Skaperkraft
TANKESMIEN

Deepfakes - Hva gjør vi når vi ikke lenger kan tro det vi ser?

Notat i Skaperkrafts prosjekt Sannhet 2.0
Desember 2020

Catharina Rodenburg Drejer

Hva er sannhet?

Dagens digitalt sammenvevde verden er en helt annen verden enn for bare tretti år siden. Vi har levd i årtusener uten umiddelbar tilgang til kunnskap og global kommunikasjon. Men vi lever nå med et globalt nettverk med lag på lag av relasjoner, bedrifter og grupper. Internett er en revolusjon av kunnskap og kontakt mellom mennesker, så mye at det først var overveldende og vanskelig å tro, men deretter plutselig helt vanlig.

Så vanlig at vi glemmer hvor uvanlig det er. Så vanlig at vi kan overse hva dette innebærer.

Internett har svar på tusenvis av spørsmål, har løsningen på utallige problemer og har skapt et rom for kreativ utfoldelse for millioner. Det frie og åpne globale internett lar alle snakke, handle og i stor grad gjøre som de vil. Det kan virke som en utopi; for hva om det vi tror er sant på nettet, ikke lenger er sant, eller vi ikke lenger kan vite om det er sant? Hva om den digitale teknologien er i ferd med å endre det som definerer sannhet?

Bokmålsordboka definerer sannhet følgende

sannhet m1, f1
1 det å være sann
tvile på sannheten i en påstand / medføre sannhet være sann, til å stole på */ i sannhet et prektig menneske* sannelig, virkelig
2 utsagn, beretning som stemmer med virkeligheten
holde seg til sannheten / si, fortelle (en) sannheten også, si sin mening uten omsvøp */ absolutt sannhet / filosofiske sannheter*

et sannhetens ord (si) et refsende, alvorlig ord (til noen)

3 oppriktighet, ærlighet
tjene Herren i sannhet / leve i sannhet

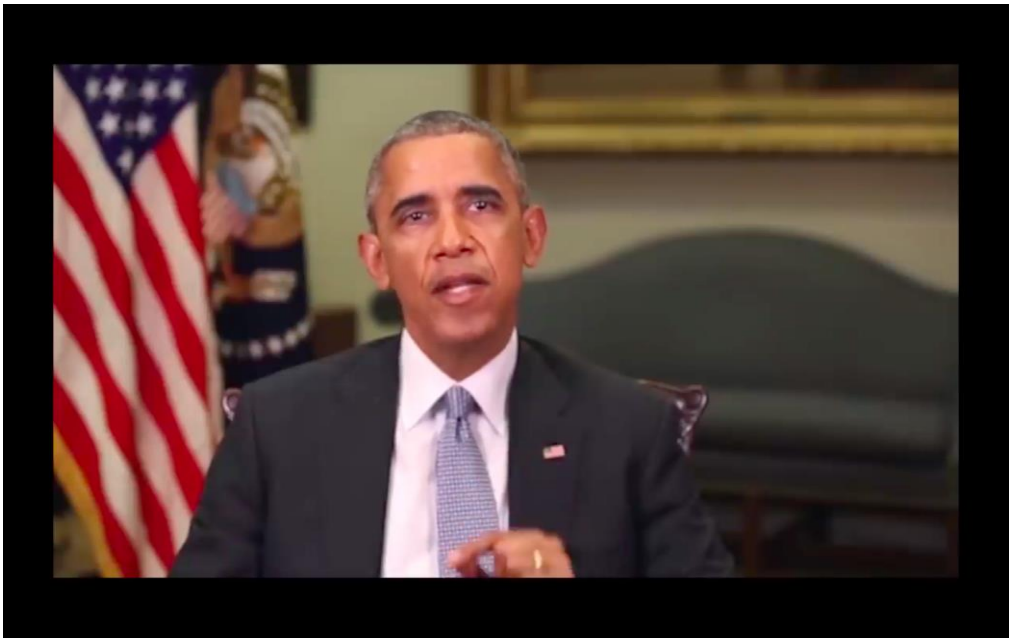
Jeg kan ikke tro mine egne øyne

I 2017 ble de første kjente deepfake-bildene publisert. En bruker på Reddit hadde da manipulert ansikter av berømte kvinner, som for eksempel Scarlett Johanson, Taylor Swift og Emma Watson, inn i pornofilmer. Brukeren publiserte disse under navnet "Deepfakes". Året etter kom den første app'en til smarttelefon, FakeApp, som lot brukere bytte ansikt med hverandre. Siden den gang har deepfake-manipulasjon blitt kommersialisert og tatt i bruk av mange bedrifter til ulike formål.

I dag refererer deepfake stort sett til videoer hvor noens ansikt, stemme eller kropp egentlig er en fiktiv versjon laget med kunstig intelligens. Uttrykket kan i tillegg brukes om bilder eller lydklipp av samme art. Deep, eller dyp på norsk, viser til dyp læring som er en form for kunstig intelligens hvor man forsøker å etterligne menneskelige tankeprosesser med såkalte kunstige nevrale nettverk.

Fake, eller forfalsket på norsk, sikter til ett aspekt som dyp læring muliggjør, nettopp det å lage realistiske, men forfalskede videoer, lydklipp og bilder. I kort handler deepfake – eller dype forfalskninger – om etterligninger, manipulasjoner eller sammenstillinger av video og lyd ved maskinlæring. Ansiktsbytter, kroppsmanipulasjon, stemmekopi og produksjon av helt nye stemmer og bilder faller alle i samme kategori.

Det er ikke bare pornografiske videoer som er blitt laget. Deepfake-videoer er også brukt i politikken, som for eksempel da Barack Obama i 2018 ble brukt av Buzzfeed for å illustrere hvordan teknologien kan forfalske videoer av kjente personer. Klikk på linken for å se en deepfake av Barack Obama.



Denne typen manipulasjon, som også kalles "falske nyheter på steroider", har potensial til å grunnleggende endre vår mulighet til å avgjøre hva som er sant og ikke. Tillit og sannhet er helt grunnleggende i å opprettholde demokratiet, og deepfake-manipulasjon må som sådan tolkes som en trussel mot demokratiet. Spesielt er det i valg og valgkamp stor fare for at forfalskede videoer av politikere blir spredd på internett.

I april 2018 ble det i et tre minutter langt videoklipp, tilsynelatende filmet i BBC News Studio, rapportert om utbruddet av en atomkrig mellom NATO og Russland, etter at et russisk fly ble skutt over Østersjøen. Klippet inkluderte BBC-logoen og et britisk nyhetsanker. Klippet gikk viralt på sosiale medier og BBC måtte gå ut offentlig og ta avstand fra det. Videoen var selvfølgelig helt falsk, men mange ble bekymret.

Hvor utbredt er deepfakes og i hvilken grad truer de demokratiet? I 2019 publiserte den nederlandske organisasjonen Sensity (tidl. Deeptrace) en rapport over det totale antall deepfakes på internett. Organisasjonen fant at siden desember 2018 antallet deepfakes hadde doblet seg hver sjette måned, altså eksponentiell vekst i publiseringer. Per 20. juni 2020 hadde de identifisert nær 50.000 videoer, en økning på 330 % det siste året.

Fremdeles er majoriteten av publiserte videoer innen underholdningssegmentet, mest blant pornografiske filmer. Bare fire prosent er rettet mot politikere ifølge rapporten. Mer foruroligende er det i følge Britt Paris, ekspert hos Data & Society og medforfatter av rapporten "Deepfakes and Cheap Fakes", at forfalskede videoer og bilder brukes mot kvinner, fargede og mennesker som stiller seg i opposisjon til mektige systemer og regimer. Hun hevder at de mest utsatte for deepfake-manipulasjon er de som ikke evner å kontrollere hva som brukes som bevis mot dem.

Dette er viktig. Overvekten av mediedekningen har vært på politisk satire og den overnevnte trusselen mot valg. Staten California har eksempelvis innført forbud mot deepfake-manipulasjoner – både å lage dem og å distribuere dem. Hensikten har vært å beskytte velgere fra desinformasjon, men loven har vist seg vanskelig å håndheve. Ifølge Paris ligger uansett ikke hovedutfordringen i de politiske videoene, men i de amatørmessige, ofte pornografiske, «cheapfakes» variantene som lages med enkle app'er og programvare på smarttelefoner. Hvilket som helst bilde kan brukes i slike videoer og hvem som helst kan rammes.

Historien til australske Noele Martin illustrerer alvoret. I 2016 oppdaget hun at ansiktet hennes ved hjelp av photoshop var redigert inn i pornografiske bilder på internett. Som følge av dette ble hun utsatt for krenkelser og trakassering av vilt fremmede. Noele Martin var bare 18 år da hun ved å bruke bilder av seg selv i søk på google oppdaget at et bilde av henne var publisert på pornosider på internett. En ondsinnet person hadde stjålet en selfie fra Facebook-feeden hennes, redigert ansiktet inn i

pornografiske bilder og spredd dem på internett. Bildene sirkulerer fortsatt. Ofte er slike bilder brukt til utpressing. Dette er en stor psykisk påkjenning for en tenåring.

Politiet kunne ikke hjelpe henne og anbefalte å kontakte webmasteren til nettstedet. Som svar fikk hun flere bilder i retur med nye krav om penger. Politiet manglet åpenbart kunnskap og rutiner til å håndtere slike saker, og sendte offeret inn i en enda mer sårbar og potensielt farlig situasjon. Det er et problem at lovverket kun er nasjonalt og relativt smalt. Selv om politiet greide å fjerne bildene fra australske nettsider, greide de ikke å fjerne dem fra andre lands.

Dette er kun ett eksempel på hvordan noen blir ofre for cheapfakes eller forfalskede bilder. Det er ikke kun økonomisk utpressing som er et problem. Det kan også være svindel rettet mot eldre hvor man ringer med forfalskede stemmer, og for eksempel utgir seg for å være et familiemedlem som trenger penger.

Ikke bare negativt

"Technology is neither good nor bad; nor is it neutral."

Professor i teknologihistorie Melvin Kranzberg (1986)

Som med all teknologi har brukeren en viktig rolle i om teknologien skal brukes til godt eller vondt. Vi kan likevel slå fast at teknologi sjelden er nøytrale verktøy, men avhenger av eller kan utnyttes av mennesker med våre skjevheter, ideer og ukjente forutinntattheter.

Deepfakes kan dermed også brukes med gode hensikter.

Forskere ved Samsungs lab for kunstig intelligens i Moskva har laget en "levende" versjon av det berømte maleriet Mona Lisa av Leonardo Da Vinci. Ved hjelp av dyp læring får de maleriet til å bevege øyne, hode og munn.

I tillegg har Dali-museet i St. Petersburg, Florida, laget en utstilling ved navn Dalí Lives, hvor de har gjenskapt den avdøde kunstneren Salvador Dali med deepfake teknologi. Ved hjelp av mer enn 1000 timer maskinlæring lærte de opp programmet til å snakke som Dalí og faktisk reprodusere sitater fra den kjente kunstneren.

Deepfake-manipulasjoner har vist seg som utmerkede historiefortellere, og det er ikke bare kunsthistorie som blir formidlet. Det er blant annet laget video av den kjente fotballspilleren David Beckham som gir informasjon om malaria på ni ulike språk.

Videoen er utviklet av Synteshia, et selskap som bruker deepfake-teknologi til å lage overbevisende dubbing gjennom automatiserte ansiktsanimasjoner.

I følge forskere ved institutt for medisinsk informatikk ved universitetet i Lübeck, kan deepfake-teknologi også brukes til å forbedre kreftdiagnostisering. De skriver at fordi algoritmene er spesialiserte til å gjenkjenne mønstre i bilder, kan man mer presist og effektivt oppdage svulster eller abnormiteter fra røntgen CT, MR eller vanlig røntgen. Per i dag er det personvernregler som står i veien for å videreutvikle slik praksis, siden opptreningen av et program vil trenge store mengder bilder av ekte pasienter.

Dette fører oss, som i de fleste diskusjoner om teknologi, til spørsmålet: Hvor mye skal man regulere? Det finnes positive og negative bruksområder, og begge sider kan argumentere for viktigheten av eller faren ved reguleringer. Samtidig står personvern og demokrati på spill. Derfor trenger vi en bred debatt som involverer teknologer, politikere, politi og ikke minst sivilsamfunnet.

Hvor mye er vi villige til å gi for å bedre kreftdiagnostiseringen? Hvilke data vil vi gi slipp på for å stoppe overgrep av barn på nett? Disse spørsmålene er vanskelige, men like fullt viktige og avgjørende for hvilken fremtid vi går inn i. Vi kan ikke ta lett på slike spørsmål. Dette handler ikke bare om hvorvidt man skal regulere, men hvilken effekt det vil ha av å la kunstig intelligens ha tilgang til stadig mer av våre data. Noe som høyst sannsynlig gir tilgang også til land og regjeringer verden over. Vi må tenke 10, 20, 50, 100 år fremover og basere våre valg på de lange og strukturelle konsekvensene av valgene får.

De sanne fakta - Hva kan vi gjøre?

Politikere som sier ting de aldri har sagt, forfalskede bilder og lydklipp, utpressing og misbruk av sårbare mennesker, deepfake-teknologien er kommet for å bli og tas stadig mer i bruk. Det er avgjørende å følge utviklingen av dette og diskutere etikk og reguleringer. Hvis teknologen skal reguleres, hvordan kan det gjøres?

Teknologiselskaper har et viktig ansvar for hva som deles på deres plattformer. Dette ansvaret må lovfestes og det må utvikles protokoller for å merke og fjerne forfalsket innhold. I takt med utviklingen må også politiet oppdatere sin kunnskap på feltet.

For her er virkelig kunnskap og kompetanse avgjørende. Hvordan vurderer politiet muligheten for om videobevis er forfalskede? Myndigheter må stimulere til forskning og teknologiutvikling som raskt kan avgjøre om bilder eller videoer er forfalskninger. Slik teknologi må være tilgjengelig ikke bare for politiet, men for alle i samfunnet.

Det er også viktig å foreta en kritisk gjennomgang av lovverket for å se hvordan det kan brukes til å hindre misbruk av deepfake teknologi. Myndighetene må både klargjøre og gjøre det allment kjent hvor grensene går for hva som er lov og ikke lov i forhold til å bruke og spre bilder av andre enn deg selv.

Vi må i tillegg forstå at vi lever i en verden hvor det blir stadig vanskeligere å skille sannhet fra løgn og forfalskninger. Vi må lære å være mer kritisk til det vi leser i nyheter, sosiale medier, til kildene som er brukt og sammenhengen informasjon presenteres i. Skolen er et viktig sted å starte. Det er viktig at kommende generasjoner blir opplært i kritisk kildesjekk og at de forstår konsekvensene av teknologien og hvordan den blir brukt i dag. Ekspertise må lage gode undervisningsopplegg, og skolene må ta disse i bruk. Skoleprosjektet Tenk, fra faktisk.no, er et eksempel på godt arbeid på dette feltet. Her utvikles det ressurser til bruk i skolen om kritisk mediebruk og kildebevissthet.

Uavhengig av alle slike tiltak, må vi forstå at vi ikke kan sitte å vente til alt er lagt til rette for oss – lovverk, opplæring, osv. - vi må som borgere selv ta tak i dette.

Tenk på det som en type selvforsvar, for deg og familien din. Vi må utfordre oss selv til å alltid søke sannhet, og huske at det vi ser ikke alltid er det vi tror det er. I vår tid kan vi ikke lenger tro våre egne øyne – i hvert fall ikke alltid.