel aan de criteria voldoet. De discriminator probeert dat bedrog te ontdekken.

In dit voorbeeld probeert de generator dus een foto te maken van een hamburger.
Het generatieve systeem [18] begint hierbij met ruis, *random noise*; het heeft als het ware geen idee waar het moet beginnen. In het voorbeeld van de hamburgerfoto's begint de generator daarom met een wirwar van pixels.

De discriminator heeft vervolgens de rol van beoordelaar van de output, als een soort scheidsrechter. Dat netwerk bepaalt of de output die is gecreëerd door het generatieve netwerk, realistisch genoeg is in vergelijking met de oorspronkelijke dataset van trainingsvoorbeelden. De discriminator bepaalt of in de gegenereerde hamburgerfoto de hamburger kan doorgaan voor 'echte output' of 'nep output'. De discriminator toetst zo de kwaliteit van de output door die te vergelijken met de oorspronkelijke input. Het discriminator systeem wijst in het begin heel veel output af, omdat de generator immers begint met *random noise*.

Deze beide neurale netwerken zijn in constante wisselwerking met elkaar. Er is een voortdurende terugkoppeling, een feedback die als rivaliteit kan worden beschouwd. De discriminator dwingt de generator alsmaar tot het leveren van een nóg hogere kwaliteit output. De generator doet daarom zijn uiterste best om het alsmaar nóg beter te doen. De wisselwerking tussen de twee netwerken, de generator en de discriminator, is als de tweestrijd tussen een kunstdetective en kunstvervalser, waarbij die laatste voortdurend probeert de detective te slim af te zijn. Door de continue feedback zal de generator aan het eind een foto genereren van een hamburger die door de discriminator wordt beoordeeld als 'echt bestaand', maar dus in feite 'gegenereerd' is.

Credit: Advanced Machine Learning (BigGAN) Generating images of a cheeseburger.
https://www.reddit.com/r/pics/comments/9y112m/advanced_machine_learning_biggan_generating/

Op deze wijze ontstaat er een ingenieus systeem dat uiteindelijk realistische voorbeelden kan maken; voorbeelden waarbij het resultaat net zo goed is als de originelen uit de database, maar die niet tot de voorbeelden uit de dataset behoren! Het systeem creëert zijn eigen invulling. Het laat nieuwe ideeën zien van bestaande voorbeelden. De nieuwe voorbeelden zijn geen letterlijke kopieën van voorbeelden in de originele dataset, maar volkomen nieuwe variaties. Het zijn foto's gecreëerd uit de 'verbeeldingskracht' van de generator. Het GAN systeem kan op deze wijze worden beschouwd als ideeënmachine.

Met deze laatste zinnen wordt het vernieuwende van deze GAN-technologie ook duidelijk. Een generator creëert output die afkomstig zou kúnnen zijn van de dataset van de originele input, maar dat dus niet is. Op deze wijze kunnen op basis van bestaande data gemakkelijk nieuwe afgeleiden, variaties, en invalshoeken worden geschapen. Dat is mijns inziens ook wat wij als mensen doen wanneer wij onze verbeeldingskracht aanspreken. Wij bedenken nieuwe beelden op basis van beelden die we kennen. Met GAN technologie krijgen machines ook een vonk van onze menselijke verbeeldingskracht.

Vraag een GAN-systeem om een afbeelding te maken van 'een vogel met rode veren, een zwarte kroon en een kleine snavel', (daarover later in het rapport meer) dan creëert het vele variaties. Het is alsof aan honderd mensen wordt verzocht om allemaal een tekening te maken van een vogel die voldoet aan die beschrijving. En hoewel er gelijkenissen zullen zijn, zullen er ook veel verschillen in de interpretaties zijn. Dat is juist de kracht van GAN-technologie. Daarbij dient wel vermeld te worden dat een GAN systeem niet volledig out-of-the-box kan creëren. Wanneer een systeem is getraind op het creëren van foto's van auto's, zal het nooit een koelkast maken. Wanneer het GAN-systeem getraind is met foto's van bestaande mensen, maar waarvan hun oren bedekt zijn, kan het systeem niet uit zichzelf de oren erbij 'bedenken'. De machinale verbeeldingskracht is op dit moment dus nog niet zo uitgebreid als de menselijke.

En belangrijk om te noemen: deze ontwikkeling bevindt zich weliswaar in een pril stadium, maar de computer heeft een aantal voordelen boven ons mensen. Die werkt over het algemeen sneller, vierentwintig uur per dag en zeven dagen per week onvermoeibaar door. Dat is op geen enkele wijze te vergelijken met hoe wij als mensen werken. Wij zouden eerst letterlijk naar een foto moeten kijken om deze te kunnen beoordelen, bij een GAN-systeem gaat dat natuurlijk op computersnelheid.

Edited image from source: These AI-Generated Cars Will Melt Your Brain | Gizmodo Australia. https://www.gizmodo.com.au/2018/12/these-ai-generated-cars-will-melt-your-brain/

GAN-systemen zijn vooral goed in het bedenken van nieuwe invalshoeken wanneer het gaat om visuele data. Andere generatieve AI-software is weer beter in het creëren van teksten of audio (maar daarover later meer). Welbekende GAN-voorbeelden van variaties zijn de gezichten van mensen die nooit hebben bestaan [19], fotomodellen die nooit hebben bestaan [20], slaapkamers die nooit hebben bestaan, auto's die nooit hebben bestaan [21] enzovoort. Nogmaals: de generator maakt daarbij nieuwe variaties op een

bestaande dataset die zo soms zo goed zijn dat ze voor 'origineel' zouden kunnen doorgaan binnen de dataset. *Maar dat zijn ze niet!* Je zou ze ook kunnen beschouwen als nieuwe ideeën.
Door de kwalitatieve groei van generatieve AI-software en GAN-technologie in het bijzonder vervaagt in de loop van de tijd dus steeds meer de grens tussen reële en door machine gegenereerde content

Wanneer u zich bedenkt dat veel zaken in onze fysieke wereld zich digitaal laten weergeven en dat generatieve AI-systemen in de toekomst daar wellicht nieuwe variaties, nieuwe invalshoeken op kunnen bedenken, dan ontvouwen zich als vanzelf de nieuwe mogelijkheden [22]. Mogelijkheden die onze verbeeldingskracht stimuleren. Die hebben niet alleen betrekking op foto, video en audio, maar misschien ook op nieuwe medicijnen, nieuwe materialen, nieuwe smaakvariaties voor frisdranken of voorspellingen van het verloop van een ziekte.

Versnellen van creatieve processen

Het is niet moeilijk om enthousiast te zijn over dit nieuwe landschap van mogelijkheden. Wij zullen als mensen op het gebied van creativiteit en innovatie steeds meer de interactie aangaan met kunstmatig intelligente machines [23]. Op sommige vlakken zullen zij het creatieve proces enorm versnellen door met grote snelheid nieuwe ideeën te genereren [24]. Sommige ideeën lijken wellicht onnozel of zelfs hallucinogeen, maar soms zijn ze ook verrassend vernieuwend en inspirerend.

Machines met verbeeldingskracht waren jarenlang onderwerp van sciencefictionboeken en Hollywood-films. Maar de deksel is van de doos, de geest is uit de fles. Wereldwijd groeit de aandacht voor generatieve AI-software en GAN-systemen in het bijzonder. Wetenschappers, bedrijven, technologie-enthousiastelingen, designers, kunstenaars en technologiefilosofen kijken met bijzonder veel interesse naar de mogelijkheden en onmogelijkheden van deze technologische trend.

Sommige GAN-technologie genereert bijvoorbeeld voor wetenschappers ideeën en invalshoeken en haalt verschillende opties daarmee in tijd naar voren: kunstmatig intelligente GAN-systemen fungeren daarbij als een creatieve assistent. Een GAN-systeem is dan een "hypothese creërende machine". [25]

Om zoveel mogelijk concreet te maken wat een GAN-systeem kan, maak ik gebruik van velerlei voorbeelden. Allereerst behandel ik de mogelijkheid van dit soort systemen om synthetische visuele beelden te creëren. Daar zijn GAN-systemen immers heel goed in. Vervolgens richt ik mij in het rapport op andere generatieve AI-systemen, waaronder systemen die teksten kunnen genereren of audio zoals stemgeluid [26] en muziek [27].

" Wij zullen als mensen op het gebied van creativiteit en innovatie steeds meer de interactie aangaan met kunstmatig intelligente machines."

1.2 | SYNTHETISCHE BEELDEN

Ik denk dat een van de allerbekendste voorbeelden van de opkomst van GAN-technologie het moment betreft toen onderzoekers van NVIDIA een demonstratie gaven van hun StyleGAN-systeem [28]. Dat systeem kon foto's creëren van mensen die niet daadwerkelijk in de fysieke wereld bestaan. De demonstratie gaf veel mensen voor wie de technologie tot dan toe onzichtbaar was, een inkijkje in de mogelijkheden.

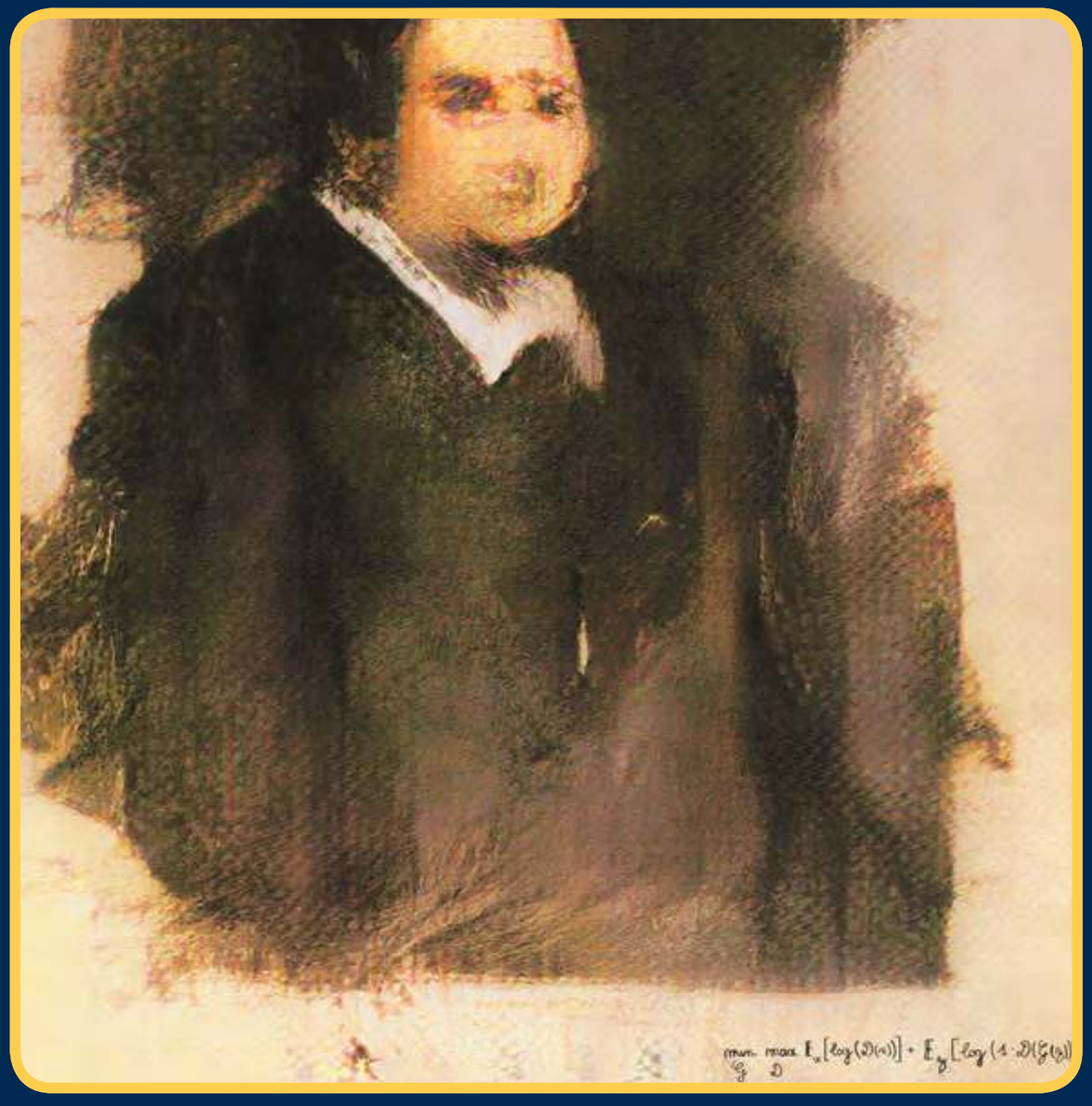
Edited image from source: This AI Creates Photo-Realistic Faces of People Who Don't Exist.
https://petapixel.com/2017/11/07/ai-creates-photo-realistic-faces-people-dont-exist/

Deze GAN-software bestond overigens al langer, maar sinds vorig jaar creëert deze software een veel gedetailleerder en geloofwaardiger resultaat. In een handomdraai levert die synthetische mensen, niet meer van echt te onderscheiden. De variaties die het systeem genereert als het bijvoorbeeld aankomt op haardracht, huidskleur, sproeten en bril is bijzonder geloofwaardig. De video *"A Style-Based Generator Architecture for Generative Adversarial Networks"* [29] laat zien dat er ook variaties worden aangebracht in het resultaat door het systeem te beïnvloeden met andere parameters of nieuwe input.
Het StyleGAN-systeem laat zien wat al vaker genoemd is in dit rapport. Het systeem creëert realistische, maar volledig nieuwe afbeeldingen, invalshoeken, ideeën op bestaande data en daarmee krijgt slimme AI-software een vonk van verbeelding.

Al snel na de release van het StyleGAN-systeem dook de website thispersondoesnotexist.com [30] op om in te springen op de belangstelling voor deze software. Die website gebruikt ook de StyleGAN-software van NVIDIA. Op de website van het bedrijf 3DUniversum staat de Deepfact Quiz [31], waarmee u kunt testen hoe goed u bij gezichten echt van nep kunt onderscheiden.

Putting the ‘art’ back into Artificial Intelligence

Een ander bekend voorbeeld van GAN-technologie betreft het schilderij ‘Edmond de Belamy’ [32]. Dat schilderij uit 2018 werd gemaakt door een GAN-netwerk [33] ingezet door het Franse kunstcollectief Obvious.



Credit: Artist: Obvious. Edmond de Belamy - Wikipedia. https://en.wikipedia.org/wiki/Edmond_de_Belamy

Het GAN-netwerk werd getraind met 15.000 voorbeeldportretten van oude schilderijen, waarna het niet bestaande nieuwe afbeeldingen kon creëren. Een van die afbeeldingen, Edmond de Belamy, werd vervolgens op canvas geprint en in New York geveild door Christie’s. De inschatting vooraf was dat het schilderij zou worden verkocht voor $ 10.000, maar de biedingen liepen op tot uiteindelijk $ 432.500 [34]. Een recordbedrag.

De verkoop van dat kunstwerk heeft een interessante discussie [35] doen oplaaien over eigenaarschap en het artistieke productieproces. Kunstcollectief Obvious gebruikte bij het maken van het schilderij namelijk bestaande GAN-software en het bijbehorende onderzoek van andere mensen. Het collectief heeft zelf weinig tijd en moeite gestoken in de technologische ontwikkeling van de GAN-software. Wel streek het de volledige winst van het schilderij op.

CycleGAN image translation

Om nog even in de artistieke hoek te blijven: de CycleGAN-architectuur [36 | 37] is een voorbeeld van een *image translation*-softwaresysteem. CycleGAN kan bijvoorbeeld alledaagse foto's transformeren naar de artistieke stijl van schilders als Claude Monet, Vincent van Gogh en Paul Cézanne. Het softwaresysteem heeft geleerd wat de specifieke stijl van deze schilders is en past die stijl toe op een echte foto. Van uw vakantiefoto kunt u nu dus een Van Gogh-interpretatie maken.

In het onderdeel over deepfake-technologie zal ik hier veelvuldig op terugkomen, want deze GAN-systemen zorgen ervoor dat wij online onze ogen niet meer kunnen vertrouwen. We weten niet meer of wat we zien is gegenereerd door een computer of daadwerkelijk heeft plaatsgevonden in de realiteit. De vraag of de software van CycleGAN op dit moment feilloos nieuwe beelden kan creëren, is wat mij betreft niet zo relevant. Het is slechts een kwestie van tijd totdat dit soort systemen resultaten leveren die niet meer van echt te onderscheiden zijn.

Edited image from source: CycleGAN Project Page.
https://junyanz.github.io/CycleGAN/

Ook kan het systeem foto's die zijn gemaakt in de zomer transformeren naar foto's in het winterseizoen. En andersom. Het heeft geleerd van digitale input en is vervolgens zelf in staat nieuwe output te genereren. Het CycleGAN-systeem maakt zo een 'vertaling' van een bestaande foto van het zomerseizoen naar een sneeuwlandschap. Ook kan het bijvoorbeeld in video's paarden transformeren naar zebra's en omgekeerd.

Deze voorbeelden demonstreren de mogelijkheid van dit soort systemen om visuele content te manipuleren, te verdraaien.

Een ander voorbeeld van image translation betreft BicycleGAN [38] , een systeem dat een lijntekening kan transformeren naar een foto van een daadwerkelijk fysiek object. Dat systeem tovert bijvoorbeeld een lijntekening van een handtas om naar een geloofwaardige foto van een handtas. Het kan tevens binnen de gestelde kaders (de lijnen) enorm veel variaties aanbrengen in kleur en textuur. Datzelfde geldt bijvoorbeeld voor schoenen. Met relatief weinig moeite creëert BicycleGan vele variaties van een bestaand schoenenmodel.

Een video [39] vervaardigd door een team van onderzoekers aan Berkeley en het Adobe Creative Intelligence Lab geeft een goed idee van de mogelijkheden die GAN-systemen bieden aan ontwerpers. Het is heel gemakkelijk om het genereren van nieuwe kleuren en texturen door een kunstmatig intelligent systeem te laten uitvoeren. Dat zullen we in de toekomst veel vaker zien: mee-fantaserende machines. Die samenwerking beperkt zich niet tot

Edited image from source: BicycleGAN.
https://prostheticknowledge.tumblr.com/post/168187711191/bicyclegan-machine-learning-research-from-berkeley

kunstenaars en ontwerpers, want GAN-software kan in interactie met de mens nieuwe invalshoeken en nieuwe afgeleiden bedenken binnen een grote verscheidenheid aan bestaande thema's. Dat zal een boost zijn voor innovatie en creativiteit. Dat merkte ook Marcel van Gerven, Hoogleraar artificiële intelligentie. Donders Institute for Brain, Cognition and Behaviour, in Nijmegen. Hij gebruikt GAN-technologie in hersenonderzoek.

INTERVIEW: MARCEL VAN GERVEN

Kan je me iets meer vertellen over je onderzoek?
Mijn interesse bevindt zich op het snijvlak van de natuurlijke intelligentie en de artificiële intelligentie. Ik hou me met name bezig met de vraag hoe het brein problemen oplost en of we van deze mechanismen gebruik kunnen maken bij de ontwikkeling van slimmere machines. Hiernaast hou ik me bezig met de nieuwste ontwikkelingen op het gebied van artificiële intelligentie, in het bijzonder de machine learning, zoals GAN-technologie, om daarmee toepassingen te realiseren die het welzijn van mensen kunnen verbeteren.

Je houdt je bijvoorbeeld bezig met het combineren van neurotechnologie en artificiële intelligentie.
Wat onderzoek je dan?
Ik onderzoek hoe we mensen met specifieke hersenaandoeningen kunnen helpen door middel van brein-computer interactie, waarbij we informatie tussen hersenen en computers kunnen uitwisselen. Door informatie uit het brein uit te lezen kunnen we misschien ooit weer communiceren met mensen die dit vermogen zijn verloren, zoals bij volledige verlamming. Door informatie naar het brein te sturen kunnen we mogelijk ooit blinde mensen weer een stukje visuele waarneming teruggeven.

Kun je een voorbeeld geven van de rol van GAN-technologie in dit domein?
Zeker. Een voorbeeld van een onderzoek is dat we een GAN-systeem beelden laten genereren op basis van hersenactiviteit. Dat gaat zo: proefpersonen worden in een MRI-scanner gelegd die hun hersenactiviteit meet terwijl ze foto's te zien krijgen van gezichten. Met deze data hebben we vervolgens een GAN-systeem getraind dat leert om specifieke hersenactiviteit aan specifieke gezichten te koppelen. Vervolgens gebruiken we het GAN-systeem om beelden te creëren op basis van nieuwe hersendata. Het GAN-systeem probeert dan te visualiseren wat de proefpersonen daadwerkelijk hebben gezien. Dit werkt eigenlijk best al wel goed: De identiteit van de afgebeelde persoon is in de GAN-visualisatie bijvoorbeeld redelijk goed waar te nemen.

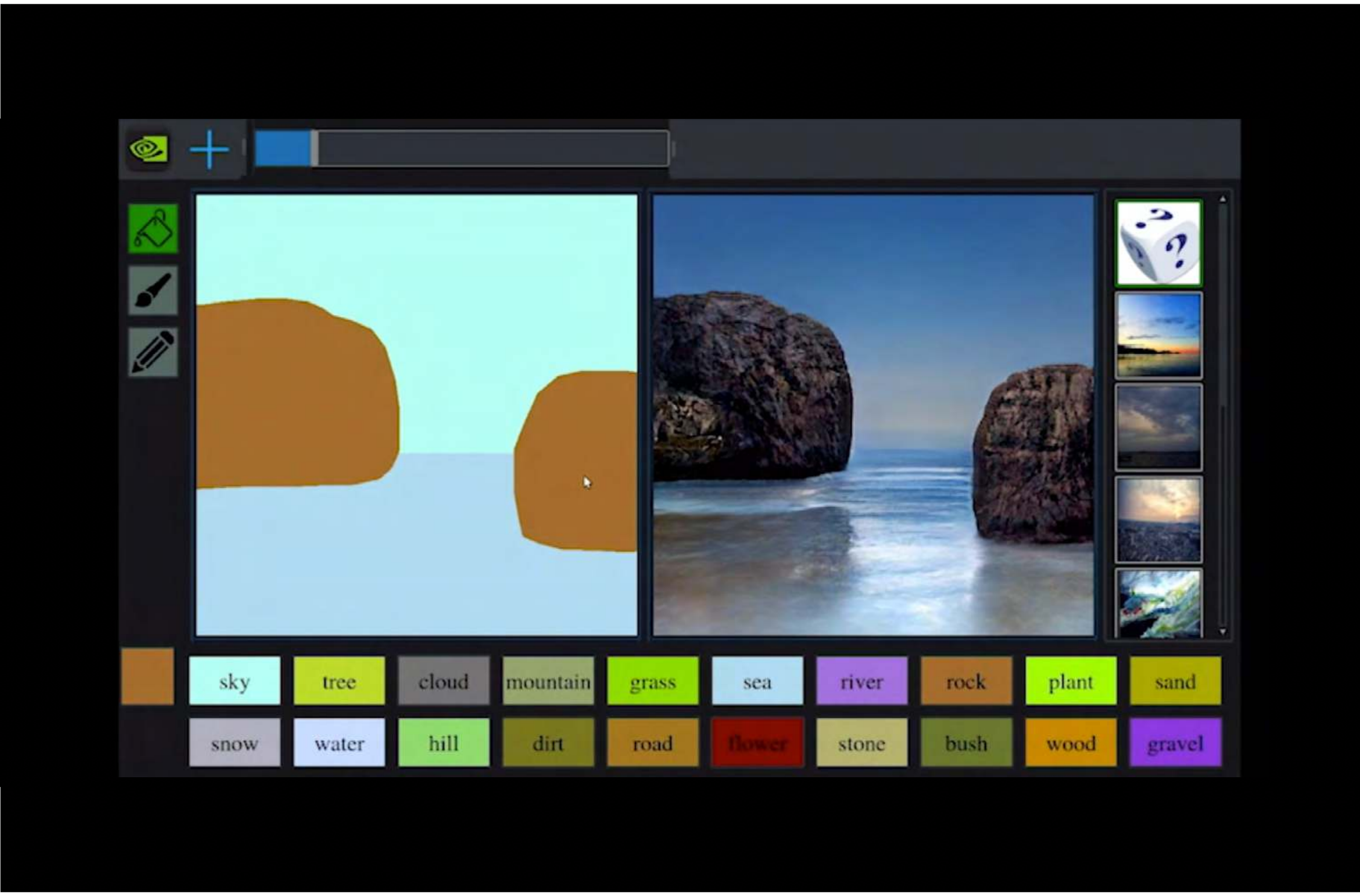
Interessant!
Wat wil je in dat onderzoek bereiken?
De doelstelling is om met behulp van AI brein-computer interactie verder te verbeteren. We zijn namelijk nog ver verwijderd van het daadwerkelijk tot stand brengen van communicatie of het herstellen van zintuiglijke waarneming. Hier gaan nieuwe ontwikkelingen in de AI hand in hand met nieuwe ontwikkelingen op het gebied van de neurotechnologie. Je moet hier denken aan de ontwikkeling van nieuwe hersenimplantaten waarmee we op een veilige manier grote hoeveelheden informatie kunnen sturen en uitlezen. Dit laatste is een van de doelstellingen van Neuralink; een bedrijf dat door Elon Musk is opgezet.

Brein-computer interactie...
Wat zijn de ethische vraagstukken?
Ons doel is om met behulp van deze technologie mogelijkheden te bieden aan mensen die we nu nog niet goed kunnen helpen gegeven de huidige stand van de medische wetenschap. Tegelijkertijd is het van belang dat er goede wetgeving komt waarin duidelijk vastgelegd is hoe en wanneer deze nieuwe technologie mag worden ingezet. Dat is niet fundamenteel anders dan de ethische goedkeuring die vereist is bij het vrijgeven van nieuwe medicijnen.

Zie je nog een andere rol voor GAN-technologie binnen jouw vakgebied?
Vanuit mijn vakgebied is er een mooie analogie tussen het brein en een GAN-systeem te maken. Laat me dat uitleggen: het brein maakt voortdurend voorspellingen van hoe de wereld om ons heen eruit ziet. We zien bijvoorbeeld met onze ogen alleen het middelste deel van ons gezichtsveld scherp terwijl de rest wordt ingevuld door ons brein. Toch zijn we meestal goed in staat om werkelijkheid van fantasie te onderscheiden. De analogie is dat je met GAN ook een nepwereld kunt creëren: volledig synthetisch, volledig gecreëerd. En tegelijkertijd dat een GAN-systeem ook een onderscheid probeert te maken tussen wat echt is en wat nep. Dit is een interessante invalshoek om verder te gaan onderzoeken.

Foto's maken met GauGAN

Generatieve AI-software zal een boost geven aan innovatie en creativiteit. Dat kan ook gezegd worden van het GauGAN-systeem [40] van NVIDIA. Dat systeem kan een eenvoudige tekening, die afkomstig lijkt uit het simpele bitmaptekenprogramma Paint dat Microsoft in 1985 introduceerde, transformeren naar fotorealistische output.



Credit: GauGAN Turns Doodles into Stunning, Realistic Landscapes.
https://blogs.nvidia.com/blog/2019/03/18/gaugan-photorealistic-landscapes-nvidia-research/

Concreet gezegd: met digitale potloden

en kwasten maakt u in een eenvoudig tekenprogramma een tekening van een landschap en de GauGAN-software vult uw input geloofwaardig in met realistische foto-output. U kunt dan heel snel verschillende fotorealistische landschappen genereren en daarin variaties aanbrengen met behulp van deze software. In de video: "Changing Sketches into Photorealistic Masterpieces" [41] ziet u de demo.

De GauGAN software doet mij denken

aan het televisieprogramma van landschapsschilder Bob Ross. Deze kunstenaar en televisiepersoonlijkheid werd eind vorige eeuw bijzonder populair met zijn televisieprogramma *The Joy of Painting*, waarin hij demonstreerde hoe hij zijn landschappen schilderde. Het gemak waarmee Ross schilderijen produceerde, vaak in minder dan een half uur, en het soms verbluffende resultaat is vergelijkbaar met

wat de GauGAN-software doet. Daarbij moet wel worden gezegd dat de GauGAN-software weliswaar sneller een resultaat produceert, maar over het algemeen ook minder ‘samenhang’ heeft.

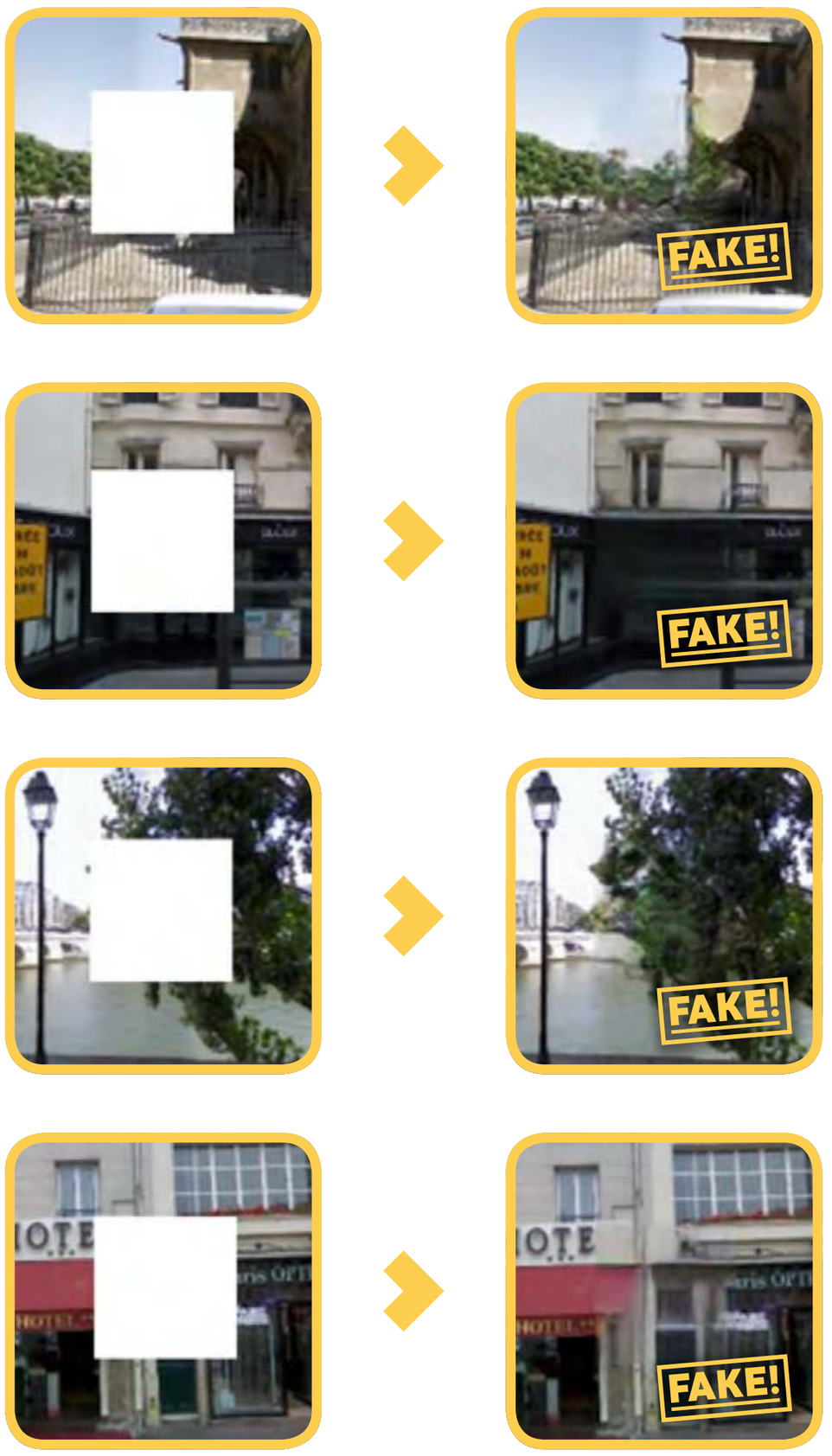
Deze software werkt nu nog voor het genereren van landschappen, maar wat als dit mogelijk wordt voor stedelijke omgevingen, binnenhuisarchitectuur, winkelinrichting of weg- en waterbouw? U kunt overigens ook zelf aan de slag met dit programma [42]. Met een beetje oefenen genereert u al snel geloofwaardige landschapsafbeeldingen.

NVIDIA, het Amerikaanse technologische bedrijf dat vooral bekendstaat om zijn grafische kaarten, kent sowieso een sterke ontwikkeling op dit gebied. Zo kan hun software korrelige foto’s weer scherpte geven [43]. Het kunstmatig intelligente systeem ‘raadt’ dan hoe de foto er uit moet zien in *high definition* en vervangt onscherpe gebieden door scherpe variaties. Zo kunt u bijvoorbeeld onduidelijke of oude foto’s transformeren naar foto’s met een hoge resolutie.

Image inpainting

Het *image inpainting*-softwareprogramma van NVIDIA gaat zelfs een stapje verder. Hoewel het niet specifiek werkt met GAN-technologie, is het wel degelijk de moeite waard van het beschrijven. Met *Image Inpainting* kunt u met een digitaal gummetje op uw beeldscherm complete objecten wegpoetsen uit een foto waarna het generatieve AI-systeem de ontstane leegte geloofwaardig invult. Dat wordt concreet gemaakt in een videodemonstratie [44] waarin een steen, een touw, een persoon en een vlag worden weggepoetst waarna het generatieve AI-systeem de leegte invult door het nieuw gecreëerde fotomateriaal feilloos te laten integreren met de bestaande achtergrond. De software gebruikt een soort vebeeldingskracht om te raden hoe de omgeving eruit ziet.

Wanneer u niet weet dat daar ooit objecten hebben gestaan, wordt dat uit de nieuwe foto niet helder. Het is een supergemakkelijke en vernieuwende vorm van ‘photoshoppen’ die waarschijnlijk binnen een paar jaar functioneert op een app op uw smartphone. Daarmee kunnen foto’s in realtime worden gemanipuleerd. U maakt een selfie en de buurjongen loopt net door het beeld? In de toekomst laat u hem wellicht met een paar vegen uit de foto verdwijnen. Onderzoekers van de Universiteit van Californië in Berkeley zijn overigens ook bezig met dit soort software [45]; hun Inpainting-systeem kleurt uitgegumde witte vlakken uit stadsgezichten in met geloofwaardige nieuwe vormen.

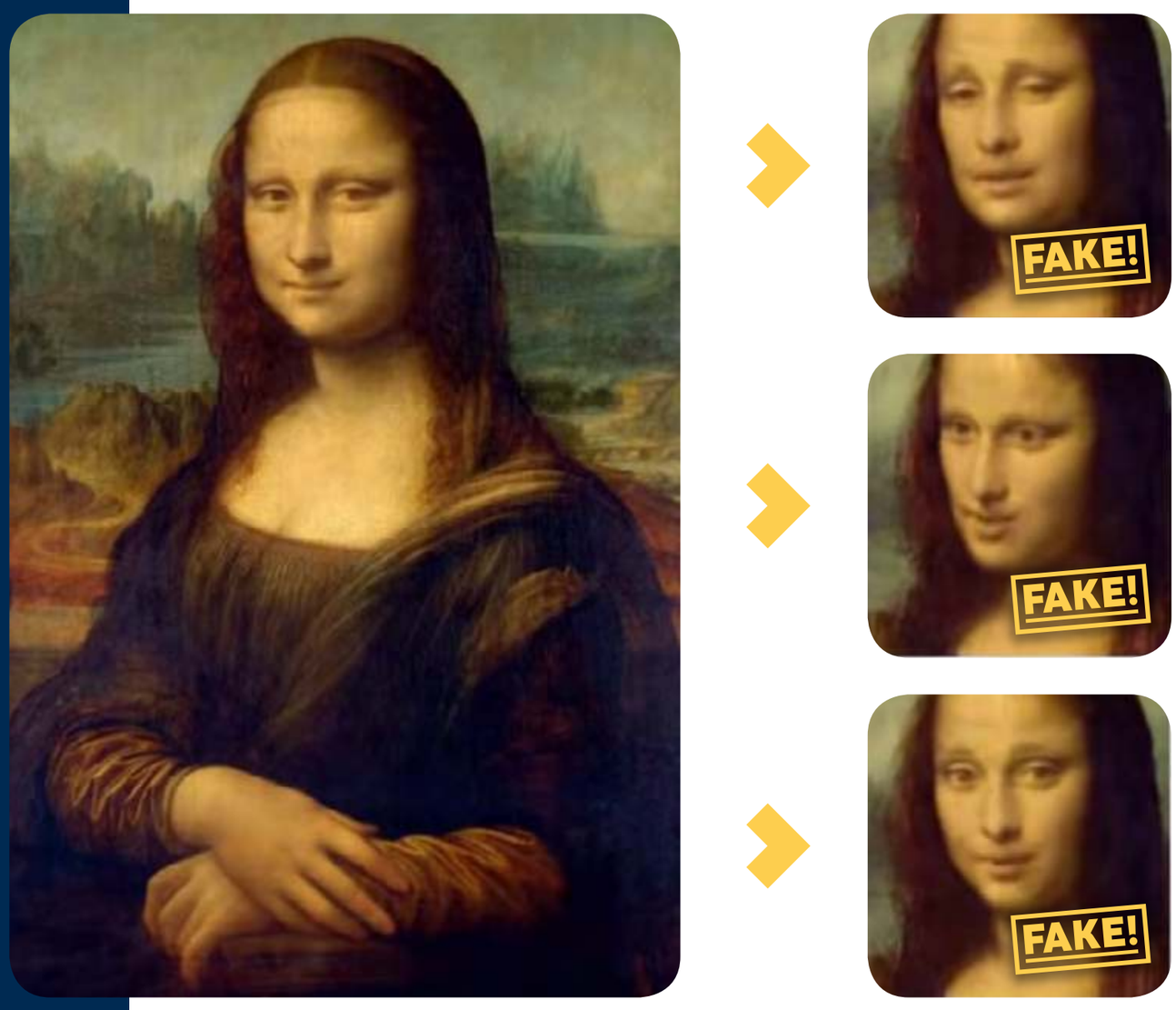


Edited image from source: Unsupervised Feature Learning by Image Inpainting using GANs. https://github.com/pathak22/context-encoder

Video van Mona Lisa

De ontwikkelingen op het vlak van beeldmanipulatie en -generatie gaan alsmaar sneller. Er zijn steeds minder data nodig om een geloofwaardig resultaat te genereren en de resultaten zullen steeds sneller tot stand komen; wellicht in de toekomst in realtime. Waar bijvoorbeeld een jaar geleden nog een flink aantal foto's nodig was om een behoorlijk realistische video te kunnen maken van een pratend hoofd, laten onderzoekers zien dat zij dit ook kunnen op basis van een enkele foto. In het onderzoek "Few-Shot Adversarial Learning of Realistic Neural Talking Head Models" [47] ziet u bijvoorbeeld dat Mona Lisa, Marilyn Monroe, Salvador Dalí en Albert Einstein hun gezicht en mond bewegen, en in alle vier gevallen was er slechts één foto van die beroemdheid gebruikt. Het resultaat is nog niet feilloos, we kunnen het nog onderscheiden van de realiteit, maar daardoor is het niet minder indrukwekkend.

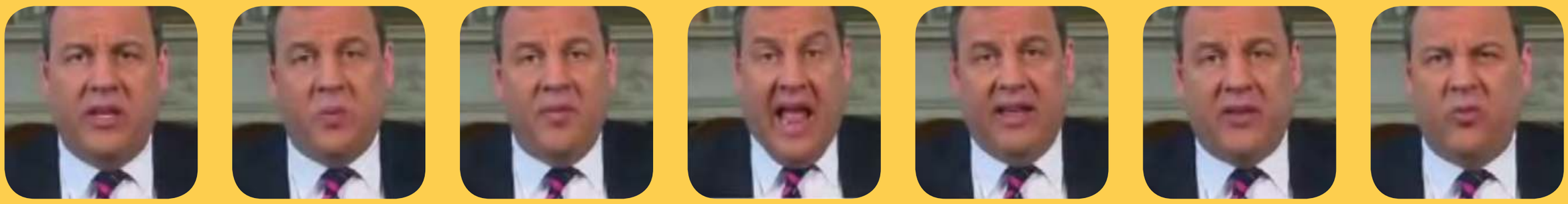
Living portraits



Edited image from source: Few-Shot Adversarial Learning of Realistic Neural Talking Head Models - YouTube
https://www.youtube.com/watch?v=p1b5aiTrGzY

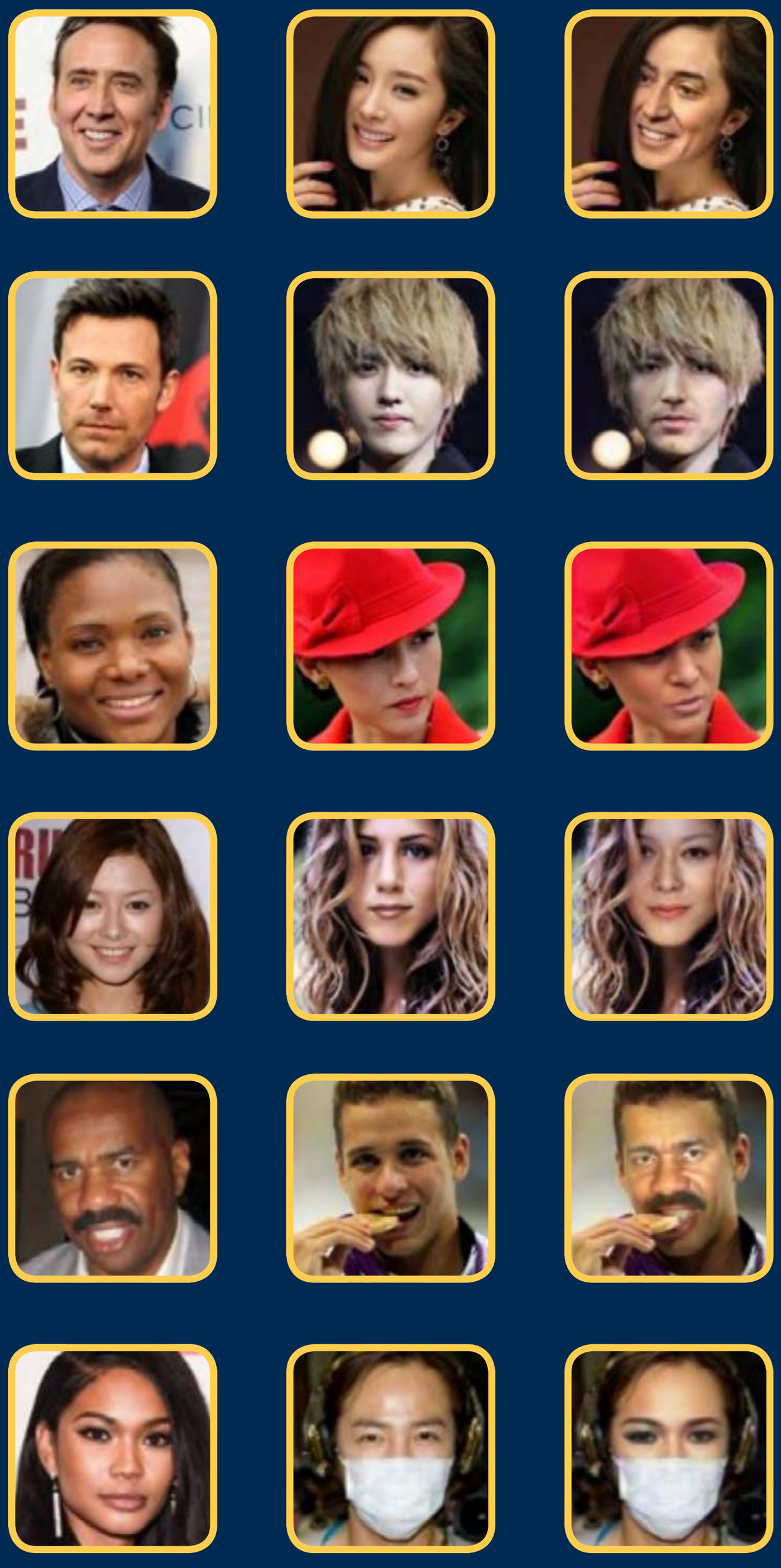
Wat ook interessant is aan dit systeem is dat het beelden kan genereren van hoe mensen er uitzien van de zijkant, terwijl het systeem enkel en alleen is gevoed met voorbeelden waarbij de persoon *recht in de camera* kijkt. Het systeem kan derhalve beelden genereren zonder dat het direct met die concrete voorbeelden wordt gevoed. Het systeem 'verbeeldt' hoe de proefpersonen er van de zijkant uitzien.

Training frames



Edited image from source: Few-Shot Adversarial Learning of Realistic Neural Talking Head Models - YouTube
https://www.youtube.com/watch?v=p1b5aiTrGzY&feature=youtu.be

Software die in video's gezichten verwisselt, *face swapping*, wordt inmiddels steeds beter in het met weinig data resultaten te genereren. Voor het gebruik van die *face swapping*-technologie, FSGAN-software, volstaat bijvoorbeeld nog maar een enkele foto om het nieuwe resultaat te kunnen bewerkstelligen. De resultaten kunnen tevens in realtime worden geproduceerd. De techniek werkt ook bij het 'verwisselen' van etniciteit. Het maken van een *face swapping*-video vereist steeds minder bronmateriaal en steeds minder programmeerkennis en het resultaat wordt met rasse schreden geloofwaardiger. Dat maakt het maken van geloofwaardige nepvideo's steeds gemakkelijker, iets waar ik dieper op zal ingaan in het hoofdstuk over deepfake.



Edited image from source: FSGAN: Subject Agnostic Face Swapping and Reenactment - Yuval Nirkin.
https://nirkin.com/fsgan/

BECKHAM

Bij een internationale campagne tegen de ziekte malaria werd gebruikt gemaakt van *Generative AI-software*. Voetbalster David Beckham sprak in deze commercial [48] vloeiend negen verschillende talen. Vanzelfsprekend was dat niet, hij kon ze niet allemaal plots beheersen. Zijn gezicht en zijn mondbewegingen waren echter gemodelleerd naar door anderen uitgesproken zinnen. De kijker ziet David Beckham, maar hoort de stemmen van anderen die in een andere taal spreken. De lipbewegingen van Beckham zijn perfect gemodelleerd naar deze tekst. Een goed voorbeeld van synthetische media, waarbij origineel en kunstmatig in elkaar overvloeien en is het tevens een originele insteek van een internationale campagne.

Generatieve AI-software biedt vele voordelen en kansen: overleden acteurs kunnen bijvoorbeeld gemakkelijker dan ooit weer ‘tot leven worden gewekt’ in nieuwe films. Acteurs hoeven bepaalde scènes niet meer over te doen en dialogen kunnen achteraf worden aangepast. Advertenties laten zich gemakkelijk aanpassen aan meerdere taalgebieden, zodat het niet opvalt dat de Nederlands gesproken reclame vanuit het Duits is nagesynchroniseerd. Stuntmannen hoeven in films minder gevaarlijk werk te doen en historische figuren kunnen in het heden digitaal tevoorschijn komen en zo bijvoorbeeld het onderwijs voor studenten en leerlingen interessanter te maken. Generatieve AI-software creëert op deze wijze nieuwe beelden.

Misschien kan deze generatieve AI-technologie digitaal zelfs overledenen visueel weer tot leven wekken en op het beeldscherm gehandicapte familieleden dingen laten doen die ze in het normale leven niet meer kunnen. Dat niet alleen: ook kunnen generatieve AI systemen – in de vorm van ‘deepfake-technologie’ – een prachtige bron zijn voor parodie en satire, een onmisbaar onderdeel van onze samenleving.

“ De kijker ziet David Beckham, maar hoort de stemmen van anderen die in een andere taal spreken. De lipbewegingen van Beckham zijn perfect gemodelleerd naar deze tekst.

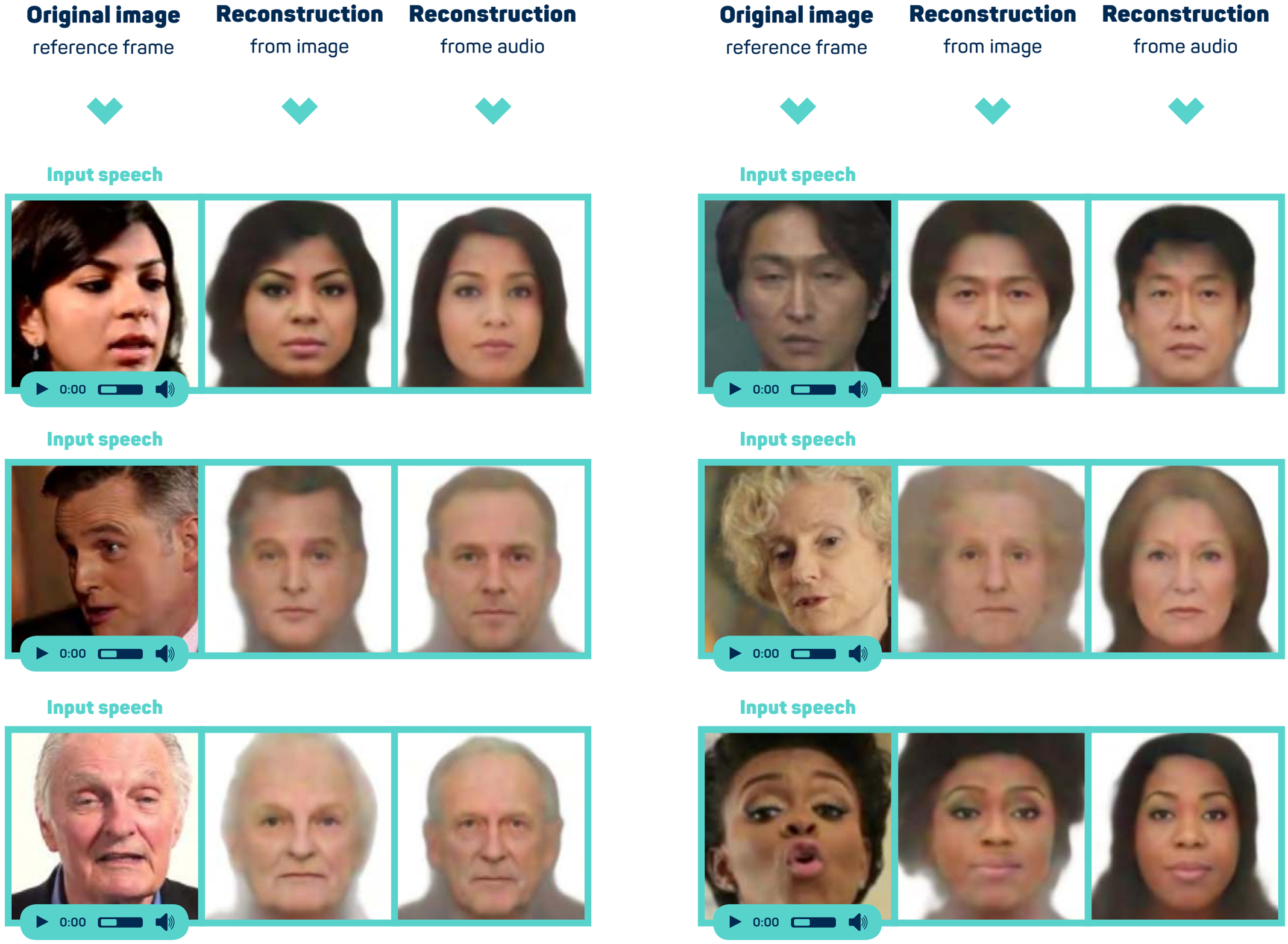
Credit: Using AI deepfake techniques to bring Salvador Dali back to life - YouTube.
https://www.youtube.com/watch?v=BxlPCLRfk8U

De kunstenaar Salvador Dalí is al in 1982 overleden, maar sinds kort is hij weer visueel tot leven gewekt. Het Salvador Dalí Museum in Sint Petersburg, Florida heeft een realistische deepfake-kloon gemaakt van de kunstenaar [49]. Het computersysteem is getraind met videomateriaal van verschillende interviews en deze gezichtsexpressie is geplakt op het gezicht van een acteur die qua lichaamsbouw lijkt op de kunstenaar. Met dit project hoopt het museum dat bezoekers een levendig beeld krijgen van hoe de kunstenaar er uitzag en hoe hij zich bewoog. In het museum is er dankzij dit project meer dan 45 minuten aan nieuw videomateriaal van Dalí te bekijken.

MET VERBEELDINGS-KRACHT KIJKEN NAAR STEMMEN

Een goed voorbeeld van ‘machines met een vonk van verbeeldingskracht’ is te zien bij het Speech2Face systeem. Dat is een generatief AI-systeem dat in staat is om een geloofwaardige afbeelding te genereren van een gezicht op basis van een korte, willekeurige stemopname. En het systeem kan deze gezichtsafbeelding al genereren op basis van een stem-opname van maar een paar seconden. Onderzoekers van MIT hebben de Speech2Face software getraind op basis van miljoenen YouTube video’s van talking heads, sprekende mensen. Het AI-model heeft tijdens de Youtube trainingsuren de correlatie ontdekt tussen de verschillende karaktertrekken van de stem en die van het gezicht. Het Speech2Face systeem herkent bijvoorbeeld leeftijd, geslacht en etniciteit op basis van een stem.

Wanneer je kijkt naar de getoonde afbeeldingen, is het resultaat opvallend goed te noemen. Een prachtig voorbeeld van hoe generatieve AI-software iets laat zien van wat normaal gesproken onzichtbaar is. En laat het ‘verbeelden van het onzichtbare’ nou ook een aspect zijn van de menselijke verbeeldingskracht.



Edited image from source: CVPR’19| Speech2Face: Learning the Face Behind a Voice. https://speech2face.github.io/supplemental/index.html#fig5

Credit: The First Interactive AI Rendered Virtual World - YouTube.
https://www.youtube.com/watch?v=ayPqjPekn7g

We zien inmiddels ook dat generatieve AI-software een volledige driedimensionale wereld kan genereren op basis van voorbeelden. Afkomstig uit het onderzoek Video-to-Video Synthesis [50|51], is de video "The First Interactive AI Rendered Virtual World" [52]. In deze video zien we synthetische beelden waarbij het lijkt of er is gefilmd vanuit een dashboardcamera van een auto die door de stad rijdt. Hoewel de weergave overduidelijk niet fotorealistisch is, zijn wel de vormen, kleuren en omtrek van vele stadsobjecten zeer goed te herkennen.
Er is geen twijfel over mogelijk dat dit de binnenstad van een Amerikaanse stad is. Het generative neural network heeft deze niet bestaande wereld geleerd te reproduceren op basis van videomateriaal.

Generative design

Kunstmatig intelligente systemen als ideeënmachine die interessante output genereren op basis van de input die mensen leveren: we zien dit soort processen steeds meer toegepast worden in de ontwerpfase van productieprocessen. Een voorbeeld daarvan is *generative design*.

Generative design is een ontwerpproces, waarbij verschillende soorten generatieve AI-software helpt bij het genereren van verschillende ontwerpopties in bijvoorbeeld industriële of technische designprocessen. Bouwbedrijven, ontwerpbureaus, autodesigners, vliegtuigbouwers, architectenbureaus en constructiebedrijven gebruiken deze software volop. *Generative design* vindt als het ware nieuwe oplossingen voor hun bestaande ontwerpproblemen. Deze computertechnologie simuleert de fysieke wereld en komt razendsnel met allerlei suggesties voor nieuwe producten, invalshoeken of constructies. Er is dan niet langer sprake van kunstmatige intelligentie die de mens bij het ontwerpen assisteert, maar van de mens die de kunstmatige intelligentie bij het ontwerpen de helpende hand toesteekt, een soort van *'human-assisted AI-design'*.

Dat werkt als volgt. Ontwerpers of ingenieurs voeden de ontwerpsoftware met allerlei parameters zoals prestatie-eisen, ruimtelijke beperkingen, gebruik van bepaalde materialen, productiemethoden en eventuele kostenbeperkingen. De generatieve AI-software genereert op basis daarvan vele verschillende ontwerpen met bijbehorende alternatieven. Het is als het ware een ideeënmachine voor de ontwerper en ingenieur. Omdat deze software is losgekoppeld van het menselijk denken, komt het regelmatig op ideeën waar mensen niet zo gemakkelijk op kunnen komen. Of de software berekent nieuwe constructies die veel minder gewicht bezitten, maar desondanks wel kunnen omgaan met stevige belastingen.

Voor de ontwerper genereert de software vervolgens vele opties op een presenteerblaadje, als hamburgers bij McDonald's. De ontwerper kan dan met de beste hamburgers weer aan de slag om daarmee nieuwe opties te creëren. Verschillende industriële bedrijven en bouwbedrijven gebruiken deze software al. De generatieve AI-technologie is zonder overdrijving de toekomst van het ontwerpproces en kan zo een belangrijke bron van inspiratie zijn.

Hiernaast zie je uitkomsten van een generative design zoektocht naar een nieuw soort stoelbeugel waar de veiligheidsgordels worden vastgemaakt. Het gekozen onderdeel is 40 procent lichter en 20 procent sterker dan het originele onderdeel.

Credit: GM and Autodesk are using generative design for vehicles of the future | Off Grid Energy Independence. https://www.offgridenergyindependence.com/articles/14248/gm-and-autodesk-are-using-generative-design-for-vehicles-of-the-future

Edited image from source: The Volkswagen Type 20 concept Generative Design - Bricsys CAD Blog.
https://blog.bricsys.com/the-volkswagen-type-20-concept-generative-design/

Autobedrijf Volkswagen heeft bijvoorbeeld deze generative design-software gebruikt om een nieuwe vormgeving te creëren voor een elektrische versie van de begin jaren zestig geïntroduceerde T2-bestelwagen van Volkswagen, het model dat zo populair werd als hippiebusje[53]. Zoals vaker als dergelijke software in deze context wordt gebruikt, kregen heel veel vormen een organische structuur, zoals bijvoorbeeld de wieldoppen, het stuur en de spiegelophanging. Jelmer Frank Wijnia, werkzaam bij Van Wijnen, werkt met deze software.

INTERVIEW: JELMER FRANK WIJNIA

GENERATIVE DESIGN LEAD. VAN WIJNEN

“Generative design-software is echt een ideeënmachine”

Hoe kijk jij aan tegen generative AI-software en generative design?
Ik ben heel erg geïnteresseerd in deze technologische ontwikkelingen. Momenteel ontwikkelen we *generative design* voor het maken van stedenbouwkundige plannen. *Generative design* is een revolutionaire manier van ontwerpen. De software creëert daarbij voor ons een legio aan ontwerpopties. Het levert niet alleen ontwerpopties, maar bespaart op sommige plaatsen in het proces ook veel tijd.

Wat levert jullie dat dan op?
Voor het vervaardigen van een stedenbouwkundig plan zijn veel partijen gemoeid. Denk aan de opdrachtgever, bijvoorbeeld een woningbouwcorporatie, de toekomstige bewoner en uiteraard de bouwende partij. Al deze partijen streven naar het vervullen van ieders behoefte in het plan, en dat is een tijdrovend proces. Met *generative design* kunnen we het ontwerpproces deels versnellen. Dat resulteert uiteindelijk in een holistisch(er) ontwerp waar bij de potentie van het plan sneller zichtbaar is voor alle betrokken partijen.

Kun je iets vertellen over de werking van *Generative Urban Design*?
Je moet het zo zien; Aan het begin van het ontwerpproces stel je verschillende eisen op waaraan het uiteindelijke ontwerp moet voldoen; niets anders dan het Programma van Eisen. Dit kan bijvoorbeeld de hoeveelheid zonne-energie zijn die een woning moet opleveren. De software houdt in dat specifieke geval dan rekening met de hellingshoek van de panelen ten opzichte van de zon, zodat het meest gunstig rendement kan worden behaald. Maar denk naast de hoeveelheid zonne-energie ook aan andere doelen zoals de bouwkosten, de uiteindelijke omzet, de tuingrootte en de zichtlijnen. Bij dat laatste bedoel ik: wat zie je eigenlijk vanuit je woning. Je zou overigens gemakkelijk kunnen denken dat een doelparameter als ’omzet’ altijd doorslaggevend is, maar dat is zeker niet zo. Wij gaan niet altijd voor winstmaximalisatie: denk bijvoorbeeld aan een bouwproject met een bepaald prestige, uitstraling. Die zijn voor ons ook heel interessant om in onze portfolio te hebben.

We zijn begonnen met de relatief gemakkelijk kwantificeerbare doelen. Met andere woorden: ze zijn goed in cijfers uit te drukken. Deze gegevens zitten, zeg maar, aan de rationele kant van de software. Dat is ook nodig: een computer moet immers het werk doen.

Je kunt deze doelen in een bepaalde hiërarchie plaatsen of zelfs aan- of uitzetten. Daarmee laat je de software weten wat je voor dit specifieke stedenbouwkundig plan

belangrijk vindt. Het is immers de bedoeling dat de software uiteindelijk de beste concepten genereert. En dat is ook heel interessant aan deze software. Het daagt je uit om voor jezelf steeds te definiëren: wat maakt in dit geval nou het allerbeste concept? Wat draagt daaraan bij? Bij het bouwen van A tot Z zijn er namelijk heel veel variabelen waarmee je rekening moet houden, heel veel krachtenvelden. En *generative design* vraagt als het ware aan jou als ontwerper heel expliciet waaruit het concept moet bestaan. En het geeft je ook de mogelijkheid om te kunnen spelen met deze informatie. Ik kwantificeer als het ware mijn eigen creativiteit in een softwaretool.

Helder.
Hoe gaat dit dan verder in z'n werk?
Nadat je alle doelen hebt gezet, genereert de software bepaalde 'families' van ontwerpen. Bijvoorbeeld een familie waarin het rendement op zonne-energie en de grootte van de tuin heel belangrijk zijn. Je kunt zelf als ontwerper aan de software vertellen wat jij vervolgens weer interessant vindt en wat niet. Je zorgt samen met de software dat het concept als het ware doorgroeit. *Survival of the fittest* in het ontwerpproces. Welk concept overleeft als het ware de volgende fase? Het is heel interessant om de evolutie van bepaalde concepten te kunnen volgen en te beïnvloeden. Met deze software boots je dus de evolutie van de natuur na, maar dan in het ontwerpproces. Wij als mensen doen overigens altijd de nacontrole maar soms moet je er ook voor waken dat dat te strak gebeurt. Het is interessant om de software daar ook wat in te volgen. In dat opzicht is de software echt een ideeënmachine. En omdat het zoveel ideeën kan genereren binnen een kort tijdsbestek, is het ook een tijdmachine. Het haalt bepaalde ideeën naar voren waarvoor je vroeger heel lang de tijd nodig had om te ontwerpen. Plus, je ziet per concept direct hoe er gescoord is op de verschillende doelen.

De computer neemt dus heel veel belangrijke processen van jullie over. Wat kan de software niet?
Waar wij mensen natuurlijk heel erg goed in zijn is om het romantische deel, de emotie van een bepaald ontwerp, toe te voegen. Dat is op dit moment gewoon nog erg moeilijk te kwantificeren. Dat gaat om een bepaald gevoel dat mensen bij het ontwerp hebben. Dat is iets wat de software op dit moment nog niet kan. Denk ook aan het gevoelsmatig kunnen integreren van een bepaald onderwerp binnen de cultuur, de visuele omgeving van de rest van de stad. Wij mensen kunnen aanvoelen wat mooi is, wat passend is en wat natuurlijk is. Dat kan software gewoon niet zo goed.

Gaat dat de software in de toekomst lukken?
Ik denk het wel. Ik ben er wel van overtuigd dat de mens altijd aan het eind de details blijft bijwerken van een ontwerp. Wij als mens snappen heel veel van de wereld om ons heen, dus ook de ruimtelijke wereld. Hoe voelt een ontwerp aan? Hoe voelt een omgeving aan? En daar blijft ook onze onderscheidende waarde zitten. En dat is helemaal prima: wat een computer kan, moet je door een computer laten doen.

Leeft deze ontwerprevolutie al in jullie vakgebied?
Jazeker, steeds meer en meer. Maar ons vakgebied is vervuld met romantiek van het ontwerp, van creativiteit, van het proces. Sommigen zijn wel eens bang dat dat verdwijnt. Maar ik ben juist ervan overtuigd dat die creativiteit in een andere vorm weer verschijnt. Maar nu in samenwerking met de computer. Het is daarom wel belangrijk dat mensen die in het vakgebied werken, goed snappen wat deze revolutie in het ontwerpproces precies inhoudt. Het kan namelijk maar zo zijn dat hun werk er over tien jaar heel anders uitziet. Zelfs op een manier die voor heel veel mensen op dit moment nog niet voor te stellen is.

This does not exist

Het is inmiddels wel duidelijk. We staan aan het prille begin van een krachtige technologische impuls op het gebied van content-generatie en -manipulatie. Videobeelden en foto's zijn nu op dit gebied het meest zichtbare voorbeeld [54], maar ook stemmen [55], teksten, muziek en vele andere toepassingen (die we nu nog niet eens kunnen bevatten) gaan ons in de toekomst enorme zoete vruchten opleveren. De mogelijkheden van GAN als beeld-creatiemachine zijn schier oneindig: een stad die niet bestaat, mensen die niet bestaan, auto's die niet bestaan, virtuele fotomodellen die niet bestaan [56] en daardoor gemakkelijk kunnen worden ingezet voor advertenties en modetijdschriften [57].

Edited image from source: Sunny Farmhouse near the Royal Stays - Guest suites for Rent.
https://thisrentaldoesnotexist.com/

En zelfs deze woning die rechtstreeks lijkt te komen van een profielpagina van AirBnB. *'This rental does not exist'*.

Ook hier geldt: zolang er maar genoeg duidelijke voorbeelden zijn waarvan de machine kan leren, kan de machine vervolgens ook nieuwe uitingsvormen van dezelfde content genereren, variaties op bestaande voorbeelden die passen binnen de dataset waarmee het systeem is getraind. Nogmaals; een GAN-systeem getraind op voorbeelden van AirBnB pagina's creëert nieuwe, verfrissende, vreemde AirBnB pagina's en niet opeens schilderijen van landschappen die nog nooit iemand bedacht heeft, zoals van Gogh en Dalí dat ooit deden.

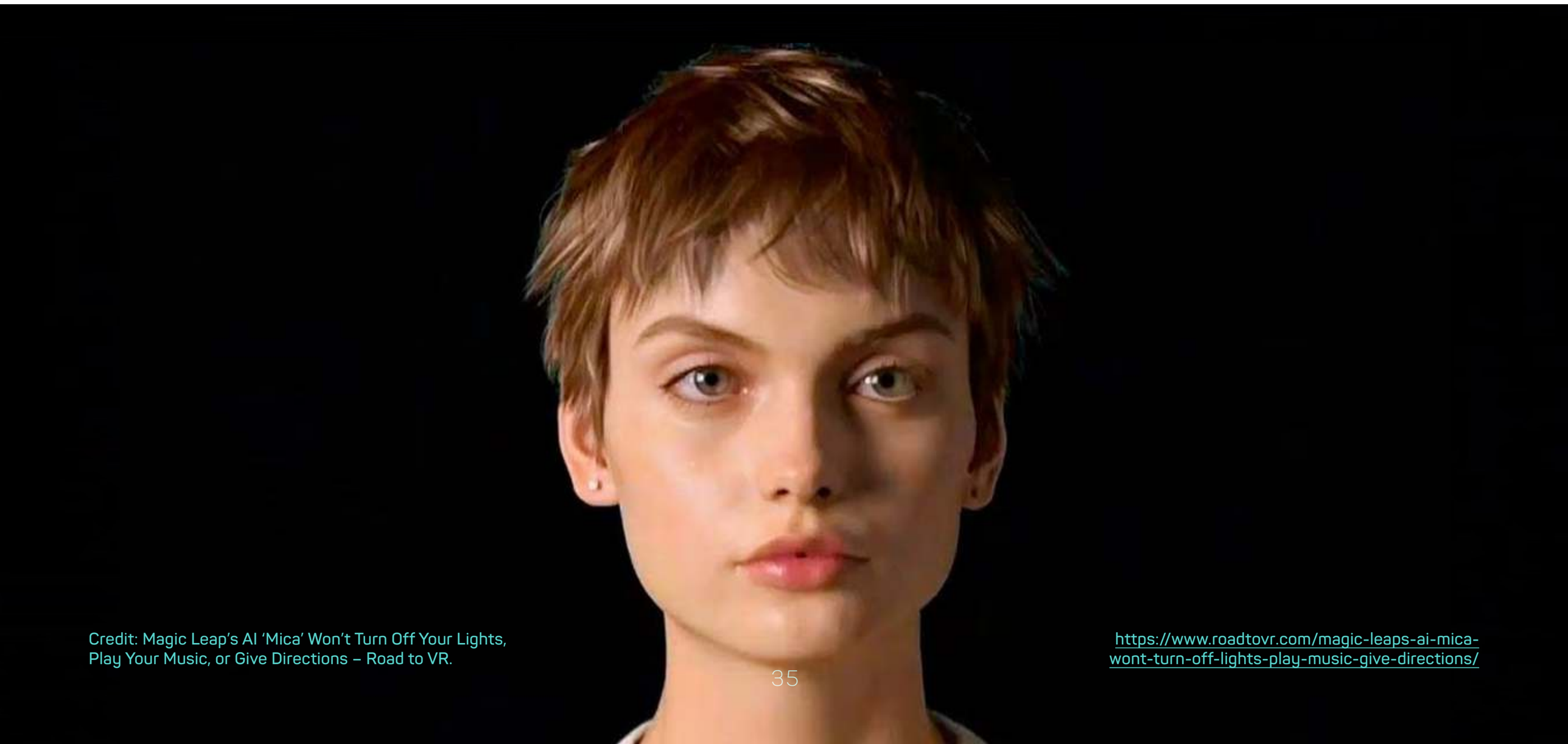
Ongetwijfeld kunt u zich nu levendig voorstellen hoe in de nabije toekomst deze GAN-software ook actief zal worden gebruikt in het domein van de binnenhuisarchitectuur en bij het inrichten van nieuwe winkels. De GAN-software zou in de toekomst op basis van een tekening van een plattegrond of op basis van een aantal foto's van de desbetreffende ruimte u dan mogelijk allerlei suggesties kunnen doen over hoe u uw woning kan inrichten. Of hoe uw nieuwe winkel eruit kan komen te zien. Machines helpen ons met onze verbeeldingskracht.

Waar nu nog mensen handmatig de moeite doen om in een plakboek of in een Pinterest-board al hun ideeën te verzamelen en inspiratie op te doen, wordt dat in de toekomst gedaan door generatieve AI-software. De komende jaren zien we zeer waarschijnlijk de opkomst van vele bedrijven die 'does not exist' als inspiratiesoftware gaan aanbieden.

Ik noemde al even kort het voorbeeld van fotomodellen die niet bestaan. Dat is een domein binnen de digitale contentcreatie dat enorm sterk in opkomst is. Nu is dat nog in handen van professionele animatiestudio's en reclamebureaus, maar in de toekomst zal het steeds gemakkelijker en goedkoper worden om goed lijkende fotomodellen te creëren.

Het is nu overigens nog lastig om hypergeloofwaardige driedimensionale fotomodellen te maken met behulp van GAN. Daar zijn de systemen nog te weinig verfijnd voor. De kwalitatief hoogstaande synthetische avatar Mica, zoals u die hieronder ziet, is gemaakt door animatieprofessionals bij het bedrijf Magic Leap. Voor dit niveau van avatarcreatie zijn GAN-systemen nog onvoldoende gedetailleerd.

Het is toch belangrijk om deze synthetische avatars in dit rapport te benoemen en te beschrijven. In de toekomst worden GAN-systemen waarschijnlijk zo geavanceerd dat synthetische avatars makkelijker te creëren zijn en daarom meer gemeengoed worden. En niet alleen gaan synthetische avatars in de toekomst een veel grotere rol spelen in ons dagelijks leven, maar ook vormen ze een duidelijk voorbeeld van onze interactie met de huidige en toekomstige synthetische wereld. En ze zijn een voorbeeld van de steeds dunner wordende scheidslijn tussen echt en nep, tussen fysiek en digitaal.



Credit: Magic Leap's AI 'Mica' Won't Turn Off Your Lights, Play Your Music, or Give Directions – Road to VR.

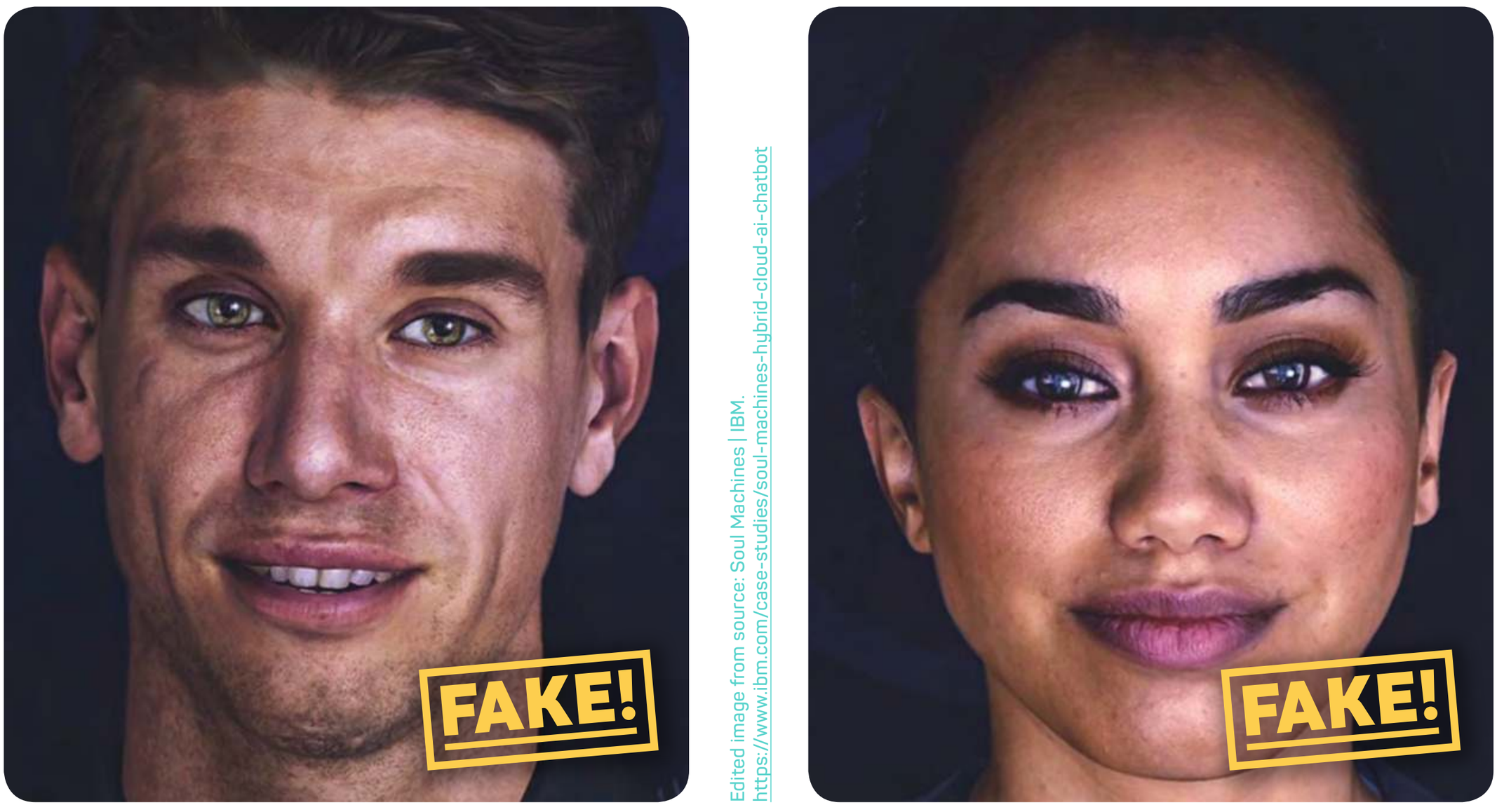
https://www.roadtovr.com/magic-leaps-ai-mica-wont-turn-off-lights-play-music-give-directions/

Synthetic avatars / computer generated Imagery

Synthetic avatars zijn door professionals ontworpen avatars die erg menselijk lijken, maar dat niet zijn. Deze *synthetic avatars* zijn het werk van animatie-professionals die werkzaam zijn bij professionele bedrijven, zoals Soul Machines [58], FaceMe [59], Facebook [60] en Magic Leap [61]. Zij gebruiken professionele software om driedimensionale digitale avatars te maken die in de toekomst kunnen worden ingezet in de klantenservice, bij trainingsdoeleinden of gewoon in gesprekken binnen een *virtual reality*-wereld. De visuele weergave van deze avatars heeft een hoge kwaliteit.

Kijk bijvoorbeeld naar de digitale avatars van het bedrijf Soul Machines. Die zijn met bijzonder veel precisie samengesteld en inmiddels kunnen ze redelijk naturel overkomen. Dat is een prestatie die we niet mogen onderschatten. Het Nieuw-Zeelandse bedrijf heeft als doel om bedrijven de mogelijkheid te geven om digitale avatars in te zetten in hun klantenservice. Digitale medewerkers kunnen worden getraind in het beantwoorden van de 'veel gestelde vragen' die bij de klantenservice binnenkomen. Deze medewerkers zijn schaalbaar; ze kunnen met gemak tienduizend gesprekken per minuut voeren. Ze zijn vriendelijk, dienstbaar, ze volgen altijd de protocollen, zijn nooit moe of chagrijnig, ze vermijden conflicten en hebben steeds meer inlevingsvermogen.

De potentiële markt voor dit soort software, waarbij digitale avatars de meest gestelde vragen voor een bedrijf gaan beantwoorden, is immens. Wanneer deze software goed werkt, en wanneer mensen kunnen wennen aan het idee dat ze met een machine in gesprek te zijn, zal deze oplossing gemeengoed worden. Mensen wiens werk het is om beroepsmatig eenvoudige vragen van klanten, zullen daarbij gedeeltelijk of in het geheel worden vervangen door deze software. Zover is het echter nog niet. Om daadwerkelijk een geloofwaardige conversatie te kunnen voeren met een persoon, is er nog wel een kwalitatieve groeispurt nodig. Wel wordt het uiterlijk van deze digitale medewerkers steeds geloofwaardiger. De verschijningsvorm van synthetic avatars wordt steeds menselijker.



Edited image from source: Soul Machines | IBM.
https://www.ibm.com/case-studies/soul-machines-hybrid-cloud-ai-chatbot

Credit: The New KFC Colonel Is a Computer-Generated Instagram Influencer - Eater.
https://www.eater.com/2019/4/10/18304550/kfc-colonel-instagram-influencer

Brand avatars

De komende jaren zullen bedrijven door de kwalitatieve toename in generatieve AI-software en GAN-software in het bijzonder steeds gemakkelijker en goedkoper in staat zijn om hun eigen brand avatars te maken. Dat zijn digitale modellen die 'het gezicht naar buiten' zijn en geheel gevormd of gekneed zijn naar de wensen van het bedrijf. Dat lijkt nu nog wat futuristisch, maar het gaat denk ik in de toekomst gebeuren. Bedrijven kunnen op deze wijze een unieke verpersoonlijking maken van de uitstraling die zij graag willen voor hun bedrijf.
Voor deze digitale menselijke uithangborden geldt hetzelfde als voor de digitale klantenservicemedeweker: ze kunnen gemakkelijk worden gereproduceerd, zijn altijd inzetbaar, hebben geen moeilijk karakter en laten hun handelen niet beïnvloeden door bepaalde ethische of morele overwegingen. Fastfoodketen KFC heeft bijvoorbeeld een digitale avatar gecreëerd van hun 'Colonel KFC' [62].

In China zagen we het afgelopen jaar zowel een mannelijke [63] als vrouwelijke versie [64] van een digitale nieuwslezer bij de staatsomroep, de Xinhua News Agency. Deze digitale personages zijn gevormd naar de gelijkenis van presentatoren van vlees en bloed, Qiu Hao en Zhang Zhao. Zij zijn ingescand en vervolgens digitaal gekneed tot een driedimensionale weergave. Ook dit is het werk van professionals en is het eindresultaat niet gegenereerd door GAN-technologie. Het is overigens gemakkelijk om bij deze digitale versies het verschil met het echte menselijke origineel nog te zien. Vooral de stem van de digitale presentatoren is nog behoorlijk robotachtig.

Human anchor

Los daarvan is het voordeel voor de Chinese staatsomroep natuurlijk helder: bij *breaking news* zijn de digitale versies meteen inzetbaar, ze kunnen 24 uur per dag, zowel online als op de televisie, het nieuws voorlezen. De nieuwslezers zijn niet moe, chagrijnig of ongehoorzaam. En natuurlijk is dit nog maar een experiment, maar alle in dit hoofdstuk genoemde bedrijven hebben een geloofwaardige toekomstvisie: replica's van bestaande mensen óf volledig nieuwe digitale mensen die ons kunnen voorzien van nieuwsberichten of online gesprekspartner kunnen zijn bij de klantenservice – en als beroemdheden in de entertainmentindustrie.

AI-Anchor

Edited image from source: World's First AI News Anchor Which Is Likely To Replace Human Anchors. https://techgrabyte.com/worlds-first-ai-news-anchor/

1.3 | COMPUTER-GENERATED CELEBRITIES

De eerste puur digitale influencers zijn trouwens allang een feit. De meest bekende door de computer-gegenereerde beroemdheden zijn nu weliswaar nog niet door een GAN gegenereerd, maar dit is in de toekomst niet ondenkbaar. Er zijn immers al door GAN gegenereerde mensen die niet bestaan en fotomodellen die niet bestaan. Een populair voorbeeld van een digitale influencer is Lil' Miquela [65], een interessante digitale weergave van een jonge vrouw met meer dan 1,6 miljoen volgers op Instagram. Ze is het product van een bedrijf met een heldere toekomstvisie: digitale beroemdheden.

Edited image from source: Miquela (@lilmiquela) • Instagram photos and videos.
https://www.instagram.com/lilmiquela/?hl=en

Lil' Miquela verschilt in haar Instagramfoto's niet eens zo veel van de gemiddelde *upper class* tiener in de VS. Ze gaat uit eten, bezoekt de bioscoop, gaat naar een festival of maakt een selfie voor de spiegel. Maar we zien haar ook op de cover van het modetijdschrift *Vogue* en als promotor van een nieuwe Samsungtelefoon; dat kenmerkt de influencer. Na korte tijd te verblijven op haar Instagramprofiel vervaagt voor de kijker langzaamaan de scheidslijn tussen echt en nep.

Miquela is het levende bewijs van de steeds dunnere wordende scheidslijn tussen realiteit en illusie, tussen de virtuele, digitale wereld en de fysieke wereld. En in de wereld van Instagram, waar alles bedacht en gemanipuleerd is, is dit project een voorbode van de synthetische toekomst, een toekomst waarin echt, nep, gemanipuleerd, gefilterd en bewerkt door elkaar lopen. Lil Miquela is inmiddels al lang niet meer de enige; Blawko22 [66] en BermudaisBae [67] zijn ook bekende *computer-generated celebrities*.

De eerste modellenbureaus met digitale modellen kunnen dan niet achterblijven. The Diigitals is zo'n voorbeeld. Ze bieden volledig digitaal gecreëerde personages aan die ontworpen zijn voor modellenwerk. En dat doen ze bij The Diigitals behoorlijk verdienstelijk. Wanneer u hun portfolio kijkt, ziet u grote automerken naast gerenommeerde modemerken.

Edited image from source:: The Diigitals Smart Car.
https://www.thediigitals.com/smart

Ook huidige beroemdheden van vlees en bloed onderzoeken hoe ze een digitale dubbelganger [69] van zichzelf kunnen maken. Op deze wijze kunnen ze in de toekomst bijvoorbeeld met heel veel fans tegelijkertijd in contact blijven. Siri Beerends, Cultuursocioloog, onderzoekt dit soort fenomenen en houdt zich bezig met wat 'echt' is en wat 'nep' is in de digitale wereld.

INTERVIEW: SIRI BEERENDS

CULTUURSOCIOLOOG BIJ SETUP

Hoe kijk jij als cultuursocioloog aan tegen deepfake-technologie?
Wanneer ik iets hoor over deepfake-technologie komt het mij in de meeste gevallen over als problematisch. Kun je echt en nep nog van elkaar onderscheiden? En voedt deze technologie niet het algehele wantrouwen wat er al is ten aanzien van nieuws? Ik maak me soms zorgen dat mensen in de samenleving onverschillig worden ten aanzien van nieuws. Dat ze denken: 'ik kan toch niet meer zien of iets echt of nep is dus het interesseert me niet meer'. Het moderne leven is voor velen al best veeleisend en ik vraag me af of mensen nog wel moeite gaan doen om zichzelf mediawijs te maken.
Mijn ervaring is dat mensen op dit vlak vaak passiever zijn dan ik graag zou willen. De vraag is ook: willen mensen wel mediawijs zijn? Want daarmee creëren ze ook een verantwoordelijkheid om kritischer te worden in hun mediaconsumptie. Ik hoop dus ten zeerste dat deepfake-technologie en synthetische media ervoor gaan zorgen dat mensen zich meer bewust worden van manipulatieve technieken in de media, maar ik weet het niet zeker.

Je bestudeert 'de authenticiteit van nep'. Kun je daar iets meer over vertellen?
Ik zie overduidelijk dat virtueel en realiteit in elkaar versmelten; dat zaken uit de fysieke wereld en de digitale wereld versmelten. Er komen bijvoorbeeld steeds meer virtuele influencers op de markt. Wij van de huidige generatie volwassenen maken nog een onderscheid tussen analoog versus digitaal en echt versus nep. Hoe zal dat zijn voor de jongere generatie? En: is het voor hen nog relevant? Deze virtuele influencers zijn een goed voorbeeld van 'de authenticiteit van nep'. Het is voor het publiek glashelder dat ze synthetisch, digitaal gecreëerd zijn en dat is in hun ogen dan authentieker dan sommige online influencers van vlees en bloed. Sommigen van deze mensen zijn namelijk heel uitgekookt, sluw en manipulatief in hun pogingen om zo authentiek mogelijk over te komen. Zij veinzen dan bijvoorbeeld hun persoonlijke kwetsbaarheden en emoties. Ik denk dat de jonge generatie wel gecharmeerd is van virtuele beroemdheden omdat ze 100% openlijk nep zijn. Echte mensen van vlees en bloed zijn bezig met Photoshop of je ziet slechts de allerbeste foto van tweeëntachtig geschoten exemplaren. 'De authenticiteit van nep' maakt een virtuele beroemdheid erg interessant en ik verwacht dat de populariteit ervan dus gaat toenemen.

Maar je laat je ook wel kritisch uit over deze ontwikkeling. Kun je daar iets meer over vertellen?
Ik verdiep mij al langere tijd in het 'surveillance-kapitalisme'. Een ontwikkeling waarbij bedrijven data verzamelen waarmee ze ons gedrag sturen, zodat ze betere gedragsvoorspellingen aan commerciële bedrijven kunnen verkopen.
Wanneer ik kijk naar virtuele influencers, dan zie ik vooral het verdienmodel van commerciële bedrijven waarbij mensen als grondstof worden gebruikt om reclame-inkomsten te genereren en artificiële intelligentie van techbedrijven te trainen. Wij zijn het product dat data genereert

en bedrijven proberen met de informatie uit onze data ons gedrag te beïnvloeden. De mogelijkheden van de digitale wereld zijn hierin veel krachtiger dan ze ooit in de fysieke wereld zijn geweest.

Kun je dat concreet maken?
Wanneer je kunt chatten met een ‘ouderwetse’ beroemdheid via internet, dan kan deze persoon van vlees en bloed maar één gesprek tegelijkertijd voeren en er worden dan ook geen voorspellingen over jouw gedrag doorverkocht. Bij een digitale variant van een influencer kan zo’n programma wel tienduizend gesprekken tegelijkertijd voeren. Deze gesprekken leveren natuurlijk enorm veel data op die het bedrijf vervolgens weer kan gebruiken. Het feit dat technologiebedrijven zoveel data van ons kunnen afplukken om voorspellingen te doen over ons gedrag en wat wij waarschijnlijk leuk vinden, creëert een onevenwichtige machtsbalans. Daar maak ik me zorgen om. Dat aan de buitenkant iets ‘gewoon een gesprek met een digitale beroemdheid’ lijkt, maar dat bedrijven uit deze gesprekken informatie proberen af te leiden. Informatie die overigens totaal niet hoeft te kloppen met de werkelijkheid.
Algoritmen analyseren onze data en plaatsen ons vervolgens in simpele categorieën en stereotyperende hokjes. Als jij bijvoorbeeld veel woorden gebruikt waaruit blijkt dat je rapmuziek leuk vind, plaatst het algoritme jou in het hokje ‘politiek conservatief’. Vervolgens krijg je alleen maar content aangeboden die politiek conservatief is. Ja, dan kun je inderdaad politiek conservatief wórden. Niet omdat ‘het bedrijf alles over je weet’ of ‘omdat het algoritme jou beter kent dan je beste vriend’, maar omdat het algoritme ervoor zorgt dat je een bepaalde kant op wordt geduwd, waardoor de voorspelling een waarheid wordt. Zo’n voorspelling werkt dus als een selffulfilling prophecy, iets wat maar weinig mensen op hun radar hebben staan maar wel heel belangrijk is.
Het data-gedreven sturen van ons gedrag en onze smaakvoorkeuren maakt dat technologiebedrijven steeds machtiger worden. Wij worden steeds meer de speelbal van hun opgedane kennis.

Waar zie je de oplossingen?
Allereerst natuurlijk dat mensen zich meer bewust worden van het feit dat techbedrijven data gebruiken om informatie af te leiden. En dat deze afgeleide informatie (de simpele hokjes en categorieën waarin algoritmen ons plaatsen) helemaal niets zegt over wie jij bent, maar ondertussen wel wordt gebruikt om jouw smaakvoorkeuren en gedrag te sturen. Ook vind ik dat consumenten vaker mogen nadenken over een nieuwsbron. Over de persoon of organisatie die bepaalde informatie verspreidt: wat is hun doelstelling, motief? Ik zou dus willen dat mensen wat meer mediawijs zouden zijn.
Ook vind ik dat we moeten nadenken over de rol van technologie om deze technologische trend te tackelen. Is meer technologie de beste oplossing? Ik vind het bijvoorbeeld niet meer dan terecht dat we over anti-deepfake-software best nog wel even kritisch mogen zijn. Hoe komt deze software tot stand? Welke data wordt gebruikt? Kun je beslissingen herleiden?
En niet te vergeten: hoe detecteren we dan valse anti-deepfake-software? Software die claimt deepfakes te herkennen, maar dat ondertussen niet doet of een computervirus installeert. Dat soort software zal er zeker komen. Je hebt anno 2019 ook heel veel programma’s die beweren computervirussen, malware en andere infecties te voorkomen maar ze zijn tegelijkertijd juist de verspreider ervan. Hoe zal dat gaan met nep-anti-deepfake-software?

Generatief algoritme vertaalt tekst naar beeld

Zoals eerder in het rapport aangegeven, kan generatieve AI-software worden gebruikt vanuit vele invalshoeken. Een van de meest interessante vind ik daar waar kunstmatig intelligente systemen geschreven teksten kunnen omzetten naar afbeeldingen of video's: *text to image* en *text to video synthesis*, ofwel de omzetting van tekst naar stilstaand en naar bewegend beeld. Wanneer deze technologie volwassen wordt, creëert die enorm veel mogelijkheden.

Wat mij bijzonder enthousiast maakt, is dat GAN-technologie [70] actief afbeeldingen kan genereren op basis van een tekstuele beschrijving. Het GAN-systeem probeert de woorden te begrijpen en zet die vervolgens om naar een afbeelding. Op dat vlak springt op dit moment het onderzoek van StackGAN [71] het meest in het oog. Door twee GAN-systemen op elkaar te stapelen, genereert dit systeem geloofwaardige en redelijk gedetailleerde afbeeldingen op basis van een beschrijving. "Deze vogel is blauw met wit en heeft een korte snavel". "Deze vogel is volledig rood, heeft zwarte vleugels en een spitse bek." "Deze vogel zit dicht bij de grond met zijn korte gele poten; zijn bek is lang en is ook geel en zijn kleur is meestal wit met een zwarte kruin".

Door de samenwerking tussen de twee GAN-systemen leveren bovenstaande beschrijvingen weliswaar kleine, maar geloofwaardige afbeeldingen op. Vanzelfsprekend zijn ze nog niet hyperrealistisch, maar dat zal mijns inziens in de toekomst zeker verbeteren.

Wanneer een kunstmatig intelligent systeem op basis van teksten volledig nieuwe afbeeldingen kan creëren, zal dat wereldwijd een enorme impuls geven aan creativiteit en innovatie. Wanneer kunstmatig intelligente systemen kunnen visualiseren wat wij beschrijven, wordt creatieve digitale technologie voor ieder mens wereldwijd veel gemakkelijker te gebruiken. Dit is nog toekomstmuziek omdat het 'begrijpen' van taal voor AI systemen nog erg lastig is. Er wordt wel volop mee geëxperimenteerd.

Computersystemen die sommige complexere beschrijvingen begrijpen en daarnaar kunnen handelen, zien we bijvoorbeeld nu al in de gamedesign-industrie. Weliswaar werkt dat nog niet met objecten die door een GAN-systeem zijn gecreëerd, ze gebruiken objecten bedacht door professionele ontwerpers. Je ziet bijvoorbeeld in deze video [73] dat het computersysteem van het bedrijf Prometheus de opdracht "ontwerp voor mij een rommelige slaapkamer van een tiener jongen uit de jaren tachtig" snapt.

Let's build a nerdy messy 80's teenager bedroom

Credit: EDITED IMAGE from Promethean AI Announcement Trailer - YouTube. https://www.youtube.com/watch?v=N5O-PDad2Ts

Op basis van een reeds bestaande database met objecten creëert het programma vervolgens geloofwaardig een driedimensionale slaapkamer. Indrukwekkend om te zien.

Het is speculatief, maar stelt u zich eens voor hoe dat in de toekomst eruit kan komen te zien. Ten eerste doe ik de aanname dat computersystemen complexere spraak- en tekstbeschrijvingen gaan begrijpen. Ten tweede zou het zo kunnen zijn dat ze op basis van deze teksten volledige nieuwe content kunnen genereren. Lang leek zoiets sciencefiction, maar het lijkt langzaamaan realiteit te worden.

Met dat soort *speech to image*-technologie zitten we straks wellicht allemaal voor ons computerbeeldscherm en maken we al pratende onze websites. Afbeeldingen, knoppen, opmaak enzovoort. Ontwerp in samenspraak met een kunstmatig intelligent systeem uw eigen website.

Die technologie zal ook handig zijn voor televisieproducenten, reclamemakers, advertentieprofessionals en mensen in de entertainmentindustrie die op deze wijze relatief gemakkelijk een storyboard voor een verhaal kunnen maken. Zij kunnen deze technologie tevens gebruiken om nieuwe ideeën te genereren. Door een bepaalde scène te beschrijven, kan het kunstmatig intelligente systeem hen mogelijk nieuwe invalshoeken laten zien of op nieuwe ideeën brengen.

Er is voldoende ruimte voor speculatie. Neemt de intelligentie van dit soort *text to image*-GAN-systemen in de toekomst dermate toe dat die een boek kunnen 'lezen' en dan op basis van de inhoud een omslag bedenken? Nu is dat zeker nog niet zo. AI systemen hebben geen echt begrip van taal, omdat je daarvoor ook heel veel van de wereld om ons heen moet snappen.

Nog even een toekomstvisie; Kunnen bloggers in de toekomst gemakkelijk met één druk op de knop interessante afbeeldingen bij hun blog laten genereren door een kunstmatig intelligent systeem, zodat ze niet alleen schrijver maar ook ontwerper zijn? Betekent dit dat fotografen en ontwerpers op het gebied van stockfotografie minder relevant worden en abonnementen op stockfotodatabases kunnen worden beëindigd?

Misschien wordt het in de toekomst zelfs mogelijk dat een filmproducent een kunstmatig intelligent systeem verschillende versies van het einde van een videoclip laat samenstellen. Of wellicht wordt een text to image-GAN-systeem dermate geavanceerd dat het op basis van gesproken aanwijzingen een andere wending in een videoclip kan bedenken. Wanneer ik nog wat verder speculeer richting de toekomst, dan kijken we misschien naar online videoclips waarbij we daarbinnen zelf een nieuwe invalshoek kunnen laten genereren. Wat denkt u ervan als het ooit mogelijk is dat een kunstmatig intelligent systeem op basis van uw persoonlijke wensen een hyper-individuele aflevering van uw favoriete serie genereert?

Die speculatieve lijn nóg verder doortrekkend zou het ook kunnen zijn dat in de toekomst een auteur een verhaal schrijft en dat daarna door een generatief AI-softwaresysteem naar een filmbestand laat omzetten, zeg maar van Microsoft Word naar MPEG-4. Zullen we in de toekomst rondlopen in onze eigen *virtual reality*-wereld en al pratende tegen de computer onze eigen omgeving vormgeven? Of kunnen we in de toekomst zelfs onze gedachten visualiseren, waarbij het virtual reality-systeem raadt wat we denken en op basis daarvan visuele content produceert? Over GAN systemen die driedimensionale werelden genereren en generative AI technologieën in de wereld van game-design sprak ik met Berco Beute.

INTERVIEW: BERCO BEUTE

PHD. SOFTWAREPROFESSIONAL. SPECIALITEITEN: SOFTWARE ENGINEERING, AI, COMPUTER-AIDED CREATIVITY.

Hoe kijk jij vanuit jouw rol aan tegen GAN-technologie?
Met name in combinatie met *semantic understanding* vind ik GAN-technologie een intrigerende manier om met AI ideeën snel en gemakkelijk te realiseren. AI als 'compiler' die ideeën omzet naar werkende systemen. De stand van zaken anno 2019 is nog steeds dat je een stevige informatica-opleiding moet volgen om het informatiesysteem te bouwen dat je ideeën verwezenlijkt. Er is dus altijd een hoge drempel om een digitaal product te maken. Met de opkomst van GAN-technologie lijken we definitief over deze drempel te kunnen stappen.

Kunstmatig intelligente systemen worden steeds beter om de intentie van mensen uit een boodschap te halen. En wanneer ze de menselijke intentie begrijpen, kunnen ze ook meehelpen om ideeën te verwezenlijken en zo bijvoorbeeld veel betere muziek of teksten genereren. Nu heb je al GAN-systemen die muziek kunnen genereren, maar die muziek is vrijwel altijd zielloos, er zit geen intentie achter. Content krijgt pas waarde wanneer je de intentie van de maker er doorheen voelt. Uiteraard kun je GAN-technologie gebruiken om ideeën en inspiratie op te doen, maar daadwerkelijk content creëren dat waardevol voelt... daarvoor moet het systeem de intentie van de maker kunnen aanvoelen en representeren.

Deze trend is reeds zichtbaar in de game-industrie, waar het zowel de creatieve mogelijkheden kan vergroten als de kosten verlaagt. Eerder was het zo dat je heel veel mensen nodig gehad om een computerspel te maken: Het vereiste veel experts om computercode te schrijven en 2D/3D-graphics te ontwerpen. Nu komen we in een tijdperk waarin het systeem zelf de computercode en graphics genereert. Wanneer je een computer data geeft in de richting die je graag zou willen zien, schrijft het zelf de regels. En die data kunnen dus ook een globale beschrijving zijn. Een kunstmatig intelligent systeem kan zowel interpreteren als samenstellen en genereren. Dat is een interessante ontwikkeling.

Kun je dat concreet maken?
Je kunt nu al tegen een AI-systeem zeggen: "Maak voor mij een driedimensionale weergave van een jongenskamer uit de jaren tachtig". En dat lukt.
Of je zegt tegen een AI-systeem: "Creëer een zonnig landschap in Frankrijk, met een korenveld en een molen". Wanneer je het zo concreet benoemt, is het binnenkort al mogelijk dat een kunstmatig intelligent GAN-systeem dit voor je genereert. Je kunt dus intuïtief zelf je content maken.
Moet je je eens voorstellen hoeveel tijd dit gaat besparen voor de game-industrie. Nu heb je heel veel mensen nodig die uitermate complexe software en 3D-werelden kunnen maken. In de toekomst ziet dat er dus heel anders uit. De software genereert dan driedimensionale beelden in samenspraak met mensen.

Hoe ziet dat er dan in de toekomst uit?
Het wordt in de toekomst bijvoorbeeld veel gemakkelijker om een film te gaan maken. Je

verzamelt een regisseur, een scriptschrijver, acteurs en de mensen van techniek, geluid en kleding en met deze groep maak je een film door het AI-systeem mondeling aanwijzingen te geven. Je kunt als team dan het computersysteem vertellen wat het moet maken:

“Het speelt zich af in Engeland in de tijd van de middeleeuwen. De eerste scène is in een zeer bosrijk gebied. Iets minder donker. Nog iets minder donker. Meer bomen aan de linkerkant. In de eerste scène spelen een man en een vrouw. De vrouw is van adel. In de scène zijn verschillende vogels zichtbaar”.

Zo moet je het voorstellen. Je werkt als team samen met het kunstmatig intelligent systeem. In realtime geef je aanwijzingen en die zie je op het beeldscherm verschijnen. De beelden worden als het ware voor je neus tevoorschijn getoverd. De kwaliteit ervan zal ook steeds beter worden zodat je in realtime realistische effecten kunt zien. Je ziet bijvoorbeeld ook al dat natuurlijke lichtinval en schaduw steeds beter wordt in games.

Als ik verder in de toekomst kijk, dan weet ik zeker dat we als mens steeds meer gaan verblijven in een digitale wereld. In een simulatie waarin je alles naar wens kunt aanpassen. Waarin de wereld zich aanpast naar jouw wensen Soms zonder dat je je daar bewust van bent. Waar het waarschijnlijk veel comfortabeler is om te verblijven. Je ziet nu al dat mensen zich soms verliezen in games of in hun smartphone omdat die wereld prettiger is. Ik denk dat veel mensen op de lange termijn er toch voor kiezen om langduriger te verblijven in de gesimuleerde, digitale wereld waarin het beter toeven is dan in de onprettige fysieke wereld. Ik geloof zelf dat de meest waardevolle ervaringen liggen in de fysieke wereld, maar wellicht dat over een paar decennia mensen hier heel anders naar kijken.

Volg je ook de ontwikkelingen op het gebied van deepfake-technologie?
Eerlijk gezegd meer van de zijlijn. Ik zie wat het kan doen en de techniek heeft wel degelijk mijn interesse, maar dat baart me natuurlijk vooral zorgen. Vanuit de studie AI weet ik wat er kan, en vanuit de studie communicatiewetenschappen weet ik wat de kracht van massamedia is, en hoe die zowel ten goede als ten slechte gebruikt kunnen worden. Vandaar dat ik hoop dat mensen zichzelf willen aanleren om nieuwsberichten en nieuwsbronnen goed te kunnen analyseren. Dat iedere burger wat journalistieke analyse-vaardigheden krijgt, zeg maar. Je moet mensen op dat vlak denk ik actief gaan trainen om met de tsunami van nepinformatie te kunnen omgaan.
Maar eerlijk gezegd heb ik daar weinig vertrouwen in. En dat is natuurlijk wel een kwalijke zaak, want hoe minder je daadwerkelijk weet, hoe meer je in je beslissingen vaart op emotie. Met de opkomst van deepfake-technologie wordt het mensen wel heel moeilijk gemaakt om waarheid en nep uit elkaar te halen.

Generatief algoritme vertaalt beeld naar tekst

Text to image vormt een toekomstige creatieve bron, maar andersom kan het ook. Beeld naar tekst *(image to text)* werkt inmiddels al best goed: wetenschappers hebben anno 2019 redelijk presterende *image to text*-systemen gemaakt. Wereldwijd proberen onderzoekers om hun computersystemen te leren te herkennen wat er op een afbeelding gebeurt, wat de activiteit of handeling is, en daar een kloppende tekst bij te bedenken. Veel mensen die dit gebied al een tijd volgen, kennen het werk van Karpathy en Fei Fei, 'Deep Visual-Semantic Alignments for Generating Image Descriptions' [74]. In het kort komt erop neer dat dit systeem objecten kan vertalen naar een concrete tekstuele beschrijving: "Een man in een zwart T-shirt speelt gitaar", "Een wegwerker in een oranje hesje werkt aan de weg", "Twee jonge meisjes spelen met Lego".

Dat is een bijzonder interessante ontwikkeling omdat een computersysteem probeert te doen wat voor ons mensen enorm alledaags en vanzelfsprekend is: het kunnen zien wat er zich in de omgeving afspeelt en daaraan context te kunnen geven.

Dat soort software zal in de toekomst standaard beschikbaar zijn voor blinden en slechtzienden om hen te vertellen wat er in de wereld om hen heen gebeurt. Wereldwijd wordt er volop gewerkt aan dergelijke projecten.

Man in black shirt is playing guitar.

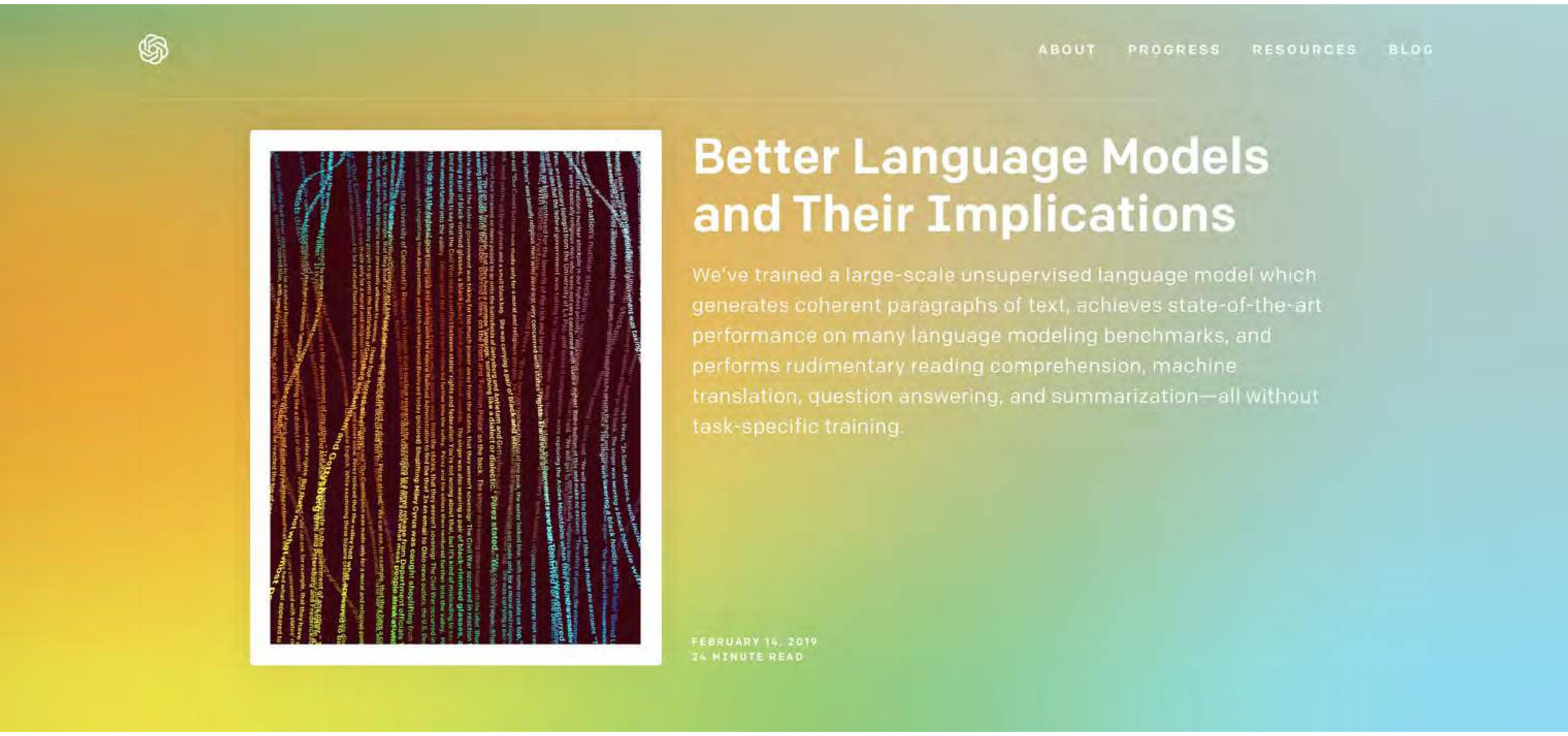
Two young girls are playing with lego toy.

Boy is doing backflip on wakeboard.

Edited image from source:
Deep Visual-Semantic Alignments for Generating Image Descriptions.
https://arxiv.org/abs/1412.2306

1.4 | TEKSTGENERATIE DOOR GENERATIEVE AI-SOFTWARE

Een logische toepassing voor generative AI-software ligt in het domein van geschreven tekst. Er is immers veel geschreven tekst beschikbaar (om als trainingsvoorbeeld te dienen), er zijn heel veel praktische toepassingen te bedenken en het is geen ondoorgrond domein: wetenschappers doen daarbinnen al decennialang onderzoek.



Credit: Better Language Models and Their Implications.
https://openai.com/blog/better-language-models/

Het Amerikaanse bedrijf OpenAI onderzoekt bijvoorbeeld al enige tijd de frontlinie van kunstmatige intelligentie. In februari 2019 publiceerde het een blog over GPT2, Generative Pretrained Transformer 2, een AI-systeem dat is getraind om zelfstandig geloofwaardige teksten te genereren. Het systeem kan dankzij de training het volgende woord in een bepaalde zin voorspellen. Daardoor kan GPT2 volledige teksten produceren. Het computersysteem heeft daarbij overigens geen weet van de 'betekenis' van woorden, maar is vooral goed in het doen van voorspellingen. Daarom kan het systeem ook op andere gebieden worden ingezet, maar daarover later meer.

Het (774 miljoen parameters) GPT2-model werd in februari 2019 overigens maar zeer beperkt beschikbaar gesteld aan derden. In eerste instantie werd toen een klein 124M-model beschikbaar gemaakt en in mei 2019 volgde een middelgroot 355M-model. De beperkte beschikbaarstelling van de software was een breuk met de *open source-traditie* van OpenAI. De beslissing daartoe werd genomen omdat de makers van GPT2 bang waren dat de software zou worden gebruikt voor negatieve doeleinden, zoals het genereren van misleidende nieuwsartikelen om de publieke mening te beïnvloeden, identiteitsfraude, nepnieuws op sociale media, e-mails voor phishing of negatieve recensies. Een belangrijke overweging daarbij was dat de schaal, de hoeveelheid van gegenereerde desinformatie,

I CONTROVERSE

Stel dat een bepaalde groepering in Nederland de politie in een kwaad daglicht wil stellen. In de toekomst kan die met een soort generatieve AI-software, een GPT2-achtig systeem, snel artikelen online zetten die zijn geschreven – of nog beter, gegenereerd – om het sentiment richting de politie negatief te beïnvloeden. De gefantaseerde headlines 'Politie grijpt hardhandig in bij vreedzame demonstratie tegen te tolerant immigratiebeleid', 'Demonstranten aangevallen door politiek gekleurde politieagenten bij vreedzame demonstratie Nederlandse Volkspartij', 'Verbale en fysieke agressie door politie bij bijeenkomst in Den Haag' kunnen dan wellicht razendsnel leiden tot geloofwaardige gegenereerde artikelen die de werkelijkheid geweld aandoen.

met behulp van GPT2 enorm kan toenemen. Op de website "Talk to Transformer" [75] kunt u niettemin stoeien met de GPT2-software. U vult een titel in en de software produceert op basis daarvan een nepartikel.

Natuurlijk ís er in het huidige medialandschap al heel veel nepnieuws. Wanneer echter een kunstmatig intelligent systeem op basis van één *headline* rondom een bepaald thema duizenden artikelen per dag kan produceren en daarbij op basis van clicks steeds beter leert hoe het kan inspelen op negatieve emoties van het publiek, dan is er een goede reden om gezond kritisch te zijn op het gebruik van die technologie.

De beslissing van OpenAI om de software niet volledig beschikbaar te stellen, leidde overigens tot flink wat controverse. Is het immers niet heel nuttig om goed te snappen hoe deze software werkt, zodat nepnieuws, gecreëerd door kunstmatig intelligente systemen, sneller kan worden gedetecteerd? Anderen beschouwden het juist als een goede zet dat OpenAI had besloten om niet het hele GPT2-model te delen met de buitenwereld. Dat kwam vooral omdat ze zich er ook zorgen over maakten dat het model bruikbaar is voor negatieve doeleinden, maar misschien nog wel belangrijker vonden ze het dat de beslissing van OpenAI de discussie opent over wanneer het wel of niet wenselijk is om bepaalde AI-software te delen.
Mag je zomaar iedere vorm van software openbaar maken en dan van anderen verwachten dat die het negatieve gebruik van de software gaan bestrijden?

Grover-systeem

Niet lang na een gedeeltelijke release van de software van OpenAI kwam het Allen Institute for Artificial Intelligence (AI2) – in 2013 opgericht door Paul Allen (1953-2018), mede-oprichter van Microsoft – met het Grover-systeem [76]. Dat is een nieuw AI-model dat losjes is gebaseerd op het GPT2-model van OpenAI. Op de website introduceert het instituut de software als 'het medicijn tegen kunstmatig intelligent gecreëerd nepnieuws'. Het Grover-model kan met kunstmatige intelligentie gegenereerd nepnieuws detecteren omdat het systeem zelf nepnieuws kan maken [77]. Het Grover-systeem is dus tegelijkertijd ziekteverwekker en medicijn; het kent zijn eigen manier van werken, gedachtesprongen en voorspellingen en kan daarom goed detecteren.
Grover is een systeem dat niet alleen nepnieuws kan detecteren dat het zelf heeft gegenereerd, maar ook het nepnieuws van andere systemen. Onderzoek laat zien dat het Grover-systeem in 92% van de gevallen slaagt in het herkennen van met kunstmatige intelligentie vervaardigd nepnieuws.

Counterintuitively, the best defense[78] against Grover turns out to be Grover itself, with 92% accuracy.

Via de website kunt u zelf een artikel genereren op basis van een zelfbedachte titel [79]. Meer hoeft u niet in te voeren. Het Allen Institute for Artificial Intelligence laat er een aantal voorbeelden van zien:

"Why Bitcoin is a great investment"
"As most of my readers know, I'm an optimist. This belief applies across my life, and to various investments as well. So I am intrigued by the success of cryptocurrencies, such as Bitcoin and Ethereum. The competition they are putting up against the gold standard looks insane, as Bitcoin goes off to the races. There's no way to fully understand what's going on in the crypto world — and I am not even sure anyone could if you tried to. Still, I can tell you that Bitcoin's recent surge is really an opportunity to buy long-term real assets. (...) There is still a lot of potential, and potentially a lot of wealth to be made. Imagine the world right now, say 10 years from now. Have you invested in things like gold, real estate, or corporations? "

"Link found between autism and vaccines"
"A paper published in the journal Genetics and Human Behavior claims to find a "significant link" between autism and a type of booster vaccine given in childhood. The paper focuses on the mercury-based preservative thimerosal and considers the possibility that the injectable vaccine contributes to autism-related symptoms. These include poor social skills, language delays and seizures, according to the paper. (...) The authors also found that the autism risk increased by twofold in children aged 3 to 10 who received the hepatitis B booster vaccine against Hepatitis B, regardless of the following vaccination schedules: aged 9 months or older with only one dose; or under 4 months old and only one dose. "

Grover kan trouwens ook nepfilmrecensies genereren, bijvoorbeeld over de film Sharknado 6. Het systeem produceert met hetzelfde gemak zakelijk nepnieuws, zoals bijvoorbeeld een artikel over een 'Uber for Dogs'-startup.

Grover is, zoals aangegeven, losjes gebaseerd op het eerder genoemde GPT2-model van OpenAI. En net zoals OpenAI heeft ook het Allen Institute for Artificial Intelligence besloten om de grote dataset van Grover niet openbaar te maken.

Toekomst

Nu kunstmatig intelligente systemen suggesties kunnen doen om teksten aan te vullen en zelfs volledige teksten kunnen genereren op basis van enkel en alleen een titel, hoe ziet de toekomst er dan uit? Het is speculeren, maar misschien komt er goede *autocomplete* software op de markt voor programmeurs. Die software doet tijdens het code schrijven goede suggesties hoe de huidige regelcode zal kunnen worden afgemaakt. Diezelfde software gaat de programmeurs voorstellen hoe de volgende regel programmeercode eruit kan zien. Waarschijnlijk volstaat eenvoudig indrukken van de tabtoets voor het accepteren van deze suggestie. Dat zal enorm veel tijd besparen.

Het programma Deep TabNine [80] doet overigens zoiets al. Het is geen GAN-technologie, maar een andere vorm van machine learning en het is getraind op twee miljoen bestanden van coderingwebsite GitHub. Deze Deep TabNine-software werkt echter nog niet perfect, het beste resultaat levert die bij het automatisch aanvullen van relatief veel voorkomende code. Het is echter een kwestie van tijd voordat dat verbetert. In de toekomst zullen kunstmatig intelligente systemen wellicht complete e-mailberichten voor ons gaan schrijven. Dan hoeven wij die berichten enkel en alleen nog door te lezen voordat we ze goedkeuren, waarna het bericht kan worden verstuurd. Het is zelfs voorstelbaar dat in de wat verdere toekomst kunstmatig intelligente systemen namens ons reageren op gemakkelijk te beantwoorden e-mails. Dat klinkt nu nog futuristisch, maar machines leren immers van voorbeelden, in dat geval onze goed- en afkeuring. De kwaliteit van dat soort *autocomplete*-software gaat met sprongen vooruit. En wie weet gaat het aanvullen van onze zinnen in e-mails, blogs en berichten op sociale media door GPT2-achtige software in de toekomst heel natuurlijk aanvoelen. Vertalen van grote stukken tekst wordt in de nabije toekomst in dat geval nagenoeg perfect en zal vrijwel in realtime plaatsvinden. Nogmaals; het is speculatief en bovengenoemde scenario's zijn niet van vandaag op morgen de realiteit, maar ze behoren tot de meer zinnige scenario's.

Ook behoort het schrijven van een samenvatting door generatieve AI-software, op basis van een lang artikel of een whitepaper, tot de mogelijkheden. Voordat deze software perfect werkt, zijn we nog wel een paar jaar verder, maar ik twijfel er niet aan dat dit soort software gemeengoed gaat worden.
Stelt u zich eens voor wat dat ons in de toekomst kan gaan opleveren. Een kunstmatig intelligent systeem zou een rapport van vijftig pagina's over een nieuwe technologische trend kunnen samenvatten op anderhalf kantje A4. Download twintig rapporten over 'decentrale autonome organisaties' en laat de software van ieder rapport een samenvatting maken. Na het lezen van twintig samenvattingen bent u dan ongetwijfeld al een stuk wijzer en kunt u tevens beoordelen wat een geschikte samenvatting is – en daarmee welk rapport waarschijnlijk goed zal zijn. U kunt er dan alsnog voor kiezen dat rapport uitgebreid te lezen. Ieder land met een sterke kenniseconomie zal enorm de vruchten van deze software plukken.

Het is speculatief, maar het zou kunnen zijn dat in de toekomst GPT2-achtige systemen plagiaat kunnen voorkomen en helpen bij het omzeilen van auteursrecht en intellectueel eigendom. In dat geval laten scholieren, studenten, onderzoekers, journalisten, bloggers en auteurs hun artikelen, verslagen en onderzoeken herschrijven door een GPT2-achtig systeem dat ze zodanig verandert dat ze niet meer herleidbaar zijn tot de originele bron, maar toch erg goed leesbaar zijn.

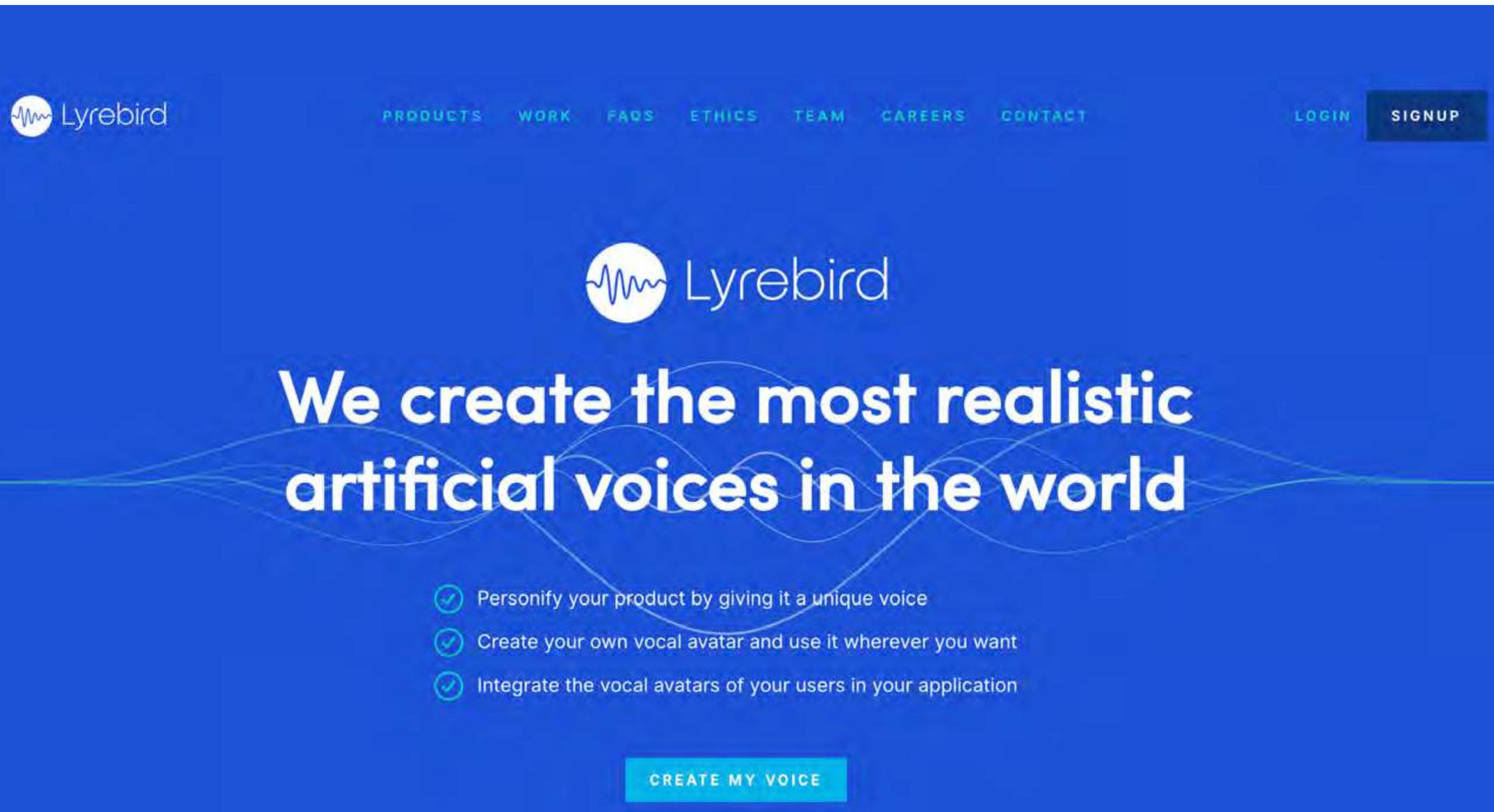
Aan de ene kant is dat natuurlijk een ontwikkeling die voor bovengenoemde beroepsgroepen positief is. Om iedere schijn van plagiaat te voorkomen, helpt het als een kunstmatig intelligent systeem teksten herschrijft [81]. Voorkomen is beter dan genezen, zullen sommigen denken, zeker als anders een bezoek aan de rechter dreigt. Tegelijkertijd kunt u zich afvragen of dat inhoudt dat auteursrecht en intellectueel eigendom nagenoeg betekenisloos worden. Ontstaat er het risico dat auteurs en journalisten gemakzuchtig stukken tekst van het internet kunnen kopiëren en er dan maar op vertrouwen dat een computersysteem het bij het rechte eind [82] heeft? Knippen en plakken is immers veel gemakkelijker dan zich grondig verdiepen in een onderwerp, duiding kunnen geven, hoor- en wederhoor toepassen en objectief daarover kunnen rapporteren.

Het is ongewis, maar misschien zullen matige auteurs in de toekomst wellicht suggesties krijgen om hun verhaal beter op te schrijven, bijvoorbeeld met een bepaalde verhaallijn of met toegevoegde dialogen [83]. Film- en scripschrijvers kunnen dan gemakkelijk dialogen laten genereren in hun script en daarmee inspiratie opdoen.

We zien nu op ons beeldscherm af en toe autocorrect- en autocomplete-software opduiken met suggestie voor nieuwe woorden of zinnen. In Gmail bijvoorbeeld het onderwerp van een e-mail en de begroeting. Dat lijkt dan maar kinderspel met wat we in de toekomst wellicht gaan gebruiken. Het is niet ondenkbaar dat journalisten en bloggers in de toekomst slechts nog een kladconcept schrijven van hun artikel, en dat vervolgens door een kunstmatig intelligent systeem laten voltooien, inclusief de correctie van taalfouten en de suggestie van twee of drie pakkende titels.
Ook kan GPT2-software [84] wellicht chatbots beter laten functioneren.

1.5 | DE GENERATIE VAN SYNTHETISCHE AUDIO

We zien niet alleen dat generatieve AI-software wordt gebruikt bij beeld- en tekstgeneratie, maar ook bij audio. Niet alleen zien we generatieve AI-computersystemen die muziek creëren, maar zelfs de menselijke stem laat zich door AI genereren. Sommige bedrijven zijn er al in geslaagd om synthetische stemmen te genereren die niet meer van echt te onderscheiden zijn, zoals Google's Wavenet [85] of Wellsaid [86|87]. Hierbij wordt overigens vrijwel altijd andere software gebruikt dan GAN-technologie en het creatieproces vraagt nog wel behoorlijk wat expertise.



Credit: Better Language Models and Their Implications.
https://openai.com/blog/better-language-models/

Een bekend voorbeeld van spraaksynthese betreft Lyrebird [88]. Met die software kunt u uw eigen stem opnemen op basis waarvan het systeem een vocal avatar, een digitale kloon van uw stem, genereert die u al typend op het toetsenbord van alles kunt laten zeggen. De kwaliteit van Lyrebird is nog onvoldoende, de gegenereerde stem klinkt robotachtig en is soms zelfs onduidelijk. De stip op de horizon staat ook hier echter glashelder.

Talking machines

In de nabije toekomst kunt u wellicht uw eigen stem klonen en die dan inzetten voor velerlei toepassingen. Uw kinderen selecteren bijvoorbeeld op hun tablet een voorleesboek en kunnen ervoor kiezen of ze dat door een van hun ouders of opa of oma laten voorlezen. Bedrijven zullen waarschijnlijk een synthetische stem op maat kunnen bestellen [89] die kan dienen als ideale stem voor een volledig digitale klantenservicemedewerker. Iedere klant die belt, krijgt dan een kort intakegesprek met 'dezelfde' klantenservicemedewerker. U kunt deze synthetische stem ook gebruiken in uw reclamefilmpjes op radio en televisie of als stem voor dienstmededelingen in de openbare ruimten van het bedrijf.

Het bedrijf Deepzen [90] biedt al de mogelijkheid aan om audioboeken te maken, waarbij kan worden gekozen voor een stem met gewenste juiste taal, accent en persoonlijkheid. Generatieve AI-softwaresystemen die een synthetische stem creëren, zullen het misschien mogelijk maken dat in de toekomst serieuze auteurs en beroemdheden hun stem laten klonen, zodat ze die relatief gemakkelijk kunnen inzetten om hun boek of autobiografie te laten omzetten naar een audioboek. Dat scheelt vele, vele uren doorgebracht in een studio achter een microfoon om een boek voor te lezen. Wellicht is het straks een koud kunstje voor een computer om binnen een paar seconden een geschreven boek om te zetten naar een audioboek.

De digitale stemmen van beroemdheden zullen waarschijnlijk tevens door commerciële bedrijven worden 'ingehuurd' om persberichten, inhoudelijke rapporten of websiteteksten voor te lezen met hun karakteristieke stemgeluid. *Voice cloning* van beroemdheden als commerciële bedrijfstak is erg interessant omdat het slechts in het begin een investering van tijd vraagt en beroemdheden daarna hun stem oneindig vaak en lang kunnen uitlenen aan derden zonder er zelf nog tijd in te hoeven investeren. En vanzelfsprekend laat deze *voice cloning*-technologie zich lenen voor criminele doeleinden. Als in de nabije toekomst iemands stem zich laat kopiëren zodat men die vervolgens van alles kan laten zeggen, dan biedt dat enorme mogelijkheden voor criminele activiteiten. Over de negatieve consequenties van voice cloning leest u meer in het tweede deel van dit rapport, dat ingaat op deepfake-technologie.

PROJECT REVOICE

Project Revoice [91] is een project dat de initiatiefnemer van de Ice Bucket Challenge zijn eigen stem in digitale vorm heeft teruggegeven. De Ice Bucket Challenge was een goede-doelenactie voor ALS-patiënten uit 2014. Mensen konden er bij deze actie voor kiezen om een emmer ijswater over hun hoofd te gieten óf een donatie te doen ten behoeve van ALS-patiënten. Velen deden overigens beide. Pat Quinn, de bedenker van de Ice Bucket Challenge, heeft namelijk zelf ook de neurologische aandoening ALS en is inmiddels zijn eigen stem verloren vanwege de ziekte. Met zijn ogen stuurt hij een computer aan en typt hij letter voor letter zijn zinnen. De computer las de teksten voor met een computerstem. Lyrebird, zoals we eerder zagen een bedrijf dat *voice cloning*-software maakt, haalde vele uren geluidsmateriaal van Quinns stem van het internet af. Vervolgens werd de generatieve Lyrebird-software getraind zodat het een reproductie kon maken van Pat Quinns eigen stem. De onprettige robotstem van zijn computersysteem werd vervolgens vervangen door een meer natuurlijke reproductie van zijn eigen stem. Mensen die letterlijk hun stem voor altijd kwijt zijn door fysieke beperkingen, krijgen zo toch een mogelijkheid om hun stem, weliswaar digitaal, weer terug te krijgen.

Het is niet ondenkbaar dat we in de toekomst dankzij geluidsopnamen uit het verleden de stem van iemand die al is overleden kunnen klonen. Er zijn verscheidene projecten wereldwijd die dat proberen te bewerkstelligen. Veelal gaat het om mensen van wie veel geluidsopnamen beschikbaar zijn, zoals televisieberoemdheden, politici, radio- en podcastmakers en professionele sprekers. Zij laten in hun werkzame leven namelijk vele uren audio achter op het internet. Wanneer het in de toekomst mogelijk is om een *voice cloning*-machine te trainen met dit materiaal, kunt u iemands stem terughalen en middels een tekst of een toetsenbord hem of haar nieuwe dingen laten zeggen.

LIP NAAR SPRAAK

Onderzoekers van Samsung en het Imperial College in Londen hebben een generatief AI-systeem ontwikkeld dat computervisie gebruikt voor visuele spraakherkenning op basis van lipbewegingen [92|93]. Dit video to speech-model heeft geleerd om audio (stemmen) te produceren door te kijken naar lipbewegingen van een spreker.
Het systeem is getraind met talking head videos, video's van pratende mensen. Zo heeft het een verband kunnen leggen tussen de bewegingen van de mond en het geluid dat er uitkomt. Het produceert uiteindelijk een synthetische stem op basis van de lipbewegingen van de persoon in kwestie. In de toekomst moet generatieve AI-software dus hoorbaar maken wat iemand zegt zonder dat u het hoort, maar wel ziet.
Een positieve manier van gebruik zou zijn om stomme ('stille') films uit het verleden opeens van spraak te kunnen voorzien.
En veiligheidsdiensten kunnen dergelijke lipleestechnologie inzetten om geluidloos 'af te luisteren', door op basis van enkel en alleen videobeelden te analyseren wat mensen zeggen. De keerzijde is dat de technologie kan worden misbruikt door criminelen, bijvoorbeeld bij bedrijfsspionage. Het geloofwaardig en gemakkelijk kunnen kopiëren van iemands stem of synthetische stemmen kunnen genereren op basis van lipbewegingen – het lijkt misschien nu nog sciencefiction, maar het is niet ondenkbaar dat dit de realiteit wordt. Daarvoor moeten nog flink wat AI-problemen worden getackeld, maar de stip op de horizon is gezet.

Credit: MuseNet. https://openai.com/blog/musenet/

Muziek gemaakt door AI

Het genereren van nieuwe muziek met generatieve AI-software wordt misschien nog wel gemakkelijker dan het genereren van een geloofwaardige stem. U las in dit rapport al over GPT2, het model dat was getraind door OpenAI om het volgende woord in een zin te voorspellen. Datzelfde model is ook gebruikt om muziekcomposities te genereren. Het GPT2-Musenet-model [94] kan daarbij wel tien verschillende muziekinstrumenten gebruiken en kan allerlei muziekstijlen combineren, van country tot Mozart en The Beatles. Net zoals bij de taalvariant werd dit GPT2-systeem niet expliciet geprogrammeerd met menselijk begrip van muziek, maar ontdekte het patronen van harmonie, ritme en stijl. Als trainingsmateriaal werd een database gebruikt van honderdduizenden Midibestanden. Ook kan zo'n GPT2-systeem verschillende stijlen in elkaar laten overvloeien en laten samenspelen. De resultaten zijn geloofwaardig en vaak niet van echt te onderscheiden.

Over creativiteit, intelligentie en verbeeldingskracht sprak ik met wetenschapsjournalist Bennie Mols.

INTERVIEW: BENNIE MOLS

WETENSCHAPS- EN TECHNIEKJOURNALIST, AUTEUR EN SPREKER

Hoe kijk jij aan tegen GAN-technologie?
Als wetenschapsjournalist in het digitale domein volg ik al heel lang de vernieuwingen op het gebied van artificiële intelligentie. Ik vind de ontwikkeling van GAN-technologie bijzonder interessant. Het feit dat kunstmatig intelligente systemen zelfstandig beelden, geluiden en teksten kunnen produceren is enorm opwindend. Machines krijgen nu een vorm van verbeeldingskracht.
Ik ben daarbij met name geboeid door het gebied waar mens en machine elkaar kunnen versterken. Ik heb bijvoorbeeld al heel interessante voorbeelden gezien waar bijvoorbeeld GAN-technologie wordt gebruikt voor het ontdekken van nieuwe, efficiëntere zonnecellen of batterijen. De software bedenkt nieuwe mogelijkheden waar de mens zelf helemaal niet aan had gedacht. Vervolgens onderzoekt de mens welke daarvan wel of niet goed werken.

Zijn er grote voordelen van de GAN-technologie?
Die zijn er zeker! Het gebruik van deze digitale technologie creëert bijvoorbeeld een enorme versnellingsstap in het bedenken en testen van nieuwe producten. En dat niet alleen: vergeet niet dat wij als mensen ook een beperkt denkkader hebben. Onze manier van kijken naar problemen en oplossingen zit vol met vooroordelen. Een machine hoeft deze vooroordelen niet te hebben en kan dus vrijuit nieuwe dingen bedenken. Neem bijvoorbeeld de go-computer van Deepmind Technologies, AlphaGo. AlphaGo deed in 2016 enkele zetten waarvan de beste menselijke spelers dachten dat het domme zetten waren, maar het bleken geniale zetten te zijn die menselijke vooroordelen over het go-spel genadeloos blootlegden.
Ik kan me goed voorstellen dat in de farmacie en de materiaalkunde deze technologie gaat worden toegepast om veelbelovende moleculen te bedenken en produceren. Die nieuwe ideeën kun je dan in het laboratorium gaan testen. Stel dat er van de dertig suggesties die de computer bedenkt vijf hele goede oplossingen bijzitten. Dat versnelt het doen van nieuwe ontdekkingen.
Op een meer fundamenteel niveau vind ik deze technologie veelbelovend omdat er een nieuw hoofdstuk geschreven lijkt te worden op het gebied van unsupervised learning. Waar kunstmatig intelligente systemen op dit moment vooral getraind worden met gelabelde data, het zogenoemde supervised learning, kun je GAN-technologie gebruiken om computers te laten leren van ongelabelde, ongestructureerde data.

Kun je daar een voorbeeld van noemen?
Nou, denk bijvoorbeeld aan een GAN-systeem dat zelf een volledige simulatie creëert van een driedimensionale wereld. Vervolgens kan bijvoorbeeld een zelfrijdende auto deze nieuwe simulatie gebruiken als trainingsmateriaal. Het GAN-systeem kan dan onnoemelijk veel denkbeeldige scenario's creëren waarmee de software van de zelfrijdende auto vervolgens moet leren omgaan. GAN-systemen kunnen dus heel goed simulaties maken waar andere slimme machines van kunnen leren. Door deze grote hoeveelheid nieuw leermateriaal hoop ik eigenlijk dat robots en zelfrijdende auto's veel beter gaan snappen hoe onze omgeving

eruit ziet. Dat ze zelf wetmatigheden in de fysieke wereld gaan ontdekken. Dat robots daardoor ook veel beter kunnen omgaan met situaties die niet vaak voorkomen.

Dat snap ik, maar hoe garanderen we dat dit nieuwe trainingsmateriaal lijkt op onze fysieke realiteit? Want wanneer een GAN-systeem honderden uren videomateriaal per dag creëert waarvan sommige onderdelen buiten de realiteit staan en soms zelfs hallucinogeen te noemen zijn, hoe houden we daar als mens dan zicht op?
Dat is inderdaad een interessante vraag. Kleine kinderen leren in de eerste paar jaren van hun leven om de wereld om hen heen te begrijpen in natuurkundige en psychologische modellen. Ze leren intuïtief hoe de zwaartekracht op voorwerpen inwerkt. Ze leren intuïtief de bedoelingen van andere mensen te begrijpen. Ik denk dat we in computers ook natuurkundige en psychologische modellen moeten bouwen waar GAN-systemen zich op gaan baseren. Idealiter zouden computers natuurlijk zelf zulke modellen moeten leren maken. De crux is dat computers niet alleen veel data nodig hebben om de wereld te begrijpen, maar ook modellen. Dat wordt in deze tijd van big data maar al te vaak vergeten.
Zulke natuurkundige of psychologische modellen werken dan als een soort scheidsrechter, een soort VAR voor de creaties van het GAN-systeem. De VAR houdt het GAN-systeem in de gaten en waarschuwt wanneer een GAN iets bedenkt dat niet strookt met het natuurkundige of psychologische model van de omgeving. Zo garanderen we dat GAN-systemen niet ongewild een loopje nemen met de werkelijkheid en onwenselijke of zelfs schadelijke dingen bedenken.

Je bent duidelijk enthousiast, maar kent deze software ook beperkingen?
Jazeker, die zijn er natuurlijk ook. Omdat de twee delen van de software, de generator en de discriminator, worden getraind met dezelfde dataset, krijg je eigenlijk altijd een mengeling van bestaande stijlen. Zo bedenkt een GAN nooit iets volkomen nieuws. Een GAN-systeem dat is getraind op foto's van honden kan bijvoorbeeld wel een hond bedenken met een geheel nieuw, niet bestaand patroon van stippen, maar het bedenkt niet ineens een totaal nieuw dier. Iets echt volkomen nieuws, ja zelfs iets geniaals bedenken, dat lukt deze machines nog niet. Picasso, Van Gogh en Dalí bedachten een geheel nieuwe schilderstijl. Einstein bedacht geheel nieuwe natuurkunde. Zover zijn machines nog lang niet, behalve in zeer beperkte domeinen als schaken of go. Voorlopig kunnen we machines heel goed gebruiken om onze eigen creativiteit te verrijken: mens en machine die samen creatiever zijn dan elk van beide afzonderlijk.

Wat zie je dan vooral het domein van de mens in dit geheel?
Wat de mens bijvoorbeeld heel goed kan is om kennis uit verschillende domeinen te combineren in een ander domein. Denk bijvoorbeeld aan een muzikant die geïnspireerd is geraakt door een film en dat vertaalt naar muziek. Of dat iemand die heel veel ervaring heeft op het gebied van tennis ook relatief gemakkelijk biljart kan spelen omdat hij of zij snapt hoe een bal zich verplaatst, hoeveel kracht je daarvoor nodig hebt en welke spin je aan een bal kunt geven. Een goed balgevoel komt in alle balsporten van pas.

Machine-intelligentie is als een laser: superkrachtig, maar op een heel beperkt gebied. Menselijke intelligentie is meer als een gloeilamp: niet superkrachtig, maar het licht schijnt wel alle kanten op. De menselijke intelligentie is voorlopig nog algemener dan de machine-intelligentie. Ook moet ik zeggen dat ik nog niet heel erg onder de indruk ben van dit soort GAN-

systemen als het gaat om genereren van teksten, zeker als het gaat om fictie. Ze kunnen simpele nep-nieuwsberichten maken, maar geen lange, consistente essays of verhalen. Om echt begrip te hebben van taal, moet je namelijk veel snappen van de wereld om ons heen. Dit soort machines kent niet de betekenis, de context van de wereld om ons heen.

Hoe kijk je naar GAN-technologie en deepfakes vanuit je rol als journalist?
Als journalist maak ik me natuurlijk zorgen dat de scheidslijn tussen nep en echt steeds dunner wordt. Ik ben getraind om kritisch naar nieuwsberichten te kijken, maar de gewone burger is dat natuurlijk veel minder. Die reageert wat sneller op emotioneel nieuws en daardoor verspreidt nepnieuws zich razendsnel op sociale media. We moeten ook niet doen alsof nepnieuws een nieuw fenomeen is. Het heeft altijd bestaan in de geschiedenis. Alleen zorgt nieuwe generatieve AI-software ervoor dat het gemakkelijker te maken en te verspreiden is. Dat je iedere vorm van content kunt manipuleren, staat voor mij als een paal boven water. Ik vind dat die constatering een direct appel doet op ons als samenleving, maar ook op bedrijven en de sociale media. We moeten een soort digitale zonnebrandcrème voor deepfake ontwikkelen. En daarbij concreet werken aan oplossingen.

Hoe zie je dat voor je?
GAN-technologie gaat gebruikt worden om allerlei typen van verdraaide informatie te creëren. Dat wordt de nieuwe realiteit. Om ons daartegen te wapenen, zijn er op allerlei niveaus acties nodig.
Overheden en media zouden burgers bewust kunnen maken van het gemak waarmee desinformatie wordt gemaakt en verspreid. Burgers zelf moeten kritisch zijn op de informatie die ze tot zich nemen: Wie is de afzender? Is deze betrouwbaar? Welk belang heeft deze afzender bij dat bericht? Hoe speelt een bericht in op de emotie?
En sociale media moeten hun verantwoordelijkheid nemen door de publicatie en verspreiding van desinformatie tegen te gaan. Ze kunnen zich al lang niet meer verschuilen achter de opvatting dat ze alleen maar een neutraal doorgeefluik zijn.

Deels gaat nieuwe technologie ons helpen het probleem van desinformatie te tackelen. Kunstmatig intelligente systemen moeten, net zoals spamfilters dat deden vanaf de jaren negentig, herkennen welke content ‘deepfake’ is en dat in onze browser blokkeren of zichtbaar markeren. Maar het allerbelangrijkste is dat we wetgeving, normen en waarden en instituties nodig hebben die misbruik van technologie voorkomen. Dat is veel belangrijker dan de technologie zelf.

Het interessante is dat landen met de meest hoogwaardige technologie grofweg gezegd ook het hoogste scoren op democratie en mensenrechten. Je kunt dus niet zeggen dat technologie de boosdoener is. Mensen zijn veel gevaarlijker dan technologie. Nederland heeft veel meer geavanceerde technologie om burgers in de gaten te houden dan Noord-Korea. Maar onze wetgeving, normen en waarden en instituties zorgen ervoor dat die technologie daarvoor niet wordt gebruikt. Noord-Korea heeft veel minder hoogwaardige technologie, maar houdt haar burgers veel scherper in de gaten en produceert veel meer desinformatie. We moeten ons daarom meer zorgen maken over wetgeving, normen en waarden en instituties dan om de technologie alleen.

Zie je hierin nog een rol voor traditionele journalistiek?
Jazeker, wanneer ik deze ontwikkeling van een afstand bekijk, zou het ook wel zo kunnen zijn dat gerenommeerde kranten juist meer waardering en autoriteit

krijgen van het publiek dan sociale media. Juist omdat de lezer de redactie van zo'n krant vertrouwt als een baken van feitelijkheden binnen de grote hoeveelheid van desinformatie. Hoe moeilijker het wordt om snel in te schatten of de content zelf betrouwbaar is, hoe belangrijker het wordt om de afzender van de content te kunnen vertrouwen. Ook zou het wel eens kunnen zijn dat voor ons de offline-wereld weer belangrijker wordt omdat de vervuiling in de online-wereld groter is.

Als je kijkt naar het jaar 2030 door de lens van GAN-technologie, hoe ziet de wereld er dan uit?
In het komende decennium gaan we met vallen en opstaan leren hoe we ons moeten wapenen tegen deepfake, zoals we dat ook hebben gedaan met spam in onze e-mail. Ik denk dat we vooral de vruchten gaan plukken van de mooie en interessante mogelijkheden die GAN-technologie te bieden heeft. Ik noemde al even de voorbeelden uit de wetenschap voor het ontdekken van nieuwe moleculen. Maar denk ook aan kunstenaars, modeontwerpers en game-ontwerpers. En ik verwacht veel van de nieuwe mogelijkheden van robots om onze fysieke wereld te snappen omdat GAN-systemen heel veel trainingsmateriaal voor hen kunnen produceren. In de afgelopen tien jaar zijn machines goed geworden in het herkennen van patronen. In de komende tien jaar gaan machines ook goed worden in het creëren van patronen. Eindelijk krijgen machines een vonk van verbeelding.

Generatieve AI-software, waarin kunstmatig intelligente systemen bewegend beeld, tekst en audio van hoge kwaliteit kunnen creëren, blijkt een fascinerende 'vonk' van digitale verbeeldingskracht. In de toekomst zullen we de zoete vruchten gaan plukken van generatieve AI-software met name GAN software. Er is echter is ook een keerzijde aan deze ontwikkeling, waarbij de verbeeldingskracht van machines met een negatieve doelstelling worden gebruikt. Creatie van synthetische media vanuit kwalijke intenties. Dat aspect belicht ik het tweede deel van dit rapport, over deepfake technologie.

DEEPFAKE TECHNOLOGIE: THE INFOCALYPSE

DEEPFAKE TECHNOLOGIE: THE INFOCALYPSE

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 [95].

Een recenter voorbeeld betreft het gerenommeerde maandblad *National Geographic* dat in februari 1982 op de cover een afbeelding van de piramiden 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 [96] 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 [97] van deepfake-technologie is inmiddels beschikbaar als app op de smartphone.

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

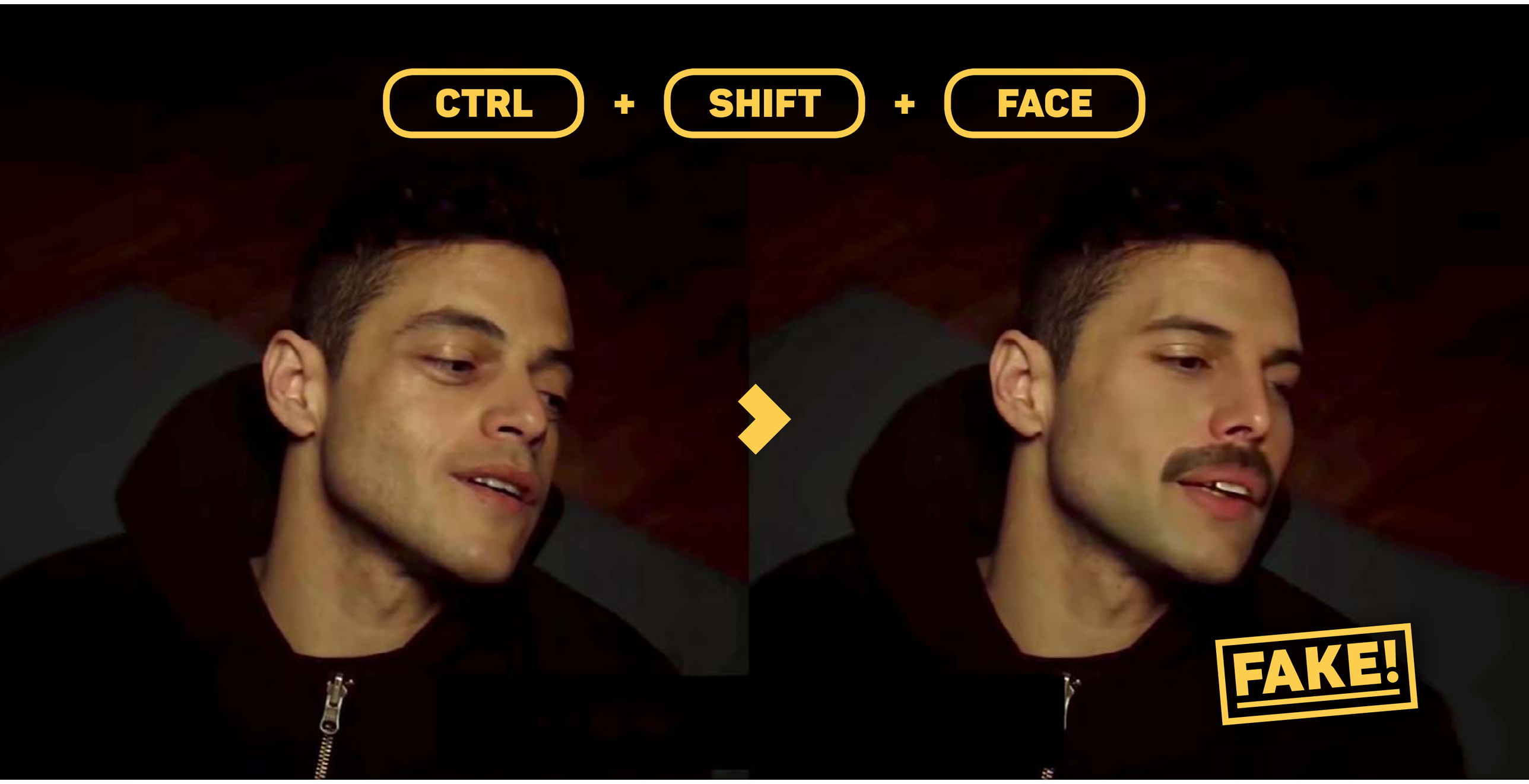
De meer serieuze en gerenommeerde nieuwsbronnen leggen vaak de focus op de mogelijke negatieve gevolgen van de nieuwe *deepfake*-technologie, bijvoorbeeld voor de journalistiek of bij toepassing voor criminele doeleinden. Veel andere nieuwsbronnen concentreren zich vooral op humoristische filmpjes vol 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. Onder deepfake-technologie wordt in dit rapport verstaan: generatieve AI-software die in staat is om digitale content te creëren zoals bijvoorbeeld gezichten, stemmen, teksten, beelden, bewegende mensen en geluiden. In het eerste gedeelte van dit rapport zijn daar meerdere voorbeelden van geschetst. Generatieve AI-software die synthetische media genereren en zo de scheidslijn tussen echt en nep, tussen de fysieke wereld en de digitale wereld steeds dunner maken. Deepfake gaat uit van negatieve intenties bij het gebruik van deze *generative AI technology*.

In dit onderdeel van het rapport leest u uitgebreid over deze deepfake-ontwikkeling. Wat is het? Waar kijken we naar? Hoe kan het dat het zo populair is geworden? Wat zijn de mogelijke risico's? Valt dit probleem volledig te tackelen? 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 [98], 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?

Daar waar het eerste gedeelte van dit rapport de voordelen van generatieve AI-softwaresystemen beschrijft en daarbij aangeeft wat er zoal voor nieuwe mogelijkheden zijn, ligt het accent in dit tweede gedeelte meer op mogelijke negatieve gevolgen van generatieve AI-software in de vorm van deepfakes. Hierbij geef ik helder aan wat er op dit moment reeds mogelijk is en speculeer ik voorzichtig over de mogelijkheden van de toekomst. Net zoals in het eerste gedeelte van het rapport blijf ik hierbij binnen de scenario's van waarschijnlijkheid.

2.1 | SOORTEN DEEPFAKES

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, meestal generative adversarial networks, 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 [99]. 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 [100] daarvan leveren de hier weergegeven afbeelding uit een video van Obama. 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

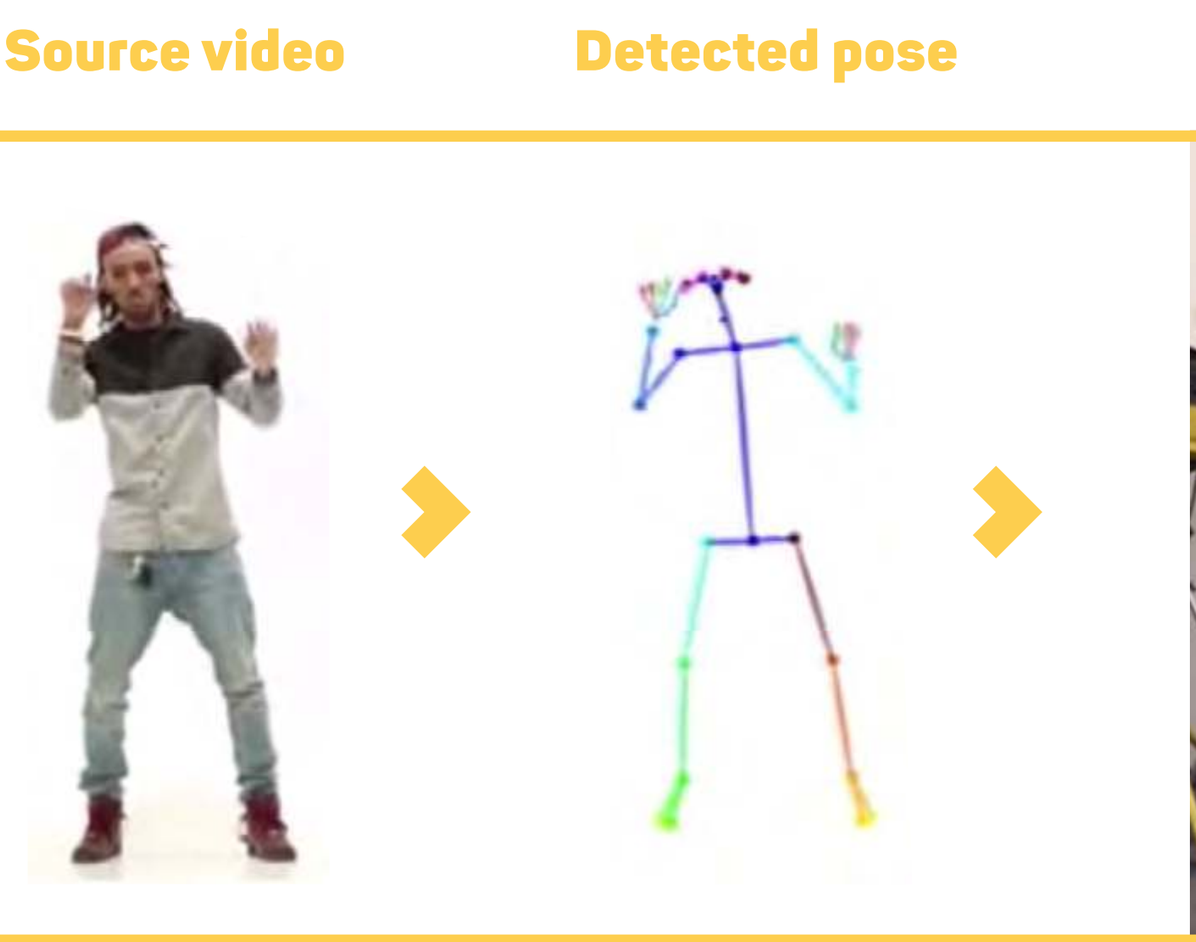
Ook in een inmiddels van internet verwijderde video van Kim Kardashian, waarbij zij zogenaamd toegaf haar volgers te manipuleren in ruil voor geld, maakte gebruik van deze *lipsync deepfake-technologie* [101]. Mede daardoor is het resultaat verbluffend goed.

Edited image from source: Bill Posters (@bill_posters_uk) • Instagram-foto's en -video's https://www.instagram.com/p/ByKg-uKlP4C/

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" [102], 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.



PERSONALIZED AVATAR CREATION

Onderzoekers van de Universiteit van Heidelberg is het gelukt om *fullbody deepfakes* [103] te maken. 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 [104], 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 [105] waren vertraagd zodat het leek alsof zij dronken was. De video [106] had vele miljoenen *views*. Pelosi heeft Facebook nog verzocht de video te verwijderen, maar dat werd geweigerd door het socialemediaplatform [107]. Op een later moment gaf Facebook-oprichter Mark Zuckerberg [108] 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 [109] werd van een deepfake-video, is onbekend. 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" [110] 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

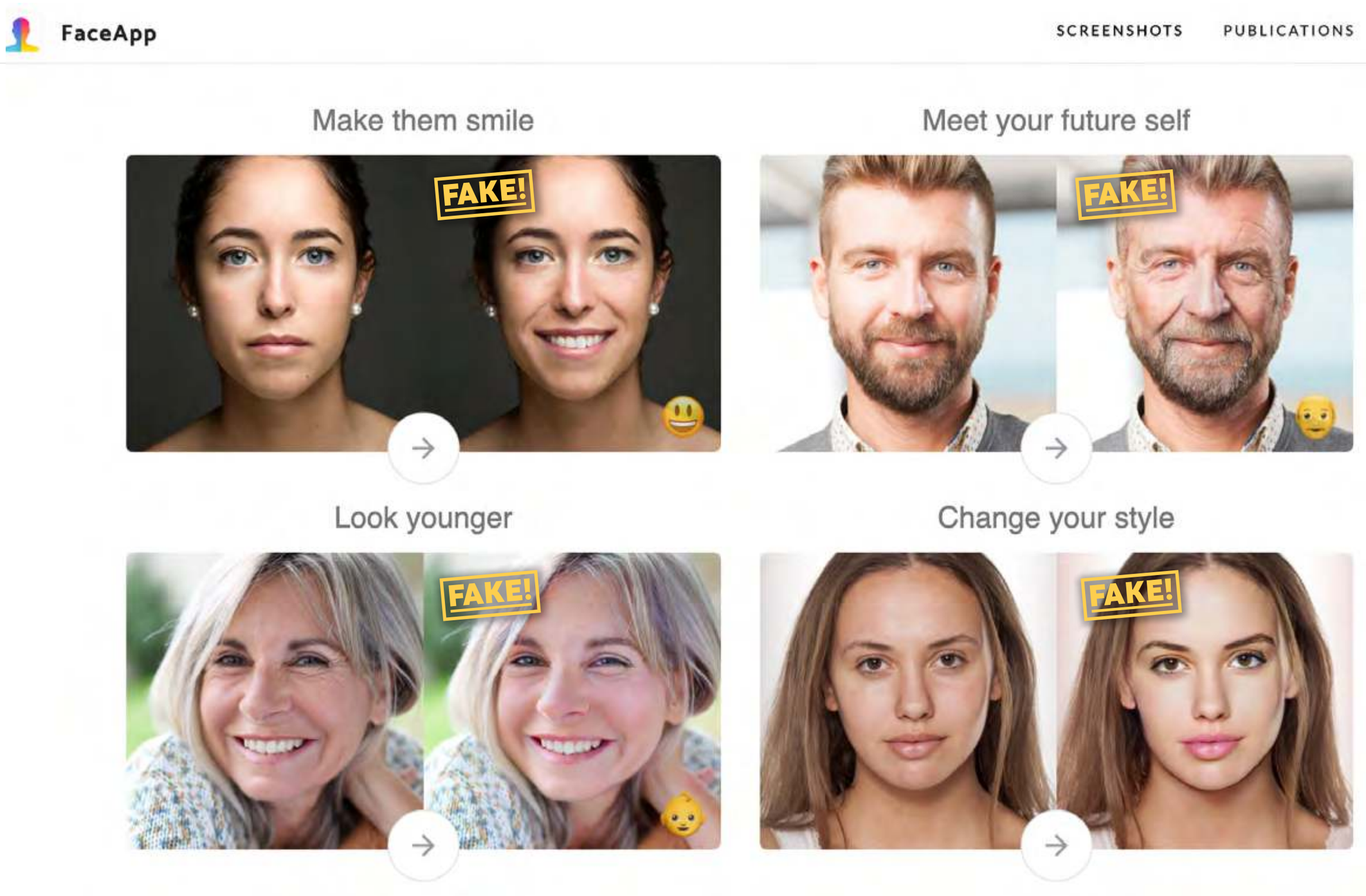
Deepfake-video's zijn, in vergelijking met een paar jaar geleden, relatief eenvoudig om te maken [111] en dat gemak gaat in de toekomst alleen maar toenemen. Op het internet kunt u moeiteloos een handleiding vinden voor het maken van een deepfake-video en u heeft daarvoor tegenwoordig al geen supercomputer 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- [112] te maken, en aan de andere kant is het moeilijk. 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 [113]. 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 [114] en of mensen zich er wel bewust van waren dat dit gebeurde. Deze gevoeligheid komt vanzelfsprekend voort uit de beïnvloeding door Rusland van de Amerikaanse presidentsverkiezingen van 2016 [115]. 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 [116] in de hand werkte.

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 [117], zeker wanneer die nieuw en negatief van aard is.

Ook zien we dat de traditionele media continu worden aangevallen door de Amerikaanse president [118] *("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 [119]. "
- James Vincent, techjournalist The Verge

2.2 | 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.

2.3 | 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

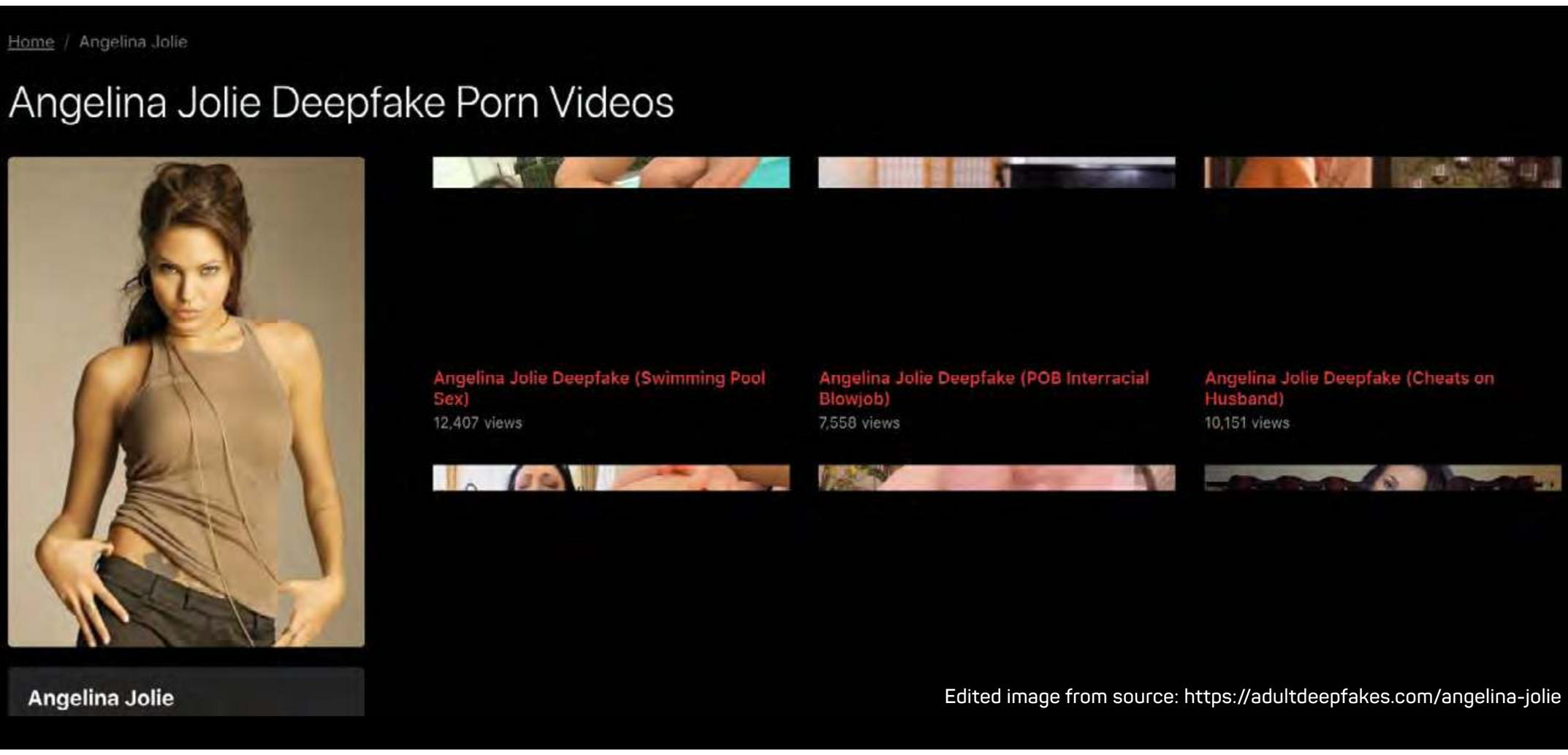
Het is niet ondenkbaar dat politici, journalisten, buitenlandse militairen, directeuren van grote bedrijven [120], klokkenluiders en financieel verantwoordelijken in de toekomst te maken krijgen met chantage met deepfake-video's. Op een dag kunnen ze dan onderstaande e-mail in hun inbox aantreffen:

" 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 [121], 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 [122] laat zien. In de recente Nederlandse geschiedenis werd ex-NOS-nieuwslezeres Dionne Stax slachtoffer van een gemanipuleerde pornografische video [123].

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 [124 | 125] 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 [126] online, aangeprezen als een *AI Powered X-Ray App* waarmee het mogelijk was om foto's van geklede vrouwen te veranderen naar naaktfoto's [127]. 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 [128] 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 [129] 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 [130] 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 [131], 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.

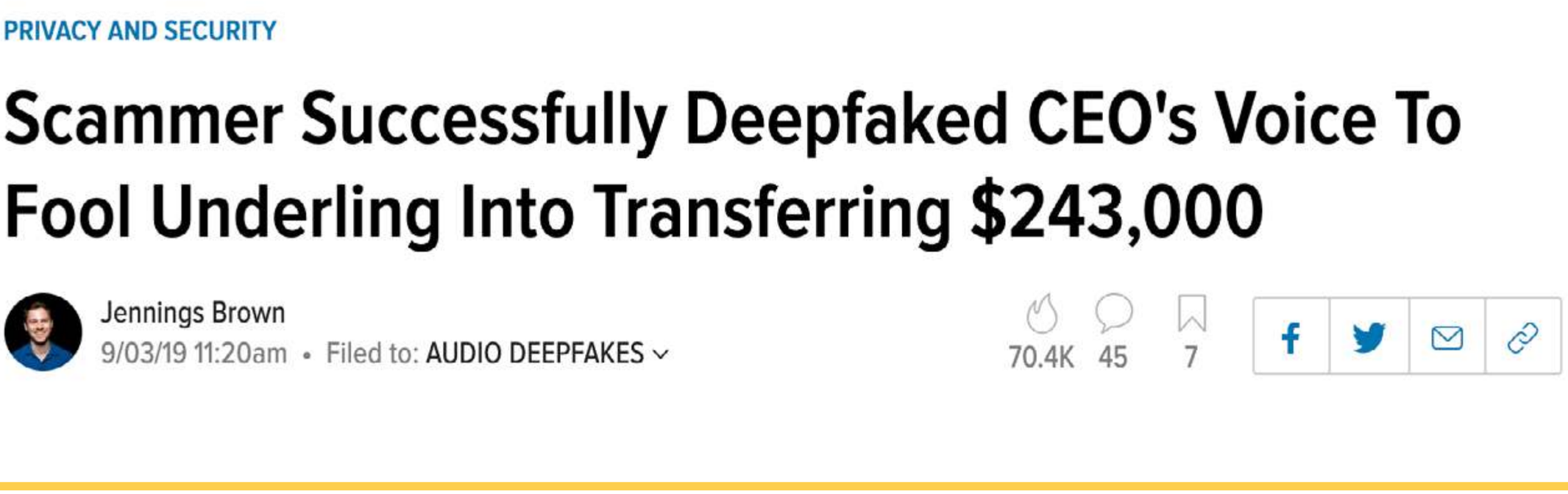
Ligt hier voor jullie bij de NOS dan ook een kans?
Dat denk ik wel. Kijk: Twitter is een prachtig medium om de allereerste nieuwsfeiten van een evenement naar je toe te halen. Er is niks sneller in de wereld op dat gebied. Maar al vrij snel begint de speculatie. En daar wordt het dus onbetrouwbaar. Ik 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.

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.

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.



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 [132].

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

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.

Zit er ook een voordeel aan deze GAN-technologie?
Natuurlijk zie ik ook wel wat meer onschuldig gebruik: dat je in de toekomst op Instagram nóg interessantere filters kunt gebruiken om jezelf nóg mooier te maken. Dat je jezelf helemaal kunnen modelleren zoals je graag zou willen zijn. Ik kan me voorstellen dat deze synthetische wereld een steeds sterkere magneet is voor met name een jongere doelgroep.
En natuurlijk zie ik nog wel een paar kleine voordelen, zoals virtualisatie voor bijvoorbeeld onderwijs of medische en therapeutische toepassingen, maar bij de opkomst van deze technologie vraag ik me wel af: ‘is dit nou wel zo’n goed idee’?

2.4 | 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: de kwaal is dan tegelijkertijd het medicijn. Generatieve AI-systemen die beelden of teksten genereren, kunnen namelijk tevens helpen ze op te sporen, zoals het Grover-systeem dat eerder in dit rapport werd genoemd.

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 [133] 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 [134]. 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 [135] 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 [136]. 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

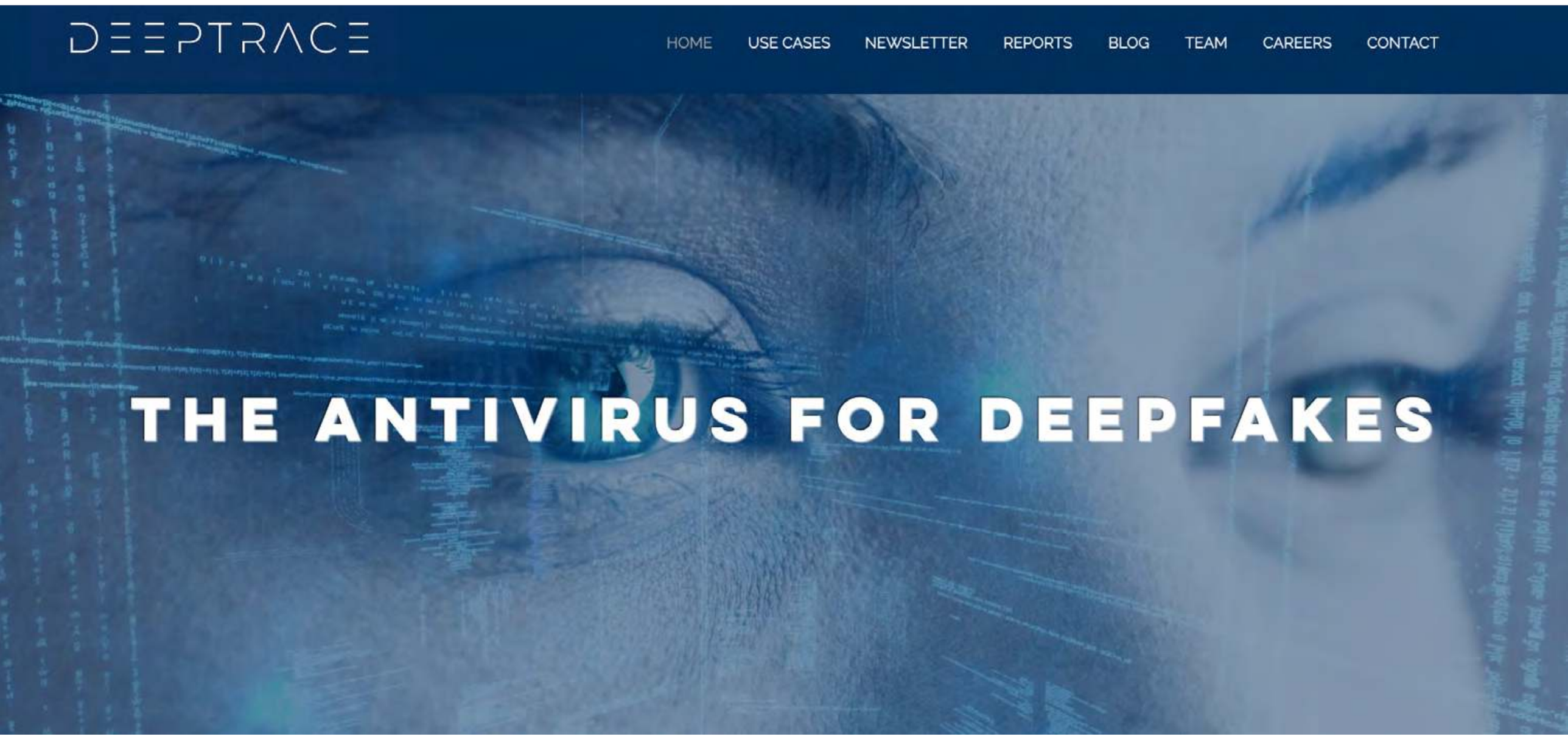
Life logging is een manier van handelen waarbij software het dagelijkse leven nauwgezet registreert. Een vorm van controlled capture dat op sommige aspecten geautomatiseerd gaat vastleggen wat iemand aan het doen is. Belangrijke politici, beroemdheden, journalisten en zakenlieden zullen in de toekomst wellicht intensief daarmee hun dagelijks leven vastleggen. Wanneer ze dat doen, beschikken ze immers altijd over een ijzersterk alibi wanneer er bijvoorbeeld belastende videobeelden [137]

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 [138] en Deepfact [139].



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 [140] 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' [141]. 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' [142] 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 [143] hebben getoond, geen bereik meer geven. In de Amerikaanse politiek [144] [145] 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 [146] 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 [147]. Voor journalisten blijft vanzelfsprekend het traditionele bronnenonderzoek daarom van groot belang. En de inzet van de eigen waarneming [148].

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 [149 | 150 | 151 | 152].

- 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 [153], werd in december 2018 ingediend en de Deepfakes Accountability Act [154] volgde in juni van 2019. In verschillende staten, waaronder Virginia [155], 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 [156], 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 [157|158]. 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* [159] sterk aanbevelen.

3 | CONCLUSIE

Dit rapport over ‘machines met verbeeldingskracht’ beschrijft de opkomst van generatieve AI-software: een verzamelnaam voor verschillende AI-technologieën die alle gemeen hebben dat ze nieuwe digitale content kunnen creëren en daarmee soms nieuwe ideeën, hypotheses en invalshoeken scheppen. In het rapport worden verschillende soorten generative AI-software systemen uitgediept die bijvoorbeeld video’s, audio, teksten, moleculaire structuren, stedenbouwkundige plannen, en zelfs volledig virtuele werelden creëren. In dit rapport is er bovengemiddeld veel aandacht voor generative adversarial networks, een relatief nieuwe AI-technologie die gespecialiseerd is in het creëren van een kunstmatige realiteit vol met nieuwe inhoud, ideeën en invalshoeken. Mede door deze technologie wordt de scheidslijn tussen de fysieke wereld en de digitale wereld steeds dunner.

Software heeft in de afgelopen jaren steeds meer menselijke vaardigheden van ons overgenomen zoals kijken, luisteren, spreken en lezen. En nu worden de eerste veelbelovende stappen gezet op het gebied van de verbeeldingskracht. Generatieve AI-software wordt dan onze creatieve assistent en onze uitvinder. Maar zoals bij iedere technologische trend kent ook deze ontwikkeling haar nadelen. Machines met verbeeldingskracht kunnen namelijk ook ingezet worden voor negatieve doeleinden. Bij deepfake-technologie wordt generatieve AI-software gebruikt om met behulp van zogenaamde synthetische media (met AI gecreëerde of gemodificeerde media) onze waarneming van de wereld op een negatieve manier te beïnvloeden. Je kunt met generatieve AI-software volledig nieuwe geloofwaardige video’s creëren, in video’s onderling gezichten verwisselen, iemands stem namaken, geloofwaardige teksten genereren en objecten wegpoetsen alsof ze nooit hebben bestaan. Zo kun je mensen dingen laten doen die ze niet gedaan hebben of dingen laten zeggen die ze niet gezegd hebben. We naderen een tijdperk waarin we online onze ogen en oren niet meer kunnen vertrouwen. Toename van risico’s als sociale onrust, geopolitieke spanningen, chantage, 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. In dat licht bekeken creëert deepfake-technologie vanaf nu voor ons een nieuwe realiteit. In het algemeen creëert generatieve AI-software een nieuwe digitale wereld met nieuwe content, nieuwe ideeën en nieuwe gevaren. Nu de scheidslijn tussen de virtuele en fysieke wereld steeds dunner wordt en illusie en realiteit steeds meer in elkaar overlopen, is het onderscheid steeds lastiger te maken tussen echt en nep. Machines met verbeeldingskracht creëren zo een kunstmatige 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=u5Ez8OEyUqo
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 **Home - Affectiva: Affectiva.**
https://www.affectiva.com/

3 **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

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

5 **These Videos Of Trump Are 'Deepfakes' - YouTube.**
https://www.youtube.com/watch?v=Ws509WASoHg

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

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

9 **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/

10 **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

11 **Verbeeldingskracht – Wikipedia**
https://nl.wikipedia.org/wiki/Verbeeldingskracht

12 **A Beginner's Guide to Generative Adversarial Networks (GANs) – Skymind**
https://skymind.ai/wiki/generative-adversarial-network-gan

13 **Meet The Man Who Makes Facebook's Machines Think**
https://www.buzzfeednews.com/article/alexkantrowitz/meet-the-man-who-makes-facebooks-machines-think

14 **An intuitive introduction to Generative Adversarial Networks (GANs)**
https://www.freecodecamp.org/news/an-intuitive-introduction-to-generative-adversarial-networks-gans-7a2264a81394/

15 **Generative Adversarial Networks**
https://arxiv.org/abs/1406.2661

16 **An intuitive introduction to Generative Adversarial Networks (GANs)**
https://www.freecodecamp.org/news/an-intuitive-introduction-to-generative-adversarial-networks-gans-7a2264a81394/

17 **A Beginner's Guide to Generative Adversarial Networks (GANs) – Skymind**
https://skymind.ai/wiki/generative-adversarial-network-gan

18 **A Beginner's Guide to Generative Adversarial Networks (GANs) | Skymind**
https://skymind.ai/wiki/generative-adversarial-network-gan

19 **Amazing This Person Does Not Exist**
https://thispersondoesnotexist.com/

20 **Amazing AI Generates Entire Bodies of People Who Don't Exist**
https://futurism.com/ai-generates-entire-bodies-people-dont-exist

21 **These cars do not exist - they where all were created by StyleGAN.**
https://www.youtube.com/watch?v=0LZ3-ZJwSu4

22 **The Age of Imaginative Machines.**
https://hubpages.com/technology/The-Coming-Democratization-of-Art-Animation-and-Imagination

23 **AI & Creativity: BigGAN as a creative engine – Merzazine – Medium**
https://medium.com/merzazine/biggan-as-a-creative-engine-2d18c61e82b8

24 **The Next Leap: How A.I. will change the 3D industry – Andrew Price – YouTube**
https://www.youtube.com/watch?feature=youtu.be&v=FlgLxSLsYWQ

25 **Generative Adversarial Networks, an episode from Colin Wright on Spotify**
https://open.spotify.com/episode/3Fo64rxx9W5MYK4upmsYmu?si=KacroBL6RXunQEanjsddcg

26 **Synthesizing Audio with Generative Adversarial Networks**
https://towardsdatascience.com/synthesizing-audio-with-generative-adversarial-networks-8e0308184edd

27 **MuseNet**
https://openai.com/blog/musenet/

28 **[1812.04948] A Style-Based Generator Architecture for Generative Adversarial Networks**
https://arxiv.org/abs/1812.04948

29 **A Style-Based Generator Architecture for Generative Adversarial Networks – YouTube**
https://www.youtube.com/watch?v=kSLJriaOumA

30 **This Person Does Not Exist**
https://thispersondoesnotexist.com/

31 **Deepfact Quiz**
https://deepfact.3duniversum.com/quiz

32 **Edmond de Belamy – Wikipedia**
https://en.wikipedia.org/wiki/Edmond_de_Belamy

33 **Can Artificial Intelligence be creative?**
https://goldmund-wyldebeast-wunderliebe.nl/wp-content/uploads/2018/11/Article-long.pdf

34 **AI Art at Christie's Sells for $432,500 – The New York Times**
https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html

35 **How three French students used borrowed code to put the first AI portrait in Christie's – The Verge**
https://www.theverge.com/2018/10/23/18013190/ai-art-portrait-auction-christies-belamy-obvious-robbie-barrat-gans

36 **CycleGAN: Software that can generate photos from paintings, turn horses into zebras, perform style transfer, and more.**
https://github.com/junyanz/CycleGAN

37 **A Gentle Introduction to CycleGAN for Image Translation**
https://machinelearningmastery.com/what-is-cyclegan/

38 **Toward Multimodal Image-to-Image Translation**
https://junyanz.github.io/BicycleGAN/

39 **Toward Multimodal Image-to-Image Translation – YouTube**
https://www.youtube.com/watch?v=JvGysD2EFhw

40 **Nvidia-research-mingyuliu**

http://nvidia-research-mingyuliu.com/gaugan/

41 **GauGAN: Changing Sketches into Photorealistic Masterpieces - YouTube**
https://www.youtube.com/watch?v=p5U4NgVGAwg&t=0s

42 **NVIDIA**
http://nvidia-research-mingyuliu.com/gaugan/

43 **AI Can Now Fix Your Grainy Photos by Only Looking at Grainy Photos - NVIDIA**
https://news.developer.nvidia.com/ai-can-now-fix-your-grainy-photos-by-only-looking-at-grainy-photos/

44 **Research at NVIDIA: AI Reconstructs Photos with Realistic Results - YouTube**
https://www.youtube.com/watch?v=gg0F5JjKmhA

45 **Context Encoders: Feature Learning by Inpainting**
https://people.eecs.berkeley.edu/~pathak/context_encoder/#extraResults

46 **Feature Learning by Image Inpainting using GANs**
https://github.com/pathak22/context-encoder

47 **Few-Shot Adversarial Learning of Realistic Neural Talking Head Models**
https://arxiv.org/abs/1905.08233v1

48 **Synthesia**
https://www.synthesia.io/

49 **Behind the Scenes: Dali Lives**
https://www.youtube.com/watch?v=BIDaxl4xqJ4

50 **Video-to-Video Synthesis**
https://arxiv.org/pdf/1808.06601.pdf

51 **NVIDIA Invents AI Interactive Graphics**
https://news.developer.nvidia.com/nvidia-invents-ai-interactive-graphics/

52 **The First Interactive AI Rendered Virtual World – YouTube**
https://www.youtube.com/watch?v=ayPqjPekn7g

53 **Autodesk News GAN Volkswagen**
https://adsknews.autodesk.com/news/autodesk-volkswagen-generative-design-electric-showcase-vehicle

54 **Inside the world of AI that forges beautiful art and terrifying deepfakes – MIT Technology Review**
https://www.technologyreview.com/s/612501/inside-the-world-of-ai-that-forges-beautiful-art-and-terrifying-deepfakes/

55 **Imitating people's speech patterns precisely could bring trouble - Cloning voices**
https://www.economist.com/science-and-technology/2017/04/20/imitating-peoples-speech-patterns-precisely-could-bring-trouble

56 **DataGrid Model generation AI – YouTube**
https://www.youtube.com/watch?v=8siezzLXbNo&amp=&t=0s

57 **CGI Agency - About**
https://www.cgiagency.fr/

58 **Soul Machines**
https://www.soulmachines.com/

59 **Solutions — FaceMe - Digital Humans**
https://www.faceme.com/our-solutions

60 **Facebook Can Make VR Avatars Look—and Move—Exactly Like You – WIRED**
https://www.wired.com/story/facebook-oculus-codec-avatars-vr/

61 **Say hello to Mica - Magic Leap's Mixed Reality AI – Digital Bodies**
https://www.digitalbodies.net/mixed-reality/say-hello-to-mica-magic-leaps-mixed-reality-ai/

62 **KFC's 'Virtual Influencer Colonel' is pretty damn hot**
https://mashable.com/article/kfc-virtual-influencer-colonel/

63 World's first AI news anchor unveiled in China | World news | The Guardian
https://www.theguardian.com/world/2018/nov/09/worlds-first-ai-news-anchor-unveiled-in-china

64 Meet The World's First Female AI News anchor, Xin Xiaopeng
https://interestingengineering.com/meet-the-worlds-first-female-ai-news-anchor

65 Miquela (@lilmiquela) – Instagram
https://www.instagram.com/lilmiquela/?hl=en

66 BLAWKO (@blawko22) – Instagram
https://www.instagram.com/blawko22/

67 Bermuda (@bermudaisbae) – Instagram
https://www.instagram.com/bermudaisbae/

68 The Diigitals
https://www.thediigitals.com/about

69 Digital Human Services – Photoreal Digital Doubles – Eisko
https://www.eisko.com/

70 DA-GAN: Instance-level Image Translation by Deep Attention Generative Adversarial Networks (with Supplementary Materials)
https://arxiv.org/abs/1802.06454

71 StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks
https://arxiv.org/abs/1612.03242

72 Image Synthesis From Text With Deep Learning
https://www.youtube.com/watch?v=rAbhypxs1qQ&

73 Promethean AI
https://www.prometheanai.com/

74 Deep Visual-Semantic Alignments for Generating Image Descriptions
https://arxiv.org/abs/1412.2306

75 Talk to Transformer
https://talktotransformer.com/

76 Grover - A State-of-the-Art Defense against Neural Fake News
https://grover.allenai.org/

77 Defending Against Neural Fake News
https://rowanzellers.com/grover/

78 Defending Against Neural Fake News
https://arxiv.org/abs/1905.12616

79 Grover - A State-of-the-Art Defense against Neural Fake News
https://grover.allenai.org/

80 Autocompletion with deep learning | TabNine
https://tabnine.com/blog/deep/

81 The Next Five Years of Synthetic Media: A Slice of Tomorrow's Society in 2024 – HubPages
https://hubpages.com/technology/The-Next-Five-Years-of-Synthetic-Media-A-Near-Future-of-AI-Generated-Music-Animations-Stories-and-More

82 The Next Five Years of Synthetic Media: A Slice of Tomorrow's Society in 2024 – HubPages
https://hubpages.com/technology/The-Next-Five-Years-of-Synthetic-Media-A-Near-Future-of-AI-Generated-Music-Animations-Stories-and-More

83 The Next Five Years of Synthetic Media: A Slice of Tomorrow's Society in 2024 | HubPages
https://hubpages.com/technology/The-Next-Five-Years-of-Synthetic-Media-A-Near-Future-of-AI-Generated-Music-Animations-Stories-and-More

84 Hello, It's GPT-2 - How Can I Help You?

https://arxiv.org/pdf/1907.05774.pdf

85 **Wavenet: a generative model for raw audio.**
https://arxiv.org/pdf/1609.03499.pdf

86 **WellSaid**
https://wellsaidlabs.com/

87 **WSLTTS vs WaveNet – YouTube**
https://www.youtube.com/watch?time_continue=27&v=akc1Ddt7rX4

88 **Lyrebird – Ultra-Realistic Voice Cloning and Text-to-Speech**
https://lyrebird.ai/

89 **Modulate Voice Skins SDK Overview**
https://modulate.ai/voiceskins

90 **DeepZen**
http://deepzen.io/

91 **Home - ALS Project Revoice**
https://www.projectrevoice.org/

92 **New GAN Can Lipread and Synthesize Speech – NVIDIA Developer News Center**
https://news.developer.nvidia.com/new-gan-can-lipread-and-synthesize-speech/

93 **End-to-End Speech-Driven Realistic Facial Animation with Temporal GANs**
openaccess.thecvf.com/content_CVPRW_2019/papers/Sight%20and%20Sound/Konstantinos_Vougioukas_End-to-End_Speech-Driven_Realistic_Facial_Animation_with_Temporal_GANs_CVPRW_2019_paper.pdf

94 **MuseNet**
https://openai.com/blog/musenet/

95 **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

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

97 **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

98 **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/

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

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

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

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

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

104 **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/

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

106 **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/

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

108 **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

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

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

111 **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

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

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

114 **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

115 **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

116 **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

117 **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-news-2018-03-08

118 https://twitter.com/realDonaldTrump/status/1171046015106990081?s=20

119 **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

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

121 **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/

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

123 **Dionne Stax duikt op in pornovideo**
https://www.shownieuws.nl/rubrieken/sterren/2019/dionne-stax-duikt-op-pornovideo/

124 **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

125 **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/

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

127 **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

128 **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

129 **2018 Russia–United States summit – Wikipedia**
https://en.wikipedia.org/wiki/2018_Russia%E2%80%93United_States_summit

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

131 **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

132 **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

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

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

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

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

137 **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/

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

139 **Deepfact**
https://3duniversum.com/product/deepfact/

140 **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

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

142 **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

143 **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

144 **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

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

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

147 **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

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

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

150 **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

151 **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/

152 **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

153 **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

154 **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

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

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

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

158 **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

159 **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