DEEPFAKES

– et undervisningsopplegg om kritisk medieforståelse

Co-financed by the Connecting Europe Facility of the European Union

Medietilsynet

Tenk.

Foto: metamorworks / Shutterstock / NTB

Pedagogisk opplegg deepfakes - forankring i fagfornyelsen.

Undervisningsopplegget egner seg på ungdomstrinnet og videregående, og det legger til rette for varierte læringsaktiviteter og refleksjon. Opplegget er lagt opp slik at man kan gjennomføre alt som en helhet, eller plukke ut deler av det.

Ved å gjennomføre dette opplegget vil elevene få øvelse i kritisk tenkning og kildebevissthet, noe som er i tråd med fagfornyelsen. Undervisningsopplegget er tydelig forankret i læreplanen gjennom overordnet del, blant annet i verdigrunnlaget, kompetansebegrepet og de tverrfaglige temaene. Opplegget er spesielt egnet i fag som norsk, samfunnsfag, KRLE og kunst og håndverk, men ettersom kritisk kildebruk er sentralt i mange fag kan det også fungere fint som et tverrfaglig undervisningsopplegg.

KOMPETANSEMÅL I FAGENE ETTER 10.ÅRSTRINN:

Norsk:

● Utforske og vurdere hvordan digitale medier påvirker og endrer språk og kommunikasjon.
● Informere, fortelle, argumentere og reflektere i ulike muntlige og skriftlige sjangre og for ulike formål tilpasset mottaker og medium

Kjerneelementer: Kritisk tilnærming til tekst.

Samfunnsfag:

● Vurdere på kva måtar ulike kjelder gir informasjon om eit samfunnsfagleg tema, og reflektere over korleis algoritmar, einsretta kjelder eller mangel på kjelder kan prege forståinga vår
● Utforske korleis teknologi har vore og framleis er ein endringsfaktor, og drøfte innverknaden teknologien har hatt og har på enkeltmenneske, samfunn og natur
● Utforske og reflektere over eigne digitale spor og høvet til å få sletta spora og å verne om retten ein sjølv og andre har til privatliv, personvern og opphavsrett
● Utforske ulike plattformer for digital samhandling og reflektere over korleis digital deltaking og samhandling påverkar forma på og innhaldet i samfunnsdebatten
● Reflektere over korleis identitet, sjølvbilete og eigne grenser blir utvikla og utfordra i ulike fellesskap, og presentere forslag til korleis ein kan handtere påverknad og uønskte hendingar
● Beskrive sentrale lover, reglar og normer og drøfte kva konsekvensar brot på desse kan ha for den enkelte og for samfunnet på kort og lang sikt

Kjerneelementer: Undring og utforskning, samfunnskritisk tenking og samanhengar

Foto: TippaPatt / Shutterstock / NTB

KRLE:
● Identifisere og drøfte etiske problemstillinger knyttet til ulike former for kommunikasjon
Kjerneelementer: Etisk refleksjon

Kunst og Håndverk:
● Utforske hvordan digitale verktøy og ny teknologi kan gi muligheter for kommunikasjonsformer og opplevelser i skapende prosesser og produkter
● Reflektere kritisk over visuelle virkemidler og eksperimentere med ulike visuelle uttrykk i en skapende prosess.


Foto: metamorworks / Shutterstock / NTB

Kjerneelementer: Håndverksferdigheter, kunst- og designprosesser og visuell kommunikasjon

KOMPETANSEMÅL I FAGENE ETTER VG1 STUDIEFORBEREDENDE UTDANNINGSPROGRAM
Norsk:
● Bruke ulike kilder på en kritisk, selvstendig og etterrettelig måte
Kjerneelementer: Kritisk tilnærming til tekst, muntlig kommunikasjon.

Samfunnskunnskap
● Utforske og presentere dagsaktuelle tema eller debatter ved å bruke samfunnsfaglege metodar, kjelder og digitale ressursar, og argumentere for sine eigne og andre sine meiningar og verdiar
● Utforske korleis interesser og ideologisk ståstad påverkar våre argument og val av kjelder, og reflektere over korleis det gir seg utslag i forskjellige meiningar
● Reflektere over eigne digitale spor, utforske kven som har tilgang til spora, og drøfte korleis data og personopplysningar kan brukast eller misbrukast
Kjerneelementer: Undring og utforsking, perspektivmangfald og samfunnskritisk tenking.

Deepfakes: Et kappløp mellom menneske og maskin

Framtidens falske nyheter blir mer livaktige enn noen kunne ha forestilt seg.

Av Martin Bergesen Publisert: 6. november 2020

I starten av Martin Scorsese-filmen «The Irishman», som dukket opp på Netflix i 2019, spiller Robert DeNiro en 36 år gammel mann som snart skal bli leiemorder for mafiaen.

DeNiro, 74 år gammel da innspillingen fant sted, ser definitivt ikke 36 år gammel ut i filmen.

Joda, Hollywoods fremste effektfolk har prøvd å forynge ham, men tross millionbudsjett har 36-åringen på skjermen gammelmannsøyne og dype furer i munnvikene der han kommer rullende i lastebilen sin.

Det siste innen filmmagi har sviktet. For å finne den nye magien må vi til YouTube-kanalen Sham00k.

Her ser vi samme scener fra samme film, med én viktig forskjell: Draget rundt munnen er mildere, og DeNiros blikk har gløden til en mann som ennå har 1990-klassikeren «Goodfellas» til gode. For å gni det litt ekstra inn er klippene sidestilt, så man kan sammenligne med originalen.

Dette er ikke utført på millionbudsjett, men på gøy.

Slik er kraften i deepfake-teknologien – et ektefødt barn av kunstig intelligens.

Lærer ansikter utenat

Ved hjelp av avansert maskinlæring trenes en algoritme opp på tusenvis av bilder av en person, helt til den kan alle fasettene av ansiktet utenat. Deretter kan den lime dette ansiktet over et annet i en video.

Dette kan gjøres ved hjelp av en vanlig datamaskin og fritt tilgjengelig programvare som, i sine enkleste former, kan brukes av hvem som helst etter en liten kom-i-gang-leksjon på YouTube.

Bruksområdene og kvaliteten på deepfakes utvikler seg i en rivende fart, og de blir stadig vanskeligere å avsløre.

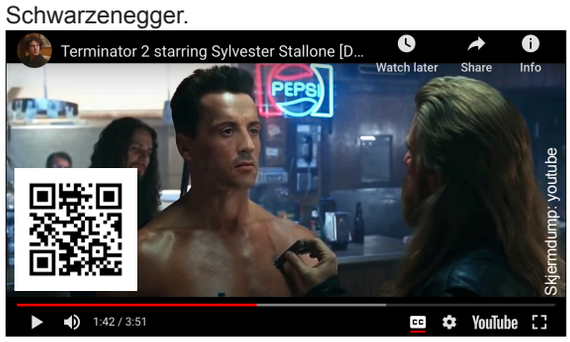
Ta nettstedet This Person Does Not Exist, som poster bilde etter bilde av maskingenererte ansikter.

Om vi så dem uten å vite at de var falske, ville vi kanskje ikke stilt spørsmål ved hvorvidt folkene på skjermen faktisk eksisterer. Rett og slett fordi de ser ut slik vi forventer at et ansikt skal se ut.

Vi står midt i et kappløp mellom teknologien og sanseapparatet vårt. Hvor bekymret bør vi være?

Gode og dårlige bruksområder

Det finnes mange gode bruksområder for deepfakes – alt fra nyskapende kunst og satire til billigere filmeffekter. Ofte er det gøy, som når vi kan se Sylvester Stallone spille i «Terminator 2» i stedet for Arnold Schwarzenegger.



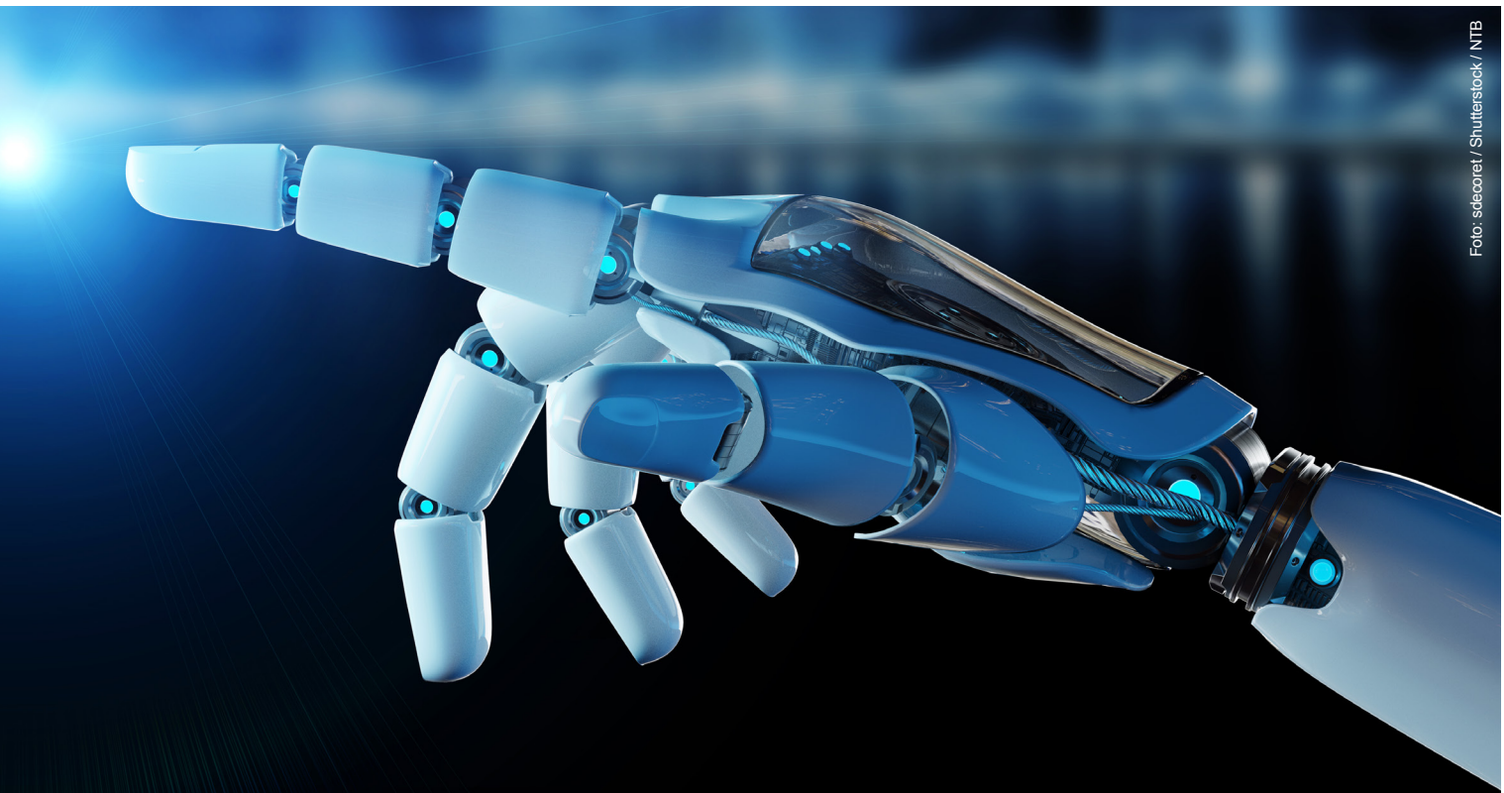


Foto: sdecoret / Shutterstock / NTB

Når deepfakes likevel skaper bekymring, er det på grunn av de mange problematiske brukstilfellene, som utpressing, mobbing og villedning.

Kjendiser og ekskjærester kan fremstilles i seksuelle situasjoner uten samtykke, mens ansiktet til et barn kan limes inn i et voldelig videoklipp og sendes som trussel til foreldrene.

Og, der enkeltmennesker kan traumatiseres, kan det på på samfunnsnivå påvirke både juridiske og demokratiske prosesser.

Et «avslørende» klipp av en politiker delt i sosiale medier et par dager før et valg rekker kanskje ikke avkreftes før etter valget. Og hva skjer med bevisførsel når man kan påstå at et videoklipp av politivold er forfalsket?

Lang lureri-tradisjon

Manipulerte og manipulerende medier har vært med oss til alle tider.

Dokumentsamlingen «Sions vises protokoller» fra starten av 1900-tallet ble gitt ut som bevis på at jødene konspirerte for å ta over verden. Ifølge Store Norske Leksikon ble dokumentene avslørt som en antisemittisk fabrikasjon allerede i 1921, men ideene lever likevel videre den dag i dag.

Josef Stalin fikk på sin side fjernet tidligere medarbeidere fra offisielle bilder når de ikke lenger var inne i varmen. I dag fjerner influensere gjerne en centimeter eller to av midjemålet på Instagram.

Ikke alle falsknerier forsøker å narre oss. På nettet var mp3-filene «Bushwhacked 1 & 2» tidlige viralhits i 2001 og 2003. Komiker Chris Morris hadde klippet ut ord og fraser fra talene til daværende president George W. Bush for å stokke dem om til satiriske remikser. Her lød lovnader om å gi alle amerikanske barn tre atomvåpen hver, samt dystre beskjeder som denne:

«And tonight I have a message for the people of Iraq: Go home and die.»

Skjermdump: youtube

Å katalogisere og sammenstille lydopptakene av Bush må ha krevd en god porsjon tid og talent - det lød dønn overbevisende.

Klippet vakte da også oppsikt, men på datidens nett spredte det seg saktere enn hva som er mulig i dagens sosiale medier. Talen var også for surrealistisk til å bli tatt for noe annet enn satire.

Men hva skjer når hvermannsen kan lage sin egen «Bushwhacked» på laptopen?

Kopierer stemmebåndene

På samme måte som at maskiner kan lære seg ansiktsformer, kan de nå også klone talemåtene og særegenhetene i folks stemmer. På YouTube finner vi kanalen Vocal Synthesis, i sin helhet dedikert til å produsere deepfakes av lyd.

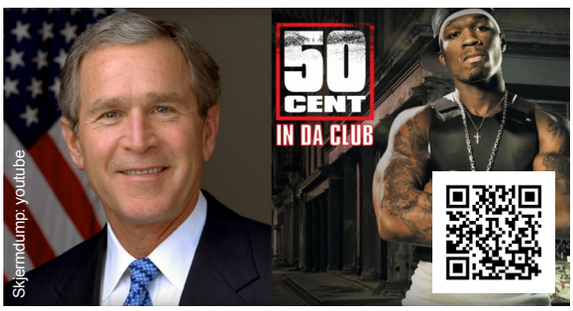
Her legges ord i munnen til alt fra rapperen Jay-Z (som forsøkte – uten hell – å få forfalskningen fjernet) til nåværende og forhenværende presidenter.

Først fores maskinen med stemmeopptak av Bush for å lære ham å kjenne. Så limes teksten til hitlåta «In Da Club» av rapperen 50 Cent inn i et tekst-til-tale-program. Og værsågod.

– I am into havin' sex, sier falske-Bush.

– I ain't into makin' love.

I lydklippet halter Bush gjennom teksten, og lyden knaser som om den er spilt inn i en bøtte – bivirkninger av at dataprosessen ennå ikke er presis nok.



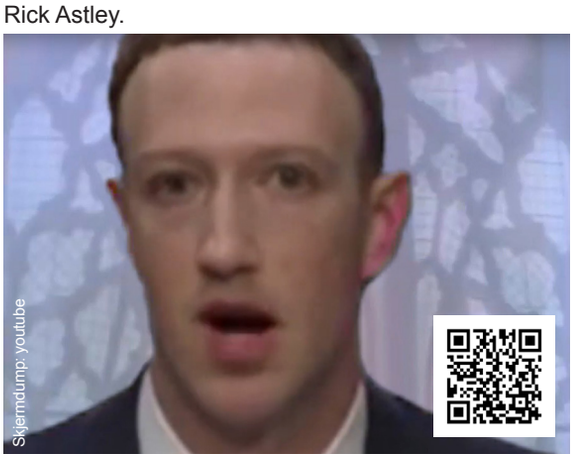
Men det er stemmen til Bush. Og dette er ikke en remiks av ting han har sagt. I stedet er stemmebåndene hans nå blitt som et piano alle kan spille på.

Bush er bare ett eksempel. Telefonnummer kan også forfalskes. Hva svarer vi når en god venn plutselig ringer og ber oss vippse over et kjapt lån?

Digitale marionetter

Ansikter kan trakteres på samme måte som stemmer. Der en god deepfake kan kreve tusenvis av bilder som kildemateriale og dagevis med arbeid, kan selv ett stillbilde få ganske livaktig mimikk.

Et eksempel på veldig lavterskel-deepfake er YouTube-kanalen Morgi Memes. Den har postet en video hvor alt fra dronning Elizabeth til Leonardo DiCaprio synger 80-tallshiten «Never Gonna Give You Up» av Rick Astley.



Videoen er i bunnsjiktet av deepfake-hierarkiet. Ansiktene strekker seg i unaturlige former, eller ter seg som stive masker.

De er lette å avsløre, fordi sanseapparatet vårt har forsvarsmekanismer mot slikt. Kunstig fremstilte ansikter og stemmer påkaller gjerne «den uhyggelige dalen»-effekten («uncanny valley») – følelsen av at noe er galt. Det er noe dødt over øynene. Stemmen er for robotisk.

Gode deepfakes har imidlertid potensial til å forbigå dette forsvaret. Jo bedre maskinlæringen blir, jo mer forsvinner de synlige kjennetegnene på forfalskning.

Hva skjer når vi ikke lenger kan stole på det øynene og ørene våre forteller oss?

Provoserende politikere

De fleste av oss har allerede en viss skepsis til ting vi ser i media, enten det er de redigerte livene til vennene våre eller nøye planlagte utspill fra politikere.

Likevel lar mange seg lure av falske nyheter, hvor selv en uærlig sammenstilling av tekst og bilde kan få sinnene i kok.

Et nylig eksempel er en post med bilde av Erna Solberg i militærhjelm sammen med en soldat, og en påstand om at hun vil ha utvidede fullmakter til å sette inn den norske hæren mot folket.

Foto: Blue Planet Studio / Shutterstock / NTB

Egentlig var det snakk om et lovforslag som ga Forsvaret adgang til å pålegge restriksjoner for egne ansatte i smittevernsøyemed.

Nå er deepfake videoer og lydklipp – gjerne i kombinasjon – modne for verktøykassen til dem som produserer denne typen falske nyheter.

Da kan det dukke opp videoer i Facebook-feeden hvor en politiker tilsynelatende sier noe vilt provoserende, noe som så spres og gjør vedkommende til gjenstand for allmenn fordømmelse.

Å undersøke og avkrefte hvorvidt noe slikt er fabrikert eller ei er mulig, men kan være både knotete og tidkrevende. Samtidig får den kjedelige sannheten sjelden samme spredning som den medrivende løgnen.

Denne dynamikken sår uvisshet, og skaper et mulighetsrom for politikere og andre som vil påstå «det sa jeg aldri», selv når ordene deres er tatt opp på video. Det samme ser vi allerede med stillbilder, som når britiske Prins Andrew avviser ektheten av bildet hvor han har armen rundt en ung jente som senere har påstått å ha blitt tvunget til å ha sex med prinsen.

Maskin mot maskin

Som med falske nyheter ellers er det å utvise kildekritikk fortsatt den beste motgiften, både når løgner spres og når sannhet benektes. Er det for godt til å være sant? Stoler vi på den som har postet dette? Hva er konteksten?

Men: Dette krever mye.

Vi skal være årvåkne og kritiske til alt av medieinntrykk vi scroller oss gjennom en sen kveldstime, og vi skal være skeptiske til venner vi ellers stoler på når de deler noe på Facebook.

Den sunne fornuften, påskrudd til alle døgnets tider - høres ikke det litt utmattende ut?

Kanskje er det bedre å bekjempe ild med ild. Maskinlæring kan nemlig også snus til å avdekke hvorvidt en video har blitt klusset med.

Microsoft annonserte nylig et verktøy døpt Video Authenticator, beregnet spesielt på valget i 2020.

Med en kunstig intelligens trent opp på å vurdere nesten identiske videoer, hvor én er en deepfake og én er originalen, vil verktøy som dette forhåpentligvis automatisere avsløringen.

Der et menneskeblikk lar seg lure, kan maskinblikket fokusere på å små uregelmessigheter i bildekodingen.

Om videoer går gjennom et slikt filter før de slippes løs på sansene våre, kan de kanskje stanses, eller merkes tydelig som manipulerte.

Men den samme lærdommen som muliggjør verktøyet, kan også brukes til å styrke deepfake-teknologien.

Kappløpet fortsetter.

Skjermdump: youtube

Deepfakes: Fra gøy til guffent på et blunk

I framtida er vi alle Billie Eilish.

Av Martin Bergesen Publisert: 6. november 2020

– Jeg tror dette er den mest ubehagelige... Å gud, nei!

Gjennom webkameraet ser vi den svenske YouTube-stjerna Felix «PewDiePie» Kjellberg stirre på skjermen sin i avsky.

Foran ham avspilles en video hvor han ser seg selv danse.

(Scan QR-koden på bildet over for å se videoen)

Rettere sagt, han ser ansiktet sitt limt på den halvnakne vrikkende kroppen til det sosiale media-fenomenet Belle Delphine, hun som er mest kjent for å ha solgt badevannet sitt for 30 dollar flaska.

Og mens Delphine vrikker, mimer PewDiePie-fjeset til den avsindige sangen hun danser. Det ser – nesten – ekte ut. Kanskje øynene hans er litt for blå, men ansiktsmimikken er nesten uhyggelig lik.

PewDiePie har allerede ledd og tøyset seg gjennom andre videoer, hvor kreative moroklumper har plassert ham i hovedrollen både i «Aladdin» og «Captain America».

Nå virker svensken et øyeblikk genuint ukomfortabel. Han gjør en grimase og snur seg bort.

– Jeg vil ikke se dette.

Avansert kunstig intelligens

Hva gjør det med oss å se videoer hvor vi gjør og sier ting vi aldri kunne funnet på? Ler vi? Eller blir vi ubekvemme? Om ikke lenge kan dette være en like normal opplevelse for oss alle som det å tøyse med ansiktsfiltrene på Snapchat er i dag.

Fjesene og stemmene våre er blitt fritt vilt. De kan brukes av hvem som helst til hva som helst, enten det er for moro skyld eller for å skade oss.

Dette er mulig på grunn av deepfakes, en teknologi som bygger på maskinlæring.

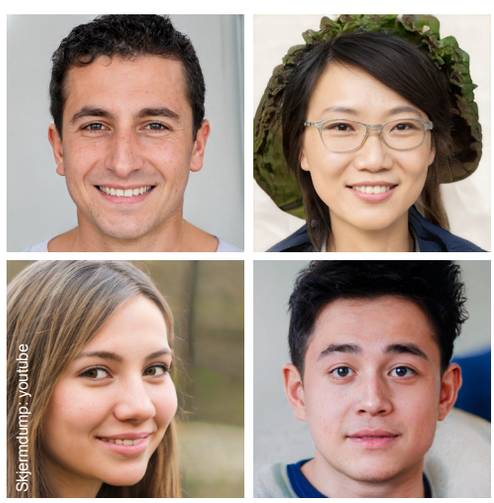
Ved å fôre en kunstig intelligens med videoer og bilder av et ansikt (for eksempel PewDiePie) kan en kunstig intelligens analysere det og lære seg hver lille nyanse og vinkel utenat.

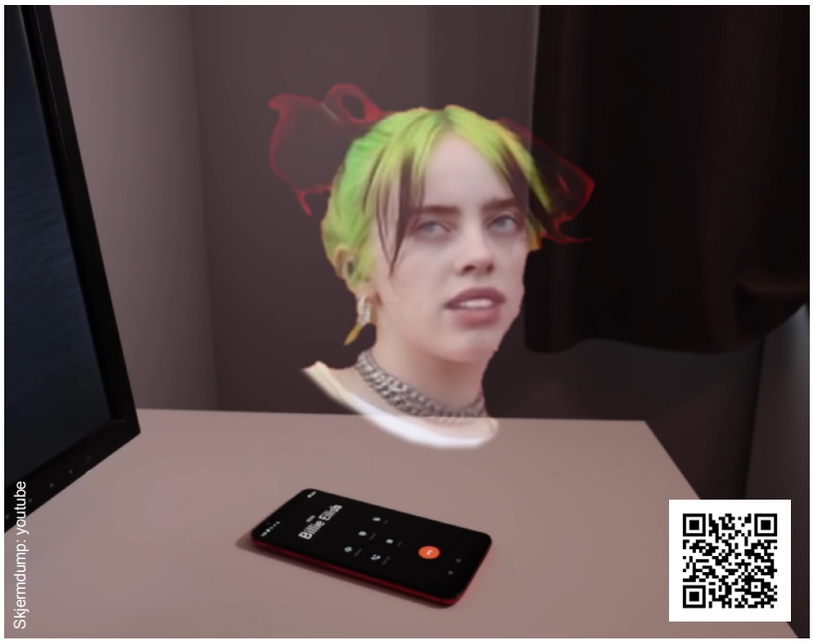
Deretter kan den spytte ut overbevisende kopier av fjeset hvor enn det skal være. Og dette begrenser seg ikke til kjendiser. Faktisk er teknologien på sitt mest imponerende når den ikke begrenser seg til virkelige mennesker.

Fordi den kunstige intelligensen vet hvordan et menneskeansikt skal se ut, kan nettstedet This Person Does Not Exist nå poste bilde etter bilde av uhyggelig realistiske dataskapte fjes.

Klarer du å se hvilke to av disse fire personene som ikke eksisterer?

(Fasit i bunnen av saken!)


Skjermdump: youtube



Billie Eilish-hologram

Det er ikke bare ansikter som kan klones – det finnes også deepfakes av menneskelig tale. Våren 2020 la den amerikanske programmereren Will Kwan ut en video av hvordan han brukte datakode som lå fritt tilgjengelig på nett til å skape en digital versjon av ansiktet og stemmen til popstjernen Billie Eilish.

(Scan QR-koden på bildet over for å se videoen)

Der ansiktet var lite overbevisende – Kwan animerte et stillbilde av Eilish ved å få det til å mime etter sitt eget ansikt – var talemålet hennes noe annet.

Kwan kjørte noen få sekunder av et Eilish-intervju gjennom en algoritme. Noen sekunders tale var alt maskinen behøvde for å lage en lavoppløselig variant av stemmen hennes. Ved å bruke et tekst-til-tale-program kunne Kwan deretter få henne til å si noen ganske tamme setninger:

– Wow. Føles godt å være i live. Takk, Will. Jeg setter virkelig pris på alt arbeidet du la inn i å få meg til å se og høres realistisk ut.

Det hele virker som et uskyldig eksperiment, helt til slutten av videoen.

Kwan er sponset av et selskap som selger nettadresser, og i den forbindelse får han Eilish-hologrammet sitt til å snakke litt mer:

– I disse dager er det vrient å finne et godt domenenavn med de populære adressene, som .com og .net, sier popstjerna, før hun går videre til å reklamere for Kwans sponsor.

Et kommersielt budskap den virkelige Eilish ikke har samtykket til, framført av hennes stemme og ansikt.

Hva ville Eilish selv tenke om hun så denne videoen? Er dette punktet hvor det hadde gått fra gøy til guffent?

Kan klone hvem som helst

Det Kwan gjør krever en del teknisk kompetanse, men denne teknologien blir stadig mer brukervennlig. Med de samme verktøyene som Kwan bruker kan også kjæresten din eller bestevennen din klones.

Hvordan ville du reagert om de ringte deg opp og sa at de var i trøbbel og trengte penger, eller kanskje sa noe forferdelig om deg?

Og hva ville du tenkt om det dukket opp en video hvor du gjør og sier ting du vet du aldri ville gjort?

Hvordan ville det vært om videoen spredte seg, og klassekameratene dine trodde på den?

Byttet ut Harrison Ford

Det er ennå en stund til det blir enkelt å lage en såpass overbevisende forfalskning, men umulig er det ikke. YouTube-kanalen Sham00k har for eksempel

byttet ut fjeset til Harrison Ford i de gamle «Indiana Jones»-filmene med actionstjerna Chris Pratt.

(Scan QR-koden på bildet over for å se videoen)

Endringen er så godt gjennomført at folk som ikke kjenner de gamle filmene lett kan tro de ser Pratts nyeste film.

Dette krever mer maskinkraft og innsats enn det Will Kwan gjorde med Billie Eilish. For å oppnå dette resultatet oppgir Sham00k å ha brukt en kraftig datamaskin og latt den kunstige intelligensen analysere over 20.000 individuelle ansiktsbilder over en periode på 128 timer.

Å nå denne graden av realisme vil imidlertid bli enklere for folk flest i årene som kommer.

Hva skjer da, om dette for eksempel tas i bruk av mennesker som ønsker å påvirke deg politisk ved å lage falske videoer av meningsmotstandere? Lar du deg lure, eller klarer du å luke ut deepfakene?

I så fall, hvordan?

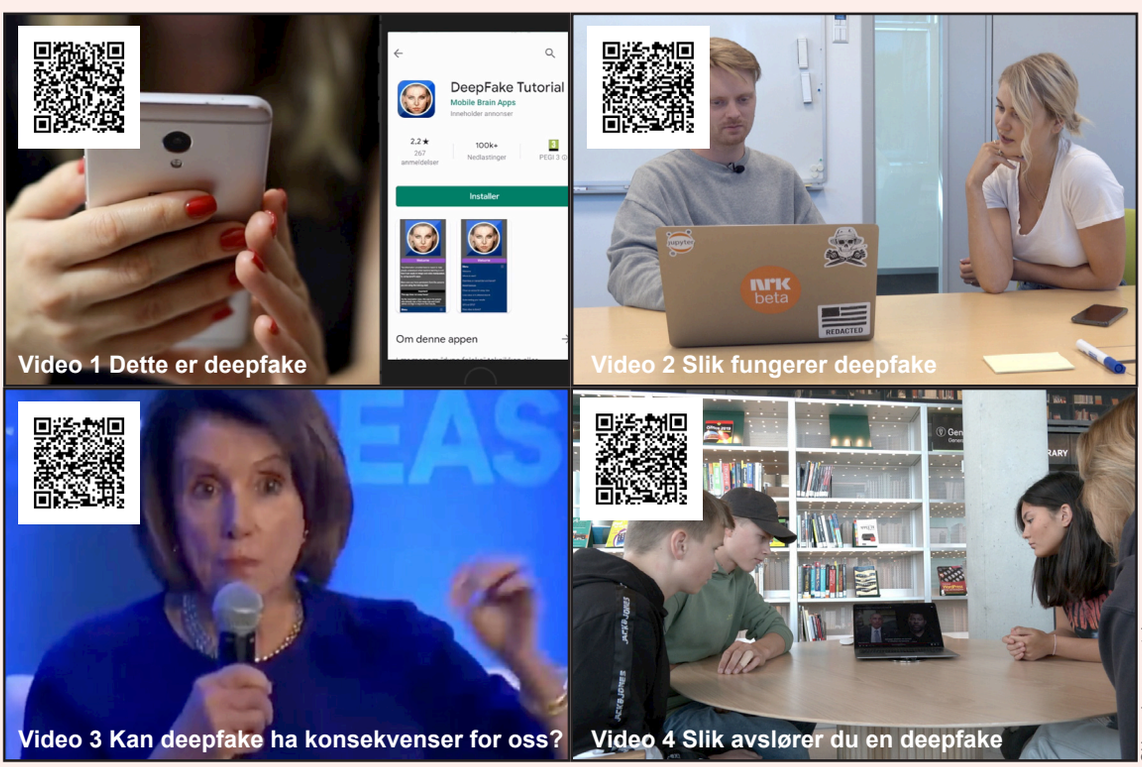
Diskusjonsspørsmål

1. Hvorfor tror du Piewdepie opplevde det så ubehagelig å se seg selv i dansevideoen som Belle Delphin?
2. Hvordan kan deepfake brukes til å ha det gøy?
3. Hvordan kan bilder fra “Thispersondoesnotexist.com” bli misbrukt?
4. Hvorfor tror du Will Kwan valgte å lage filmen av Billie Eilish?
5. Hva kan du gjøre for å unngå å bli lurt av en deepfake som ringer og ber om penger?

FASIT: Alle fire er datafjes. Bekløger. Om du vil teste deg på ordentlig kan du gå til nettstedet WhichFaceIsReal.com.

Undervisningsopplegg deepfake

Diskusjonsspørsmål til videoene:



Skjermdump: youtube

Video 1

1. Hva er forskjellen mellom en deepfake og en vanlig video?
2. Peter Brandtzæg sier at deepfakes begynte med gode hensikter, men så har det blitt oppdaget av folk med onde hensikter. Hva tror du han mener med det?
3. Hva tenker du er fordelene med deepfakes?

Video 2

1. Hva kan være positive og negative sider ved at denne teknologien kan få folk til å snakke andre språk?
2. Hvorfor tror du det lages lite deepfake i Norge?
3. Hvordan kan deepfakes brukes for å lure deg?

Video 3

1. Hvordan kunne noen brukt deepfakes til å påvirke meningene dine før et valg?
2. I videoen sies det at vi må *jobbe med*, og ikke *kjempe mot* den teknologiske utviklinga. Hvordan kan vi gjøre det?
3. Hvordan tror du vi bruker deepfakes om ti år?

Video 4

1. Hvordan kan du finne ut hvem som er avsender av en video som dukker opp i feeden din?
2. Hvordan kan vi unngå å bli lurt av falske videoer på nett?
3. Hva er den slemmeste deepfaken du kan se for deg?

Diskuter casene

Case 1:

Du er med i en gruppechat og en åpenbart manipulert film har blitt delt. Flere i klassen har fått ansiktene sine manipulert inn i filmen. Hvilke rettslige og personlige konsekvenser kan en slik film få for de som har delt videoen og de som blir utsatt for det?

Case 2:

Deepfake vil gjøre det mulig å manipulere en telefonsamtale slik at det fremstår som at en nær venn av deg ringer og ber deg vippse penger til nummeret det ringes fra. Har du noen tips til hva du kan gjøre for å unngå å bli lurt av denne svindelen?

Dilemma, fysisk i klasserommet:

Del klasserommet i to der den ene siden representerer det ene svaret og den andre siden det andre svaret. Elevene viser hva de ville valgt ved å gå til den ene eller andre siden av klasserommet. Dilemmaene har ikke noe rett eller galt svar. Poenget er at det kan skape bevissthet rundt egne valg, og være et godt utgangspunkt for en klasseromsdiskusjon rundt temaer innen kildebevissthet og kritisk mediebruk. Her er lærerens oppfølgingsspørsmål sentrale.

Ville du helst:

1. At det blir spredt en ekte video av deg som fremstiller deg dårlig, eller at det blir spredt en deepfake video av deg som fremstiller deg bra?

2. Lese en ekte nyhet som er kjedelig, eller en falsk nyhet som er fengende?

3. Være delaktig i at en deepfake av dine venner blir spredt, eller i at en deepfake som henger ut en kjendis blir spredt?

4. Ville du delt en morsom deepfake av læreren din som danser på sosiale medier, eller ville du ikke?

5. Ville du helst at all teknologisk utvikling skulle stoppet nå, eller at det fortsetter slik at vi får sett konsekvensene av det?

Ranger disse påstandene etter hva dere synes er best og verst:

Elevene kan jobbe i grupper eller par når de skal rangere påstandene. De påstandene som elevene anser som verst skal øverst og deretter rangere resten nedover. Denne aktiviteten kan være et godt utgangspunkt for en klasseromsdiskusjon der gruppene argumenterer for sine topp tre verste påstander.

- Ta videoopptak av noen uten at de vet om det
- Dele et bilde du har fått av noen uten å spørre om det er greit først
- Manipulere ansiktet til noen inn i en reklame uten tillatelse
- Lage en deepfake av stemmen til noen der de sier ting de ikke har sagt
- Spre en video med deepfake av noen, som setter dem i dårlig lys
- Ta bilde av noen når de ikke vet det, og dele bildet videre
- Spre et lydklipp med deepfake, som setter dem i dårlig lys
- Lage en deepfake av noen uten at de vet om det

Fordypningsoppgaver:

Deepfake er lyd, bilde eller video som er satt sammen med kunstig intelligens. I mange sammenhenger blir deepfake brukt til å påvirke eller lure folk. I denne oppgaven skal du jobbe med en utfordring knyttet til deepfake. Velg en av oppgavene.

1. Du har fått i oppdrag fra Snapchat å forklare for deres nye brukere hva deepfake er. Skriv en kort forklaringstekst eller lag en kort film der du forteller hva deepfake er og hva det blir brukt til. Informasjonen skal være tilpasset en 13-åring.
2. Du er statsminister i Norge og i det siste har det blitt spredt et rykte om deg, og det spres nå en deepfake som underbygger dette ryktet. Du har fått to minutter taletid på Dagsrevyen i kveld. Hva vil du si for å forsvare deg? Velg selv om du skal lage video eller skrive innlegget.
3. Hvordan kan deepfake påvirke oss? Velg om du vil lage video eller skrive en tekst der du reflekterer rundt konsekvensene av deepfake.

Quiz:

1. Deepfake er:

A: En falsk fremstilling basert på kunstig algoritme
B: En falsk fremstilling basert på kunstig intelligens
C: En falsk fremstilling basert på intelligent kunst

2. Deepfake kan omfatte:

A: Manipulering av bare video
B: Manipulering av bare lyd
C: Manipulering av både lyd, bilde og video

3. For å avsløre en deepfake kan du:

A: Bruke kildekritiske teknikker og eventuelt avansert programvare
B: Se etter om bildet er uskarpt i kantene
C: Se etter om leppebevegelser og tekst er synkroniserte.

4. Hvorfor kan deepfakes være farlig?

A: Fordi det utfordrer oss til å ikke kunne vite hva som er sant og usant på internett
B: Fordi deepfakes kun finnes på nettsider med virus som kan ødelegge din PC
C: Kun fordi mange manipulerer seg penere enn de er i virkeligheten

5. Du får tilsendt et bilde som åpenbart er manipulert av en medelev i klassen. Hvilke handlinger kan være straffbare for deg å gjøre?

A: Se på bildet
B: Ikke informere om bildet til personen som er blitt utsatt for bildemanipulering
C: Ta printscreen av bildet og be om penger fra medeleven din for å slette bildet

6. Hvordan skal vi håndtere deepfakes nå og i fremtiden?

A: Vi må prøve å slette alle deepfakes som finnes på internett
B: Vi må utvikle kompetanse som gjør det lettere å avsløre deepfakes og være forberedt på at teknologien forbedrer seg enda mer.
C: Det må bli forbudt å bruke teknologi som kan lage deepfakes

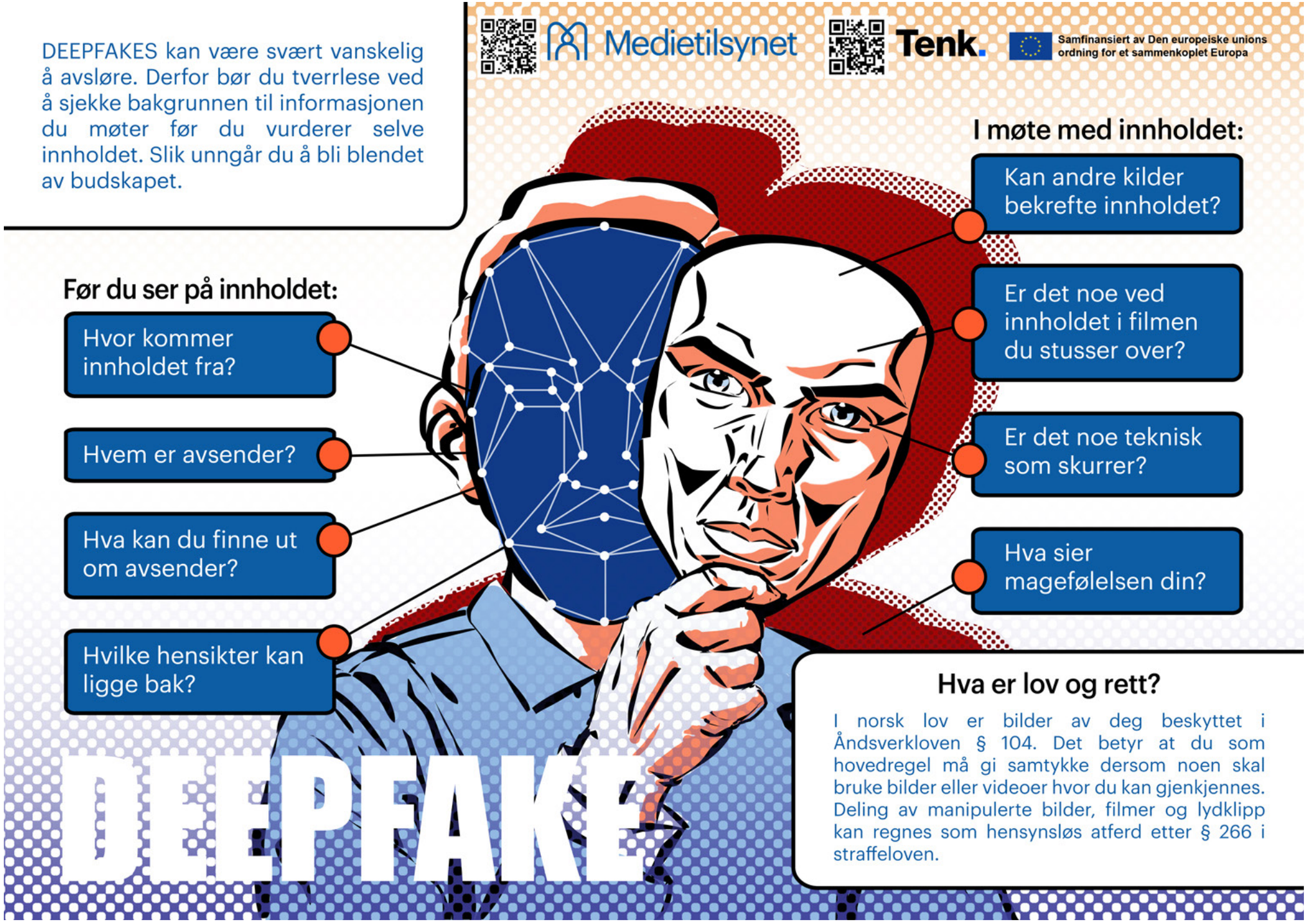
Fasit til quizen: 1b, 2c, 3a, 4a, 5c, 6b

DEEPFAKES kan være svært vanskelig å avsløre. Derfor bør du tverrlese ved å sjekke bakgrunnen til informasjonen du møter før du vurderer selve innholdet. Slik unngår du å bli blendet av budskapet.

Medietilsynet

Tenk.

Samfinansiert av Den europeiske unions ordning for et sammenkoplet Europa



Før du ser på innholdet:

- Hvor kommer innholdet fra?
- Hvem er avsender?
- Hva kan du finne ut om avsender?
- Hvilke hensikter kan ligge bak?

I møte med innholdet:

- Kan andre kilder bekrefte innholdet?
- Er det noe ved innholdet i filmen du stusser over?
- Er det noe teknisk som skurrer?
- Hva sier mageføelsen din?

DEEPFAKE

Hva er lov og rett?

I norsk lov er bilder av deg beskyttet i Åndsverkloven § 104. Det betyr at du som hovedregel må gi samtykke dersom noen skal bruke bilder eller videoer hvor du kan gjenkjennes. Deling av manipulerte bilder, filmer og lydklipp kan regnes som hensynsløs atferd etter § 266 i straffeloven.

Illustrasjon: Marius Vibe

Aktivitet knyttet til plakaten:

Jakten på deepfake

Deepfake kan være flere ting. Det kan være bilde, lyd eller film som er laget ved hjelp av kunstig intelligens. De beste deepfakes er så godt laget at det kan være umulig å se på bildet, videoen eller høre på lyden at det du har fått er falskt.

De to filmene under av Kim Kardashian og Mark Zuckerberg er begge deepfake, de er altså falske. I denne oppgaven skal du velge en av filmene og bruke plakaten "Deepfake" til å gjøre undersøkelser som kan bevise at filmen du har valgt er falsk. Oppgaven kan løses sammen med noen eller du kan jobbe alene. Hva klarer du å finne ut? Det kan være lurt å åpne lenken i et nytt nettleservindu får å få frem nyttig bakgrunnsinformasjon om filmen.

Velg en av filmene:

Kim Kardashian: https://www.instagram.com/p/ByKg-uKlP4C/

Mark Zuckerberg: https://www.youtube.com/watch?v=3f66kBwfMto

Før du ser filmen:

Det er viktig å sjekke bakgrunnen til informasjon du møter. Det kan være lett å la seg rive med og glemme å sjekke kilden. Svar på spørsmålene på venstre side av plakaten og noter stikkord om det du finner. Hvilket inntrykk får du av filmen ut fra det du har funnet ut?

Se filmen du valgte.

Etter at du har sett filmen:

Diskuter spørsmålene til høyre på plakaten. Er det noen av svarene som overrasker deg?

Felles oppsummering i klassen:

Dette er to deepfakes som er svært godt laget. Del ideer til hvordan dere kan møte deepfake i fremtiden. Er det noen av spørsmålene på plakaten som dere synes var ekstra viktige? Kommer dere på andre spørsmål man kan stille seg selv i møte med deepfakes?

Lyst på en ekstra utfordring?

Den originale videoen av Kim Kardashian er slettet fra YouTube, men videoen er ikke slettet fra Instagram. Hvorfor sletter ikke Instagram filmen? Og hvorfor sletter YouTube filmen? Hvem er du mest enig med?