Daniel Smedsvik

Bachelor's thesis

Deepfake en trussel eller mulighet?

November 2021

NTNU
Norwegian University of Science and Technology
Faculty of Humanities
Department of Art and Media Studies

NTNU
Norwegian University of
Science and Technology

Bachelor's thesis

2021

NTNU
Norwegian University of
Science and Technology

Daniel Smedsvik

Deepfake en trussel eller mulighet?

Bachelor's thesis
November 2021

NTNU
Norwegian University of Science and Technology
Faculty of Humanities
Department of Art and Media Studies

NTNU
Norwegian University of
Science and Technology

Innholdsfortegnelse

1 Introduksjon

Fotomanipulering og mediemanipulasjon er ikke noe nytt, og det har uten tvil vært manipulert bilder og fotografier så lenge fotografering har eksistert. Stalin er for eksempel kjent for airbrushe folk ut av bilder som har falt ut av smak, men fra og med 1990-tallet begynte vi å se en revolusjon innen digital bildebehandling, og mye av det har blitt drevet av digitale kameraer som er allestedsnærværende (ubiquitous) og programmer som Adobe Photoshop som har gjort det enklere og enklere å manipulere fotografier i de siste 25 år. I de siste årene har en teknologi kalt deepfake vært stadig fremtredende. Dette er det denne teksten kommer til å handle om. Hovedspørsmålet i denne teksten er om deepfake er en genuin trussel eller mulighet? Hvilke problemer som er knyttet til deepfake, men også mulige muligheter som en slik teknologi kan ha.

2 Fremgangsmåte og litteraturgjennomgang

I denne oppgaven har jeg tatt for meg et prosjekt om deepfake som omfatter et stort spenn fra politikk, medievitenskap og juridiske spørsmål. Det har vært vanskelig å finne god og sikker litteratur som kan være med på å belyse dette emnet. Deepfake er en teknologi som er relativt ny og som blir dekket av alle mulige ulike forskningsfelt noe som har gjort det vanskelig å prioritere visse og ekskludere litteratur om emnet. I denne teksten kommer jeg til å bruke Hany Farid sin presentasjon av deepfake for å nærmere informere om hvordan deepfakes blir produsert, men også bruke hans perspektiv innen «mediaforensics» til vise konsekvenser, men også teknikker for å avdekke deepfakes. Jeg kommer til å bruke litt fra boken til Sturken og Cartwright for å gå tilbake i tid og fotograferingsmyten om dens representasjon av virkeligheten. Videre bruker jeg artikkelen fra Paula Fraga-Lamas som karakteriserer «fake news», men også deepfake som kan være med på å fremme en slik misinformasjon. Nicholas Diakopoulos og Deborah Johnson sin artikkel har jeg brukt for å videre undersøke problematikken av deepfake som problematisk for spredning av «fake news» og politiske prosesser under en valgkamp. Videre har nyhetsmedier brukt frykten for deepfake for å fremme sin posisjon i samfunnet og dette er noe som Aya Yadlin og Oppenheimer skriver om i deres artikkel, dette er noe jeg syntes var viktig å ta med for å vise at deepfake som et tema kan brukes til å fremme visse ideologier. En stor problematikk knyttet til deepfake er pornografiske videoer som blir produsert og distribuert på internett uten samtykke fra kjendiser, dette er noe Chandell Goose og Jacquie Burkel snakker om i sin artikkel og belyser hvordan deepfake kan bli væpnet til å objektivisere kvinner som seksuelle objekter innen visuell kultur.

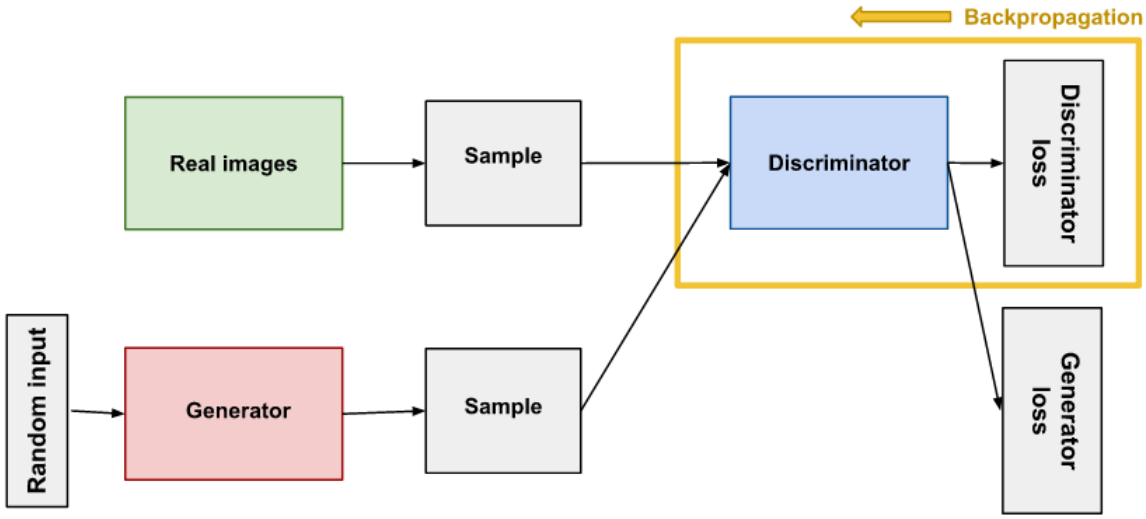
Videre i teksten kommer jeg til å drøfte noen tabu emner knyttet til deepfake sin mulighet til å «gjenopplive» døde personer til å si eller gjøre ting de aldri har sagt, eksemplene brukt er av Joaquin Oliver, skuespilleren Paul Walker og Robert Kardashian. Disse eksemplene er problematiske, men kan også bli brukt som nyttige verktøy innen underholdningsindustrien. Jeg kommer også til å forslå visse teknikker eller opplæring som må til for å avdekke eller i det minste skape en oppmerksomhet for deepfake ved hjelp av Hany Farid sin presentasjon av deepfake, men også artikkelen til Diakopoulos og Johnson som snakker om «media litteracy» som en løsning. Avsluttende kommer jeg til å gi en konklusjon av mine funn og mitt resonnement av min undersøkelse.

3.1 En introduksjon av fotomanipulasjon og deepfakes

Økningen av foto manipulering har vært brukt til å produsere artige ting, eksempelvis bytte ansiktet til president Obama med Michelle Obama. Det er relativt enkelt å gjøre dette er et program som Photoshop, og det var også mulig å manipulere videoer i midten av 2000-tallet dette har gjort det mulig for å manipulere Obama til å sparke ned en dør etter å ha holdt en tale.

Fra 2015 og videre til den dag i dag har vi begynt å se en ny tidsalder for digital manipulasjon, det er dette jeg skal snakke om i denne artikkelen. Hvis du navigerer til nettsiden thispersondoesnotexist.com, vil du bli møtt av ansikter som ikke eksisterer.

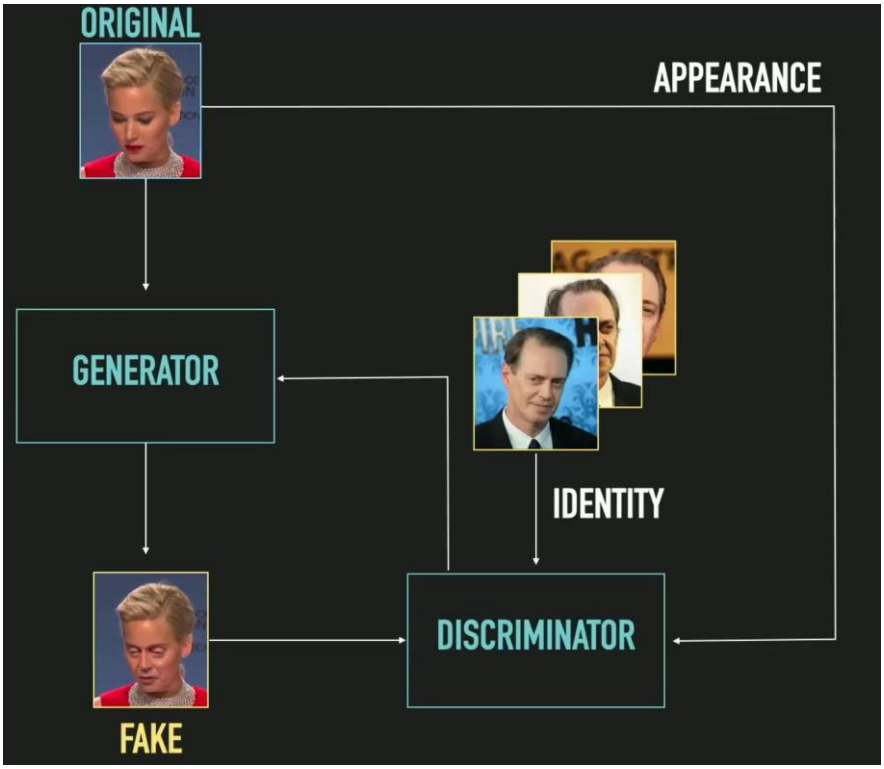
Disse ansiktene blir 100% syntetisert av en datamaskinalgoritme. De eksisterer ikke og bildene dekker alle mulige kjønn, etnisitet, alder, ansiktshår og briller. Disse bildene blir laget av en såkalt GAN (generative adversarial network). GAN Fungerer på følgende vis: man begynner med at tilfeldig bilde, og med dette mener jeg en haug med pixels i et bilde som deretter går inn i en generator. Det er generatoren sin hensikt å produsere et bilde av personer som ikke eksisterer. For å gjøre dette tar generatoren det bildet med tilfeldige pixeler og sender bildet videre til diskriminatoren (discriminator). Diskriminatoren har tilgang til bilder av hvordan fysiske personer faktisk ser ut. Diskriminatoren sin oppgave er å oppdage om bildene fra generatoren ser lik som bildene av ekte personer. Hvis svaret er nei, sender diskriminatoren bildet tilbake til generatoren og da må den prøve på nytt.



Figur 1. Illustrasjon av GAN

Når man kjører denne enkle mekanismen millioner av ganger vil den til slutt produsere et bilde hvor diskriminatoren ikke greier å se forskjellen på generatoren sitt produserte bilde og bildene av faktiske mennesker som diskriminatoren har tilgang til.

En relativt enkel teknikk ved deepfake, er en såkalt «face swap deepfake». Man kan endre to forskjellige identiteter med hverandre. Eksempelvis kan man endre identiteten av Jennifer Lawrence med Steve Buscemi. I dette scenarioet er det generatorens mål å modifisere piksler i ansiktet og gi det videre til diskriminatoren, men i dette tilfellet er det ikke generatoren sin jobb å spørre «er dette et bilde av en ekte person?» men heller «er dette Steve Buscemi?». Hvis svaret er nei, blir det bildet sendt tilbake til generatoren og sånn jobber de frem og tilbake frem til et bilde blir produsert som generatoren er tilfreds med. I en video kreves det at generatoren og diskriminatoren jobber bilde for bilde frem til videoen er laget. Til slutt får man et resultat som bildet nedenfor. Her kan man se makten deepfake kan ha, deepfake har even til å få en person til å si eller gjøre noe som den person aldri har sagt eller gjort. Eksempelvis kan denne teknologien bli brukt til å etterligne en president, presidentkandidat, en CEO av et stort firma osv. (Farid, 2020)



Figur 2. Hvordan en deeplearning algoritme kan produsere en «faceswap» deepfake

Figur 3. Steve Buscemi sitt ansikt i en «face swap» deepfake over Jennifer Lawrence sin kropp

En annen deepfake teknikk som er å manipulere en video som har en person man ønsker å få til å si noe annet i videoen. Jordan Peele brukte denne teknikken i en video han dubbet over en autentisk video av Barack Obama og fikk han til å si noe han aldri har sagt. I videoen hører vi Jordan Peele og manipulasjonen her er ikke å bytte hele ansiktet, men å synkronisere munnen i den originale videoen til å stemme overens med lyden til en annen person. Denne teknikken kalles for «lip sync deepfake». Med denne teknologien er det mulig å produsere artige videoer som blir brukt til memes, men det er også en mørkere side ved denne teknologien. Dette kommer jeg nærmere innpå senere i teksten.

En tredje teknikk er en såkalt audio deepfake. Her får en GAN gitt mange lydklipp av personen man ønsker å få til å si noe gjennom «text to speech» i personen sin egen stemme. Den canadiske psykologiprofessoren Jordan B Peterson har fått oppleve dette personlig, gjennom nettsiden «notjordanpeterson.com». Denne nettsiden brukte ai teknologi som gjorde det mulig for hvem som helst til å skrive noe og få det reprodusert gjennom Jordan sin stemme. Dette var mulig fordi Jordan har hundrevis av timer med videoer av seg selv gjennom sine forelesninger eller intervjuer. Disse videoene ble matet til en GAN som gjorde det samme, bare med lyd istedenfor bilder. Etter hvert får man en ganske autentisk rendering av stemmen man ønsker å få til å si noe vedkommende aldri har sagt. Selv om dette var mest på moro og det ble brukt til å lage memes på YouTube. Så kunne dere ha blitt brukt til mer ondsinnede formål. Jordan skrev til og med en bloggpost om hendelsen på sin nettside og skrev det følgende:

«It’s hard to imagine a technology with more power to disrupt. I’m already in the position (as many of you soon will be as well) where anyone can produce a believable audio and perhaps video of me saying absolutely anything they want me to say. How can that possible be fought? More to the point: how are we going to trust anything electronically-mediated in the very near future (say, during the next Presidential election)? We’re already concerned, rightly or wrongly, with “fake news”—and that’s only news that has been slanted, arguably, by the bias of the reporter or editor or news organization. What do we do when “fake news” is just as real as “real news”? What do we do when anyone can imitate anyone else, for any reason that suits them?» (Peterson, Jordanbpeterson.com).

En skumlere teknikk er en som er mer komplisert kalt «Neural Voice Puppetry» og kanskje den skumleste av dem alle. Her kombinerer man en video enten gjennom en «faceswap deepfake» eller en «lip sync deepfake» med en «audio deepfake», resultatet er at man har en video hvor lyden er så å si autentisk til den personen man etterligner, men man kan også få vedkommende til å gjøre ting i videoen. Dette er en teknikk som kan ha store konsekvenser ettersom denne teknologien bare kommer til å bli mer sofistikert i fremtiden. Denne makten kan ha store konsekvenser for hvert enkelt samfunn på kloden. I midten av den kalde krigen har atomvåpen nesten blitt avfyrt på grunn av svikt av ulike systemer. Hva kan skje dersom i en periode i fremtiden befinner oss i en lignende ustabil periode og det sirkulerer enn video av en statsleder som erklærer en atomkrig mot en annen nasjon?

3.2 Destabiliserende for samfunnet «fake news»

Før jeg snakker om ødeleggende effekter som deepfake kan ha for samfunnet, er det nyttig å snakke om fotografering som en representasjon av virkeligheten. Så lenge fotografering som et medium har eksistert har det hatt en nær tilknytning til en representasjon av virkeligheten. Det er en ide om at kameraet er en objektiv maskin, selv om fotografering gjennom et kamera involverer en grad av subjektive valg. Disse prosessene er involverer motiv, innramming og belysning osv. siden midten av 1900 tallet har det vært mange argumenter for og imot ideen om at fotografier er objektive representasjoner av virkeligheten. Fotografier kan for eksempelvis bli brukt til å bevise at en viss person var i live i en gitt periode. Sturken og Carthwright snakker om dette i boken «practices of looking» og foreslår at fotografering som ekte representasjon av sannhet er en myte. Ved tidlige analoge fotografier har alltid vært mulig å manipulere bilder. Et tidlig eksempel på manipulasjon av virkeligheten gjennom fotografering, er bildet av en død konføderert fra den amerikanske sivil krigen. Dette bildet ble publisert i Alexander Gardners bok fra 1865 «Photographic Sketchbook of the War». Den amerikanske sivil krigen og den første krigen som ble dokumentert av kameraer. Gardner Presenterte bilde som en scene han hadde støtt på. Senere ble det mistenkt at Gardner kludret til scenen, tilsynelatende var riflen mistenkelig plassert der for dramatisk effekt og soldaten hadde blitt dratt inn i en bedre vinkel og hodet til soldaten støttet opp slik at ansiktet ble synlig for kameraet (se bildet i figur nedenfor). På 1990-tallet ble bildet analysert av William J. Mitchell og andre forskere i en tid når digital fotograferingsmanipulasjon hadde blomstret frem via programvare som Phhotoshop. Dette gjorde det vanskelig å støtte fotografering som en ekte representasjon av virkeligheten. (s. 24-26 Sturken og Carthwright).

Figur 4. Timothy O'Sullivans Devil's Den

Innen journalistikk er det særdeles viktig at et bilde eller video som blir vist for allmennheten holder en visst integritet til journalistiske retningslinjer. Disse retningslinjene blir visket ut dersom en deepfake video blir spredt til massene. Hvem som helst som har tilgang til programvare som kan produsere deepfakes kan bli brukt til å lage overbevisende videoer av hendelser som aldri har skjedd. Falske nyheter eller «fake news» er et begrep som har blitt brukt mye de siste årene. Selv om begrepet er kritisert særlig siden den tidligere presidenten Donald Trump brukte dette begrepet til å nedsette ulike nyhetsmedier som han ikke likte. Likevel er begrepet til nytte når vi ser på deepfake sin mulighet til å spre misinformasjon. Paula Fraga-Lamas karakterer falske nyheter som: 1. innehold som er med på å manipulere/lure mottakere. 2. Innhold som er fokusert på å skape usikkerhet, fiendtlighet eller polarisering som kan forstyrre demokratiet, særlig ved demokratiske prosesser (f.eks. valg og folkeavstemninger), grunnleggende rettigheter eller rettssystemet. 3. Innhold som dekker saker av offentlig interesse (politikk, helse, miljø osv.) 4. Informasjonen formidles strategisk gjennom automatiserte og aggressive teknikker, falske kontoer, roboter, mikromålretting eller trolling. 5. Den har egenskaper som muliggjør rask og utbredt spredning. 6. Den er stadig mer motstandsdyktig mot oppdagelse siden AI, «augmented reality og «virtual reality» utvikler seg raskt, og siden identifisering av falske nyheter har fått mindre ressurser (dvs. finansiering og institusjonell støtte) enn de som konstruerer «fake news». I hennes artikkel forteller hun:

«Although the potential influence of fake news still remains uncertain, in at least a few cases (e.g., the Brexit campaign, the independence of Catalonia), they appear to have impacted significantly public behavior. An example is the so-called misinfodemics, where health misinformation (e.g., the effect of vaccines) is enabling the spread of diseases.» (s. 2-4, Lamas)

Videre skriver hun at falske nyheter har en stor virkning på to områder: 1. Personvern og databeskyttelse, altså tjenester som tilbys av platform leverandører drives i en økende grad som en inntektsgenerering for «big data». For eksempel biometriske- og rekognisjonssystemer (f.eks. ansikt, øye, stemme og følelser). Dette er med på å øke sårbarheten til personvern og er med på å avdekke personopplysninger ufrivillig. 2. Ytringsfrihet til å motta pålitelig informasjon. Fremveksten av deepfakes vil forverre svindel siden individer, bedrifter og samfunnet kan bli møtt med nye former for utpressing som kan ha stor risiko for demokratiet og nasjonal sikkerhet. Lamas forteller at på den ene siden vil den fremvoksende teknologien muliggjøre falske nyheter å bli mer suksessfull, men på den andre siden vil den samme teknologien gjøre det mulig å bekjempe dem. Ulike plattformer på nettet har gjort fremsteg ved å bekjempe falske nyheter. Google har annonsert «Google News» for å støtte sikre nyheter.

Problematikken rundt deepfake til å destabilisere samfunnet, særlig demokratiske prosesser er noe som blir diskutert mye i nyhetssaker og akademiske artikler. Fordi deepfakes kan bli brukt til å under en valgkampanje til å få velgere til å stemme annerledes ved å vrenge virkeligheten. Dette er noe som Nicholas Diakopoulos og Deborah Johnson diskuterer i deres artikkel «Anticipating and addressing the ethical implications of deepfakes in the context of elections». Bedrageri via en deepfake video

under et valg er en skade for enkeltpersoners evne til å ta en informert beslutning. Falsk informasjon om kandidater kan forvrenge en velgers beslutning under et valg. Diakopoulos og Johnson forteller at et enkelt tilfelle av bedrageri er nok til å skade et individs beslutning, men store distribusjoner av deepfakes blir fordel vil skaden forstekes. Denne spredningen blir støttet av digitale og sosiale medier, noe som gjør det vanskelig å bekjempe det (s 8, Diakopoulos og Johnson). Deepfakes som blir publisert i de siste ukene av et valg kan være umulig å bekjempe, men det vil være viktig å publisere en respons av dens falskhet. Diakopoulos og Johnson forteller at skaden ved å produsere en deepfake uten samtykke kan best forståes som «persona plagiarism», dvs. en plagiering med et fokus på kilden i stedet for innholdet (s. 11, Diakopoulos og Johnson).

I deres artikkel forteller de om at lovverket i USA knyttet til ytringsfrihet under en valgkamp er kompliser og generelt sett er det ikke noen lover knyttet til falske ytringer under en valgkamp. Likevel er det ikke lov å drive generell ærekrenkelse for individer, og dette kan føre til søksmål. De forteller «The reluctance of U.S. law to ban false campaign speech has more to do with the dangers and challenges of trying to regulate false claims than denying the harmfulness of false speech.» Regulering og sensurering av deepfakes er noe jeg kommer til å diskutere senere i teksten.

Aya Yadlin og Oppenheim skriver om hvordan nyhetsmedier har brukt frykten ovenfor deepfake til å posisjonere sin status i samfunnet, i artikkelen «Whose dystopia is it anyway? Deepfakes and social media regulation». I deres artikkel skriver de at i nyhetssaker knyttet til deepfaker er det lagt mye vekt på dens potensiale for a skade for sårbare grupper i samfunnet, undergrave oppfatningen av virkelighet, og greie å skille det ekte fra det falske. Nyheter som dekker deepfake fenomenet er ofte omringet rundt en frykt for teknologien og dens potensiale for skade. Dette fører til en stor diskusjon på hvilke steg som må til for å regulere manipulasjon og spredningen av misinformasjon. Hvis man kombinerer face swapping med den teknologiske muligheten for å etterligne noen sin stemme har man potensiale for misinformasjon på et katastrofalt nivå. Eksempler som Yadlin kommer med er Trump som kan erklære krig mot Nord Korea eller Hillary Clinton som kan bli manipulert til å lovprise Illuminati. Begrepet «falske nyheter» setter spørsmålstegn ved statusen som journalistikk har som en troverdig kilde til allmennheten i en såkalt «post-sannhets-æra». Dette overlapper med andre negative informasjon spredninger som misinformasjon. Falske nyheter er med på å viske ut tradisjonelle journalistiske standarder og normer. Yadlin og Oppenheim argumenter i sin artikkel at nyhetsmedier brukte bekymringer knyttet til deepfake til sin fordel for å ta avstand fra falske nyheter og dermed gjenopprette en troverdighet til journalister sitt arbeid. Dette fritok journalister fra kritikk om falske nyheter som kan bli spredt av dem. Dette har skapt en ideologisk og sosial avstand mellom misinformasjon og nyhetsmedier. Yadlin og Oppenheim forteller i sin artikkel:

«The technological fabrication created by deepfake applications was tied by journalists to the fake news phenomenon, where positioning deepfakes as a threat tosociety became a functional tool journalists used to renegotiate their role and importance in Western societies» (s. 11, Yadlin og Oppenheim).

The technological fabrication created by deepfake applications
was tied by journalists to the fake news phenomenon, where positioning deepfakes as a threat to
society became a functional tool journalists used to renegotiate their role and importance in
Western societies
Videre forteller de at perspektivet fra nyhetsmedier er: reguleringen av sosial medier blir sett på som den eneste mulighet for å sikre at allmennheten har en delt oppfatning av virkeligheten. Dette blir presentert fra en dystopisk synsvinkel konstruert av nyhetsmedier. Problemet med et slikt ståsted er at det overskygger en viktig samfunnsdiskurs som er nødt til å ta sted rundt media ansvar og etikk (s. 13 Yadlin og Oppenheim).

3.3 Problemer knyttet til deepfakeporn og identitet

Deepfake har allerede blitt brukt til å lage falske videoer av kjendiser som eksempel en serie av Nicolas Cage Videoer, og politikere som Barack Obama i Jordan Peel sin video. En stor problematikk med deepfake er pornografien som blir produsert av kjendiser i kroppene til forskjellige pornografiske skuespillere. Dr. Chandell Gosse og Jacquie Burkel snakker om dette og forteller at deepfake brukes til å produsere pornografi uten samtykke av kjendiser. Selve teknologien har ikke en innebygd misogynis i teknologien, men blir brukt til å produsere og distribuere pornografi på internett.
Forfalsket/manipulert pornografi har eksistert tidligere, men Gosse og Burkel forteller i sin artikkel at det er seksuelle deepfakes som for første gang har fått stor dekning i nyhetsbildet. kvinners agency blir i dette tilfellet fjernet/ svekket. I artikkelen til Goose og Burkel sier de:

«Over the last decade women's digital intimate images in particular have been weaponized and used to inflict harm. Collectively known as image-based (sexual) abuse, these practices include, among others, "relationship retribution," "sextortion," "sexual voyeurism," "sexploitation," and "sexual assault".» (s. 2 C. Goose og J. Burkel).

For det meste avhenger dette bildemisbruket tilgang til intime bilder/videoer av kvinnen som blir avbildet. I motsetning til det som Goose og burkel sitt omtalt misbruk kan deepfakes brukes til bildebasert overgrep uten tilgang til faktiske intime fotografier eller videoer av kvinnen som blir avbildet i deepfake videoen. Ved å bruke deep fake kan enhver kvinne bli brukt og fremstilt i pornografi. Det eneste som trengs her er et tilstrekkelig korpus av ansiktsbilder, en passende pornografisk film som kvinnens ansikt skal plasseres i, og litt teknisk kunnskap kan nå deep fake brukes til å utnytte kvinners rettigheter. Gjennom artikkelen til Goose og Burkel understreker de at nyhetsmedier ofte fokuserer på trusselen rundt deepfake til å bli brukt som et våpen for å spre politisk misinformasjon, og ikke like mye rundt problematikken rundt ikke samtykkende

deepfake pornografi. Innen visuell kultur har kvinners kropper lenge blitt objektivisert gjennom reklame, film, magasiner osv. Photoshop har vært med på å lage seksualiserte bilder av kvinner. Selv om det er strategier for å overvinne/oppdage deepfakes er det en form for ødeleggende omdømmemessig ovenfor ofre som blir utnyttet i en seksualisert deepfake som ikke kan repareres senere av disse oppdagelsene. For kvinner som allerede har fått ansiktet sitt byttet ut i en sånn deepfake er skaden allerede gjort.

Funnene til Goose og Burkel avdekket at det hovedsakelig var tre måter kvinner ble skadet på i en seksualisert deepfake. For det første var det en emosjonell og psykologisk bekymring.For det andre var det et tap av autonomi over ens egen kropp og omdømme. For det tredje var det en sammenheng mellom deepfake og andre mulige kriminelle handlinger. Goose og Burkel siterer noe fra en artikkel de analyserte «Some deepfake videos may create false statements of fact about a person's presence and actions that lead to a loss of reputation of that person» (s. 10, Goose, Burkel). Videre forklarer de at det er en bekymring at de som blir avbildet kan bli ofre for trakassering, ærekrenkelse og utpressing.

Videre forteller Goose og Burkel at gjennom deres analyse av ulike nyhetskilder knyttet til deepfake, at det er gitt mer oppmerksomhet og bekymring rundt deepfake som et verktøy for å spre misinformasjon. Dermed blir det klart at faren rundt deepfake er politisk og det er mindre farlig at det blir spredt falsk pornografi. Goose og Burkel forteller videre at bruken av deepfake for å spre ikke samtykkende pornografi bør bli gitt like mye oppmerksomhet i enhver diskusjon av teknologien. Bekymringene knyttet til deepfake til å spre misinformasjon vokser, men Goose og Burkel sin bekymring er at skaden forårsaket av seksuelle deepfake blir overskygget av andre problemer knyttet til teknologien.

En annen forsker Hany Farid som spesialiserer seg innen «media forensic» er enig i at det er et stort problem knyttet til bruken av deepfake for å spre pornografiske videoer på internett og dette er noe som vi er nødt til å ordne opp fremover. Farid forteller i en video om deepfake:

«… there's also a darker side to this technology and where we are seeing the biggest harm today is in the form of nonconsensual pornography taking one person's likeness and inserting it into sexually explicit material and then distributing that on the Internet. That is probably the most common use of deepfake technology today. Which is yet another example of the weaponization deepfake technology against women and something that I think we need to get a handle on.» (Hany Farid)

På lignende vis som deepfake kan bli misbrukt til produksjonen av ikke samtykkende pornografi, kan man også produsere ikke samtykkende videoer av avdøde mennesker. Hva som gjør et menneske ekte er konstruert av ens stemme, utseende og identitet. Dette blir visket ut av deepfakes og gjør at disse tingene ikke lenger er unik til ens egen identitet. Algoritmene bak deepfake gjør det mulig å manipulere ens stemme, ansikt og

kroppsspråk. Dette tar vekk en side av oss som gjør oss mennesker, vår autonomi. Joaquin Oliver døde under en masseskyting på Marjory Stoneman Douglas skole i Florida, USA. Gjennom en deepfake teknologi ble han gjenopplivet for å oppmuntre folk til å støtte politiske tjenestemenn som fremmer våpensikkerhet. Ved hjelp av foreldrene til Joaquin ble denne videoen produsert av en frivillig organisasjon kalt «Change the Ref» som lagde denne aktivist videoen. Joaquin sitt ansikt ble overlagt en «stand in» og hans stemme ble rekonstruert gjennom algoritmer som ble matet audio av hans stemme (Diaz, 2020). Selv om man kan være enig i budskapet i videoen er det noe foruroligende med måten det blir uttrykt. Juaquin har aldri sagt disse ordene og han ga ikke samtykke til å bli gjenopplivet på denne måten.

Figur 5. Joaquin Oliver gjenopplivning til et politisk budskap

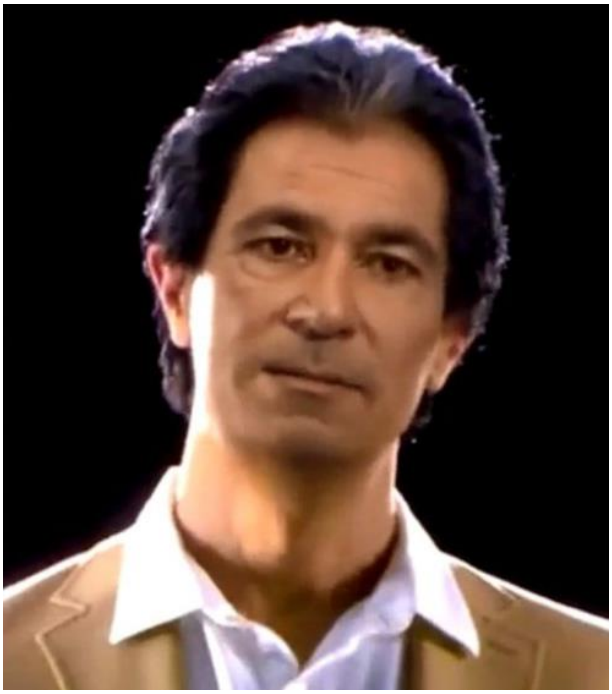
3.4 Deepfake som en mulighet

Der er ikke bare skadelig potensialet i teknologien, deepfake kan eksempelvis bli brukt i underholdningsindustrien, og mer spesifikt i filmproduksjon. Amazon prime sin nye serie som handler om tolken sitt univers «middle-earth» denne serien fra går lenge før fortellingen i «ringenes herre». Ian Mckellen var kjent for sin rolle som Gandalf i trilogi, selv om Amazon kan velge en annen skuespiller for å spille Gandalf i sin yngre form eller ikke ha Gandalf med i sin fortelling, kan de ved hjelp av teknologien eksempelvis bruke eldre filmopptak eller bilder av Ian Mckellen for å bruke han i deres fortelling. Et annet eksempel hvor deepfake kan brukes i filmproduksjon er at man har en skuespiller som befinner seg på andre siden av jordkloden ellers som er utilgjengelig for filmopptak der filmen skal spilles inn, men filmen må spilles inn på et spesifikt tidspunkt. Ved teknologien kan man nå ha en «stand in» som kan være kroppen til skuespilleren, mens man kan på en bruke en såkalt «face swap» deepfake med skuespilleren en «stand in» sin kropp.

Det er også mulighet for å bruke døde personer som ikke kan spille inn filmer ved å gjenopplive de i en ny film. Dette er tabulagt som nevnt over og som kan bli misbrukt, men hvis det gjøres på en respektert måte kan det ha nyttelser. Eksempelvis ble noe lignende gjort under innspilningen «Fast and Furious 7». Den avdødde Paul Walker ble gjenopplivet gjennom VFX artister i sin rolle i filmen. Siden flere viktige scener som ikke hadde blitt filmet så det ut som om-opptak av scener med en ny skuespiller i filmen så ut til å være eneste mulighet, men filmselskapet bestemte seg for å fullføre filmen og Paul Walker sin fortelling. Dette var mulig fordi brødrene til Paul: Caleb og Cody var med på å bli skannet og brukt i scener, som var det nærmeste filmselskapet kunne være for å gjenskape Paul i en tid før deepfakes. 260 filmopptak som involverte brødrenes skuespill hvor ansiktene ble byttet ut med CG (computer generated) versjon av Paul. For å lage sånne CG-hodeerstatninger for Paul Walker ble brødrene som nærmeste referanse. Så brukte de Paul sine opptak for den siste finpussen for å fange detaljer som hudteksturer. I nyhetsartikkelen fra Diaz «How 'Furious 7' Brought the Late Paul Walker Back to Life» forteller VFX eksperten joe latteri:

"we used a lot of Paul's footage as reference, because as close as the brothers were in style and mannerisms, they just weren't Paul when Paul played his character. We really tried to limit our interpretation of the character to things that we had seen Paul do as the character. We found performances that matched the situation that we needed to put him in, and we used that to guide us." (Diaz, 2015)

Kanye West brukte deepfake da han ga sin gave til Kim Kardashian på hennes 40 årsdag. Kanye hadde fått noen til å lage et hologram av hennes døde far Robert Karddashian. Ved hjelp av data samlet inn av hennes far kunne de produsere et hologram som kunne snakke til sin datter fra den andre siden av graven (Gorman 2020). Det er bare et spørsmål om tid før gaver som dette kan nå allmenheten. Familie arvegods og portretter kan bli erstattet med videogjenskaping av døde familie medlemer.



Figur 6. og 7. Utklipp av videoer som viser hvordan Paul Walker og Robert Kardashian ble gjenopplivet gjennom VFX effekter og deepfake.

Innen video spill har også en AI-teknologi blitt brukt for å lage en helt audio syntetisert trailer av stemmene-skuespillerne fra spillet Skyrim. En amatør filmskaper på YouTube Adriac har brukt en maskin lærings program som har samlet dialog fra spillet og har greid å reprodusere helt ny syntetisert dialog i en trailer. Uten tilgang til Bethesda sine stemme-skuespillere har Adriac i stedet brukt eksisterende dialog i spillet gjennom bruk at den AI-drevende synteseverkøyet xVASynth. Opplært fra lyd hentet direkte fra spillet kan verktøyet brukes til å sette sammen og finjustere helt skreddersydde dialoglinjer fra Skyrim. Adriac forteller at «I don't think it will ever surpass a voice actor, but I think this could be incredible for the future to keep voice actors voices alive even hundreds of years after they've passed.» (Clayton, 2021)

3.5 Hvordan kan vi bekjempe deepfake som kan være skadelig?

Gjennom denne teksten har jeg nevnt noen særlig skumle eksempler om hvordan deepfake kan bli brukt til å skade både enkelt individer, men også samfunn. Så da blir dermed spørsmålet hvordan kan vi bekjempe deepfake? Særlig siden den vil bli enda mer sofistikert i fremtiden. Ofte har nyhetsmedier vært med på sensasjonellere deepfake som noe vi nærmest ikke kan bekjempe og må dermed nesten bare regulere og sensurere det totalt. Problemet med et sånt ståsted er at man får en fremstilling av et mediedeterministisk synspunkt.

Diakopoulos og Johnson som nevnt over kommer med noen forslag om hvordan man kan bekjempe deepfakes via utdanning og media litteracy. Utdannin og opplæring innen mediekunnskap har en stor rolle og som blir nødvendig jo mer utbredt deepfake blir i fremtiden. Tanken her er at enkeltpersoner kan oppmuntres til å utvikle en bevissthet

om teknologiens evner slik at man kan tilegne seg ferdigheter i å verifisere, fakta sjekke og vurdere informasjonskilder på nettet. Dette vil hjelpe enkeltpersoner å ta ansvar for sitt eget bruk og deling av informasjon. Diakopoulos og Johnson forteller:

Even if it does little to address the sources and motivations for the production of deepfakes, by equipping viewers with critical assessment abilities this strategy is compelling because it directly mitigates the potential for deception and furthermore diminishes reputational harm by reducing amplification via social channels.» (s. 14, Diakopoulos og Johnson).

Videre glemmer også nyhetsmedier ofte å nevne at fagpersoner jobber med å oppdage og beskytte mot deepfake. Personer som «mediaforensics» forskeren Hany Farid forteller at det er teknologier som er under utvikling for å oppdage slike typer deepfakes, og beskytte oss mot mulige trusler. En teknikk er å mate en slags lignende neuralt nettverk som deepfake, hvor dens mål er å oppdage deepfakes. Dette kan gjøres ved ulike teknikker som Farid forklarer, teknikker som involver at dette neurale nettverket analyserer tidligere video opptak av den som blir målrettet i en deepfake og sammenligner det med en potensial deepfake. Farid bruker som eksempel Barack Obama til å illustrere poenget, Obama snakker på en veldig markant måte. Gjennom tidligere videoer av Obama kan man se at han har det Farid kaller for «mennerism of behavioral tic that he has, and we all have it. I raise my eyebrow when I emphasize something. We are going to use those mannerisms to try to detect deepfake.» Dette er noe man kan bruke for å se at den som manipulerer en såkalt «face-swap» deepfake video som prøver å få Obama til å si en ting, mens kroppsspråket sier noe annet. Man kan også gjøre noen grunnlegende hode- og ansiktssporing med tidligere videoer av eksempelvik Obama for å se om ansiktsuttrykkene hans stemmer overens med bildet som blir vist. Dermed kan man oppdage om noe er en deepfake eller etterligning som er veldig spesifikk til personen sine ulike væremåter.

Farid avslutter i videoen om deepfake at teknologien er såpass rask utviklende og at de kommende årene vil deepfake være helt annerledes enn den er i dag. Produksjonen av deepfake er i stadig endring og et felt i rask bevegelse. Vi må begynne å tenke seriøst og nøye om trusselen om misinformasjon i en tid hvor vi stoler mer og mer på internett for informasjon. Når vi ikke kan stole på media som vi ser, hører og leser daglig er vi i trøbbel som et samfunn og som et demokrati. Vi er nødt til å ta reguleringen seriøst. Teknologiselskapene som lar denne misinformasjonen spres gjennom nettverkene deres må bli enda mer seriøse om hvordan plattformene deres blir våpengjort. Reguleringen av disse sosiale mediene må bli mer seriøse om hvordan de kan få kontroll over misinformasjon som blir brukt til å skape sivil uro og rive samfunnet fra hverandre og forstyrre demokratiet. Alt dette må skje kjapt, men med respekt for ytringsfrihet (Farid, 2020).

4. Konklusjon

I denne teksten har jeg prøvd å belyse hva deepfake er, men også hva det kan brukes til og mulige konsekvenser eller muligheter det kan ha for oss.Dette har jeg drøftet om igjennom oppgaven. Deepfake teknologien har blitt brukt for å spre memes av massene, men også blitt brukt for å advare oss mot dens potensiale til å være ødeleggende for

individer gjennom ikke samtykkende pornografiske videoer, men også samfunnet og dens prosesser særlig under valgkamper.

Det er mye som er umulig for oss å forutse i fremtiden og hva den kommer til å bringe til teknologien deepfake, det som kan sies er at teknologien er i stadig utvikling og vil være stadig vanskelig å skille fra genuine videoer. Foreløpig er teknologien i sin første utviklingsfase og det er den tid i dag ikke særlig vanskelig for fleste å se forskjellen på ekte videoer og amatørproduserte deepfakes, men siden teknologien er i en såpass rask utvikling vil det sannsynligvis bli nærmest umulig for de fleste mennesker å skille mellom ekte og deepfake videoer. Selv om det meste jeg har diskutert ikke er stort mer enn deepfakes potensiale til å bli mer avansert og raffinert, er det en utvikling som man er nødt til å følge med på i fremtiden. Man kan produsere teknologi for å avsløre slike programmer, men også tilegne seg personlige egenskaper.

Istedenfor å følge et mediedeterministisk synspunkt som nyhetsmediene ofte presenterer dette gjennom sensasjonelle artikler, hvor de nesten fremmer en total sensurering av deepfake teknologien, noe som i seg selv er nærmest umulig å gjennomføre i praksis. Er en mer realistisk håndtering å utvikle på samme måte som deepfake, teknologien en teknologi som er rettet mot å oppdage deepfakes før de blir distribuert til massene. Dette er en form for regulering som ikke omfatter total sensur, men en regelasjon som krever et samarbeid mellom ulike nyhetsmedier, men også sosiale medieplatformer for å bli enig i hva som kan og ikke kan bli distribuert. En annen tilnærming er en konstant oppdatering innen såkalt «media litteracy» som kan delegere ansvaret til den enkelte for å greie å skille det som er sant fra det som er manipulert. Noe som kan være nyttig for alle som vokser opp i en stadig mer manipulert verden.

5 Referanser og kilder

Artikler og bøker:

Sturken, Marita og Cartwright, Lisa 2018. *Practices of Looking: And Introduction to Visual Culture*. Revidert tredjeutgave. Oxford: Oxford University Press.

Farid, Hany. 2019. *Fake Photos.* Cambridge, MA: MIT Press.

Gosse, Chandell og Burkell, Jacquelyn. 2020. «Politics and porn: how news media characterizes problems presented by deepfakes.» *Critical studies in media communication,* 37(5): 497-511. doi: 10.1080/15295036.2020.1832697

Fraga-Lamas, Paula og Fernandez-Carames, Tiago M. 2020. «Fake News, Disinformation, and Deepfakes: Leveraging Distributed Ledger Technologies and Blockchain to Combat Digital Deception and Counterfeit Reality.» IT Professional, 22(2): 53-59. doi: 10.1109/MITP.2020.2977589

Yadlin-Segal, Aya og Oppenheim, Yael. 2020. «Whose dystopia is it anyway? Deepfake and social media regulation.» *Convergence,* 27(1): 36-51. doi: 10.1177/1354856520923963

Ann-Christine Diaz. 2020 «Parkland victim Joaquin Oliver comes back to life in heartbreaking plea to voters»

https://adage.com/article/advertising/parkland-victim-joaquin-oliver-comes-back-life-heartbreaking-plea-voters/2285166

Carolyn Giardina. 2015 «How 'Furious 7' Brought the Late Paul Walker Back to Life»

https://www.hollywoodreporter.com/movies/movie-news/how-furious-7-brought-late-845763/

Alyx Gorman. 2020. «Kim Kardashian's father resurrected as hologram in birthday present from Kanye West»

https://www.theguardian.com/lifeandstyle/2020/oct/30/robert-kardashian-resurrected-as-a-hologram-for-kim-kardashian-wests-birthday

Natalie Clayton. 2021. «This fan-made Skyrim trailer uses entirely AI-synthesized voice acting»

https://www.pcgamer.com/this-fan-made-skyrim-trailer-uses-entirely-ai-synthesized-voice-acting/

Videoer:

Hany Farid. 2020. «Creating, Weaponizing, and Detecting Deep Fakes»

https://www.youtube.com/watch?v=poSd2CyDpyA&ab_channel=Databricks

Figurer:

Figur 1. https://developers.google.com/machine-learning/gan/discriminator

Figur 2. https://www.youtube.com/watch?v=poSd2CyDpyA&ab_channel=Databricks

Figur 3. https://www.youtube.com/watch?v=r1jng79a5xc&ab_channel=birbfakes

Figur 4. Devil's Den https://www.loc.gov/item/2018666313/

Figur 5. https://www.youtube.com/watch?v=m6I_wEetSck&ab_channel=ChangetheRef

Figur 6. https://www.hollywoodreporter.com/movies/movie-news/how-furious-7-brought-late-845763/

Figur 7. https://www.theguardian.com/lifeandstyle/2020/oct/30/robert-kardashian-resurrected-as-a-hologram-for-kim-kardashian-wests-birthday