



Semesterprojekt

Roskilde universitet

2. semester – Forår/Sommer 2024

Eksamensgruppenummer: S2424791990	
Deepfakes og Demokrati	
Gruppens medlemmer:	
Julia Lundager	Stud-Julial@ruc.dk
Liva Plougmann Sørensen	Stud-Plougmann@ruc.dk
Liva Adelholm Baunsgaard	Stud-Adelholm@ruc.dk
Karl Emil Krabbe	Stud-Karlk@ruc.dk
Daniel Pedersen	Stud-Danielp@ruc.dk
Johan Aksel Maribo Andreasen	Stud-Maribo@ruc.dk
Vejleder: Jens Ulrik Hansen	
Dato: 06/06 - 2024	

Indholdsfortegnelse

Abstract	4
Indledning og problemfelt	5
Teori	7
Hvad definerer AI?	8
At handle menneskeligt.....	8
At tænke menneskeligt.....	9
At handle rationelt	9
At tænke rationelt	10
Falske nyheder på sociale medier	10
Redegørelse for forskellen mellem des- og misinformation	12
Ekkokamre	12
Confirmation bias	14
Luhmann – “Theory of Trust”	15
Robert Dahl – “Democracy and its critics”	16
Metode	17
TRIN-modellen som rammeværktøj	17
Studie af den eksisterende litteratur	20
Interviews af Tanggaard og Brinkmann	23
Anvendt interviewguide	24
Analyse	26
Teknologisk analyse	26
Neurale netværk.....	27
Deep Learning	29
CNN.....	30
GAN	31
Diskriminator	32
Generator.....	32
Autoencoders.....	34
Natural Language Processing (NLP)	35
High-performance computing (HPC).....	36
Systemer	38
Sociale medier og desinformation	39
Nyheder	40
Ekkokamre & Confirmation Bias	41
Delkonklusion	41
Interviews	42
Interview 1.....	43

Interview 2.....	44
Interview 3.....	45
Interview 4.....	46
Interview 5.....	47
Konklusioner fra interview	48
Demokratisk analyse	49
Tillid i et demokrati.....	50
Dahls demokratiske kriterier.....	52
Delkonklusion	53
<i>Diskussion</i>	54
Teoretiske overvejelser	54
Metodiske overvejelser	57
<i>Konklusion</i>.....	59
<i>Litteraturliste:</i>	60
Artikler	60
Hjemmesider	62
Undervisningsmateriale:.....	67

Abstract

This paper aims to investigate the effects the deepfake technology has on citizens in democracies. It commences with a theoretical framework, which lays the foundation for the technological analysis and the analysis of the impact it has on a democracy and its citizens. The success-criteria for a democracy are in this paper defined by Robert Dahl and Niklas Luhmann, whose work respectively deals with democracy and trust. The technological analysis highlights the intricate web of systems that are required in the making of a deepfake, including GAN (Generative Adversarial Networks), CNN (Convolutional Neural Networks), NLP (Natural Language Processing), Autoencoding, and Deep Learning. Subsequently, the paper concludes that deepfakes are not a threat to democratic citizens, and their ability to participate in a democracy, unless it works together with phenomena such as social media, echo chambers, fake news and confirmation bias.

Keywords: teknologi, deepfake, algoritmer, sociale medier, demokrati, systemer, fake news, tillid, ekkokamre

Indledning og problemfelt

Teknologi findes efterhånden i samtlige aspekter af hverdagen, og udviklingen af denne foregår efterhånden i et accelererende tempo (Government et al., 2016). Skoler og uddannelser bliver mere og mere digitaliserede, omsorg og pleje efterlades i større grad i hænderne på teknologiske artefakter og systemer, og sågar politik og debat er blevet præget af den teknologiske udvikling.

”... the rate and evolutionary power of technological developments will accelerate in the years to come. Digital development will be so fast, profound and unpredictable that it will challenge and change society in ways we cannot even begin to imagine.” Således skriver den danske regering i 2016 (Ibid., s. 1, l. 19-24).

Et biprodukt af den teknologiske acceleration er den længe debatterede, men først for nyligt for alvor udviklede teknologi, *kunstig intelligens*. Kunstig intelligens er ikke i sig selv et nyt fænomen, og man har siden 1960'erne haft samtaler om netop dette (Chen & Chen, 2022). Den stigende hastighed, hvori teknologi udvikler sig, har dog også efterladt sit præg på den kunstige intelligens, og de seneste par år har det for alvor taget fart. En subkategori af kunstig intelligens er Generative Artificial Intelligence (GAI). GAI er en type kunstig intelligens der kan genere nyt og originalt indhold (Adobe, u.å). Det vil sige, at man potentielt vil kunne genere indhold, både af fiktive, men også af rigtige personer samt karakterer, som ytrer sig, eller opfører sig, på en måde der ikke er sand. Dette rejser flere problemstillinger, hvilket vi i følgende rapport vil undersøge.

Helt konkret tager vi fat i fænomenet *deepfakes*, som er videoer, billeder eller lydfiler, af folk der foretager sig handlinger, der ikke er sket i virkeligheden. Teknologien kan blandt andet benyttes til at sprede politiske budskaber, hvilket kan have potentiale til at påvirke udfaldet af valgkampe (Ledford, 2024). Dette rejser vigtige spørgsmål om teknologiens indflydelse på samfundsstrukturer og individets opfattelse af dette (Ibid.).

Forestil dig de gammeldags satiretegninger i Politiken, hvis de havde synes helt ægte.



Figur 1, politisk satireillustration, Als, 2022

Havde Figur 1 set "ægte" ud, og havde det været svært at opdage at det var en satirisk tegning, kan der være anledning til bekymring, da det ville kunne lægge udsagn i munden på danske politikere, som de aldrig har sagt. Man ville dermed kunne få politikere til at fremstå i særligt dårligt lys, og potentielt influere folks holdninger til dem og deres politik.

Selv hvis nogle mennesker godt var klar over at den var falsk, ville man så stole blindt på, at ens medborgere, der ligeledes er stemmeberettigede, og dermed indgår i et folkestyre, kan gennemskue hvad der er ægte og hvad der ikke er? Ville man stole på at en medborger ville være i stand til at forvalte sin stemmeret, og dermed have en mulig medbestemmelse over dig, hvis de ikke havde adgang til sande informationer? Ville man stole på den danske statsminister, hvis man vidste at den pågældende statsminister fik leveret usande informationer? Vi ser her et enormt potentiale for mistillid mellem medborgere, som videre vil sætte spørgsmålstejn ved borgeres evne, og lyst, til at indgå i et demokrati.

Dette er essensen af vores opgave, og vi har til følgende rapport udarbejdet en problemformulering der lyder således:

Hvilken rolle spiller deepfakes i informationslandskabet, og hvordan vil det kunne påvirke menneskers tillid til, samt evne til, at indgå i et demokrati?

Til ovenstående problemformulering, har vi udformet fem arbejdsspørgsmål, som danner ramme for vores opgave:

1. Hvad definerer begrebet deepfake?
2. Hvordan fungerer deepfake-teknologien?
3. Kan deepfake-teknologien og dens anvendelse, have en indflydelse på den enkelte borgers tillid til både hinanden, men også institutionerne?
4. Hvad definerer en borgers evne til at indgå i et demokrati?
5. Kan deepfakes og den dertilhørende desinformation, forringe borgeres evne til at indgå i et demokrati?

Vi vil i følgende rapport definere *informationslandskabet* som værende de medier, fora og andre kilder, hvor borgere i et demokrati får deres information fra. Dette inkluderer, men er ikke udelukkende, sociale medier, traditionelle nyhedsmedier og formel eller uformel dialog mellem borgere.

Teori

Vi vil i følgende afsnit redegøre for vores valgte teorier, der skal danne ramme for vores besvarelse af vores problemformulering. Indledningsvist vil vi introducere begrebet kunstig intelligens, og definere det ud fra Stuart Russel og Peter Norvigs bog fra 1995, "Artificial Intelligence: A Modern Approach". Herefter vil vi opsætte en teoretisk ramme om falske nyheder på sociale medier, der tager udgangspunkt i en tekst af Femi Olan et al., "Fake news on Social Media: the Impact on Society". Efter dette vil vi definere begrebet ekkokamre, samt følge op med en teori om confirmation bias, et begreb først introduceret af Peter Wason. Til sidst vil vi opstille to teorier, der skal danne ramme for vores definition af et effektivt og velfungerende demokrati. Disse teorier er henholdsvis Robert A. Dahls demokratikriterier fra hans bog "Democracy and its critics", og Niklas Luhmanns teori "Theory of Trust", fra hans bog "Trust and Power".

Hvad definerer AI?

En teoretisk ramme for, hvordan *kunstig intelligens (AI)* kan klassificeres og forstås, er præsenteret af Stuart Russell og Peter Norvigs bog "Artificial Intelligence: A Modern Approach" fra 1995. Russell og Norvig er to fremtrædende skikkelser inden for feltet AI, da de begge har bidraget betydeligt til forskning inden for kunstig intelligens (Smith & Zadeh Chair, u.å; Stanford University, u.å).

Ifølge forskerne, er der fire tilgange til, hvordan en maskine kan defineres som "intelligent", som giver en struktureret måde at evaluere og udvikle AI-systemer på (Russell & Norvig, 2009). De fire omtalte tilgange omfatter:

1. At handle menneskeligt
2. At tænke menneskeligt
3. At handle rationelt
4. At tænke rationelt

At handle menneskeligt

Denne tilgang fokuserer på, at en AI skal kunne udføre opgaver på en måde, der ligner menneskers måde at handle på. For at opfylde dette krav skal et AI-system være i stand til at handle på måder, som umiddelbart kan genkendes som "menneskelige" i deres adfærd. Til dette formål anvendes Turing-testen, der kan vurdere om en maskine kan demonstrere adfærd der er på et niveau der er sammenligneligt med mennesket (Russell & Norvig, 2009). For at bestå Turing-testen skal en AI-model være i stand til at:

1. Udføre naturlig sprogbehandling, for at kommunikere effektivt på et menneskeligt sprog.
2. Have vidensrepræsentation, for at opbevare den information den modtager og kender til.
3. Ræsonnere automatiseret, for at kunne svare på spørgsmål og drage nye konklusioner.
4. At justere sig efter nye forhold og identificerer, samt udvide mønstre, hvilket kræver at den er udstyret med maskinlæring (Ibid.).

At tænke menneskeligt

Hvis en AI teknologi opfylder dette kriterium, har den til formål at efterligne menneskelige tankeprocesser. Dette involverer at skabe mentale AI-modeller, der har evnen til at navigere i verden, eftersom tilgangen fokuserer på at simulere menneskelig tænkning. Kognitiv videnskab spiller en afgørende rolle i denne sammenhæng, da det kræver at maskinen skal forstå hvordan mennesker tænker, lærer, husker og løser problemer, så den kan efterligne de kognitive processer, som mennesker anvender.

For at få en forståelse for, hvad det indebærer at en maskine kan tænke menneskeligt, kræver det en forståelse for, hvordan mennesker tænker (Russell & Norvig, 2009). Den viden kan indsamles på tre måder:

1. Introspektion, med henblik på at identificere og forstå de øjeblikke, hvor menneskelige tanker dannes.
2. Psykologiske eksperimenter, for at observere menneskelige handlinger.
3. Hjernescanninger, for at observere hjernen i aktion.

Inden for det tværfaglige felt kognitiv videnskab, kombineres computermodeller fra AI med eksperimentelle teknikker fra psykologi, for at konstruere præcise og testbare teorier om det menneskelige sind. Disse teorier kan anvendes til udviklingen af maskiner der kan tænke menneskeligt (Ibid.).

At handle rationelt

Dette kriterium sigter mod at udvikle AI, der kan udføre opgaver og træffe beslutninger ved hjælp af rationel handling for at opnå optimale resultater. Det udgør ofte grundlaget for skabelsen af *autonome agenter* i forskellige programmer. En autonom agent er enhver enhed, enten fysisk eller virtuel, der er i stand til at handle selvstændigt for at opnå specifikke mål. Disse agenter kan være robotter, computerprogrammer eller andre systemer, der kan træffe beslutninger og handle uden direkte menneskelig kontrol. De kaldes *rationelle agenter*, da de stræber efter at opnå det bedste mulige resultat, eller det

bedst forventede resultat, under usikkerhed i forhold til den opgave, de er designet til at udføre. Metoder baseret på sandsynlighedsteori og maskinlæring, har muliggjort at dette har kunnet lade sig gøre (Russell & Norvig, 2009).

At tænke rationelt

Denne tilgang tager udgangspunkt i logik, til at frembringe intelligent tænkning. Den adskiller sig fra de menneskelige handlings- og tankeprocesser, idet den udelukkende bestræber sig på at systematiserer viden og drage logiske konklusioner, uden de kognitive eller erfaringsbaserede aspekter af menneskelig adfærd. Ydermere, differentierer den sig fra rationel handling, idet fokuset i langt højere grad ligger på de interne processer og deduktioner, der kommer en handling i forkøbet, hvor den rationelle handling, som nævnt, er markant mere resultatorienteret (Russell & Norvig, 2009).

Mange AI-systemer opfylder i dag flere af de ovenfor beskrevne kriterier, og da teknologien udvikler sig med accelererende fart og de neurale netværk bliver betydeligt større med fremkomsten af deep-learning, som vil uddybes i analysen, er det nu muligt at gøre maskiner væsentligt mere intelligente end hidtil. Det kan derfor vise sig, at denne teori har vanskeligt ved at kategorisere moderne kunstig intelligens, men vi har alligevel valgt at inddrage den, for at give en basal teoretisk forståelse for, hvad kunstig intelligens kan defineres som (Ibid.).

Falske nyheder på sociale medier

I følgende afsnit vil vi redegøre for den teoretiske ramme omkring begrebet fake news, samt redegøre for fænomenet, og dets evolution i en verden med sociale medier. Dette vil vi gøre gennem to tekster. Først vil vi anvende "Fake news on Social Media: the Impact on Society" (Olan et al., 2022), da denne videnskabelige artikel, afdækker dette fænomen og giver os et dybere indblik i dets kompleksitet. Udover denne, vil vi inddrage teksten "Fake news, disinformation and misinformation in social media: a review." Denne tekst

bruger vi primært til at fremhæve de mange konsekvenser spredningen af fake news på sociale medier kan have.

Vi vil i analysen argumentere for, at deepfakes er fake news, og at deepfakes derfor kan have nogle af de samme effekter, og bruges på nogle af de samme måder, som andre typer fake news, hvilket gør det muligt at bruge definitionen som en teoretisk ramme. Derfor fandt vi det nødvendigt her i teoriafsnittet, at redegøre for fake news på sociale medier, da det vil hjælpe med at perspektivere deepfakes mulige indflydelse på folks evne til at indgå i et demokrati.

Femi Olan et al. udgav i 2022 den videnskabelige artikel nævnt ovenover, hvis formål var at afdække fænomenet fake news og dets spredning på sociale medier. *Fake news* beskrives i dette tidsskrift således: "*FN (red. Fake News), although unvetted (red. uovervåget), has a credible and professional appearance, ensuring that people cannot always distinguish it from true news*" (Olan et al., 2022). Ydermere, hævder de også at fake news særlig rettes mod særlige segmenter og grupper af mennesker, for at promovere en vilkårlig ideologi, ved at stimulere stærke overbevisninger og polarisere samfund (Ibid.). Spredningen heraf, og dermed effektiviteten af fake news, er steget enormt meget i forlængelse med tilkomsten af sociale medier (Ibid.). Fake news er som fænomen ikke nyttilkommet, og det har eksisteret siden romertiden, hvor "*the first Roman Emperor had to announce fake news to encourage Octavian to destroy the republican system*" (Ibid.). Modtagerne, omfanget og naturen af fake news har dog markant ændret karakter, i takt med udbredelsen af sociale medier, da det nu kan spredes til og fra alle.

En anden artikel, skrevet af Esmâ Aïmeur, Sabrina Amri og Gilles Brassard fremhæver også adskillige konsekvenser ved den stigende mængde fake news på sociale medier (Aïmeur, et al., 2023). De fremhæver blandt andet den enorme hastighed fake news spredes, på sociale medier. De citerer en undersøgelse, der havde til formål at afdække spredningen af fake news, og fandt frem til, at på det sociale medie X, spredes falske nyheder seks gange hurtigere end sandfærdige nyheder, hvilket kan skyldes den oftest sensationelle natur af fake news. Derudover konkludere de at 70% ikke kunne adskille disse falske nyheder fra rigtige nyheder (Ibid.). De belyser også den brede vifte af discipliner, udover samfundet generelt, der oplever konsekvenserne af fake news, herunder økonomisk videnskab, politisk videnskab, psykologien, sundhedssektoren og

klimavidenskab (Ibid.). Et nyere eksempel på konsekvenserne af fake news, der ligeledes nævnes i denne artikel, er Covid-19 pandemien. Under pandemien blev sociale medier oversvømmet med nyheder, fakta eller på anden vis ”informerende” opslag, der skabte panik og røre i utroligt mange lande (Ibid.). Det er senere blevet påvist at disse opslag ikke bundede i noget videnskabeligt, og dermed kan klassificeres som fake news, men skaden var sket. Herudover forekom der også mange opslag, der hævdede at have en ”kur” mod corona, såsom at spise kogt hvidløg eller ligefrem drikke chlordioxid (Ibid.). Disse kure blevet ligeledes senere affejet som fake news, og viste sig i nogle tilfælde at kunne være sundhedsskadelige (Ibid.). Dette understreger de alvorlige konsekvenser fake news på sociale medier kan have.

Redegørelse for forskellen mellem des- og misinformation

Semesterprojektet vil tage udgangspunkt i begrebet desinformation, der har sin oprindelse fra det russiske ord “dezinformatsiya” (Eskildsen, 2022). For at tydeliggøre semesterprojektets definering af desinformation, er det nødvendigt at redegøre for, hvad begrebet i bund og grund betyder og ligeledes definere, hvordan det står i kontrast med begrebet misinformation. Ordbogen definerer desinformation som misinformation (Ordnet.dk, 2003), dog er dette misvisende, da der er en essentiel forskel på disse begreber. Misinformation består af information, der misforstås eller formidles uden nogen dårlig hensigt. Omvendt består desinformation af bevidst falsk information, der videregives eller skabes for at gøre en form for skade (Fallis, 2014).

Ekkokamre

Vi har valgt at tage udgangspunkt i Cass R. Sunsteins definition af ekkokamre, da vi finder hans tilgang særdeles relevant for vores emne. Sunstein, en amerikansk jurist og professor, har bidraget væsentligt til forståelsen af, hvordan love og regler påvirker menneskelig adfærd. Sunstein har skrevet omfattende om beslutningstagning, informationspolitik og de sociale konsekvenser af teknologi. Han har modtaget flere

prestigefyldte priser og har haft betydelige rådgivende roller i både nationale og internationale sammenhænge (Walmsley, u.å)

Ifølge Sunstein udgør ekkokamre, som grundlæggende kan beskrives som tabet af en fælles offentlig sfære, en trussel mod det deliberative demokrati (Löblich & Venema, 2021).

Ekkokamre, som navnet antyder, er rum, hvor synspunkter gentages. Dette fænomen ses ofte på diverse sociale medier, hvor brugere bevidst eller ubevidst sorterer uønskede politiske holdninger fra. Det bevidste valg sker, når brugere tilslutter sig grupper, hvor specifikke politiske holdninger debatteres og diskuteres, mens uønskede synspunkter frasorteres. Denne form for bevidst ekkokammerdannelse, kan føre til en forstærkning af eksisterende overbevisninger, da brugerne konstant bliver bekræftet i deres synspunkter, uden at blive udfordret af alternative perspektiver (Barberá, P. 2020).

Ekkokamre kan også opstå ubevidst, hvilket ses, når filteringsalgoritmer bliver så avancerede, at brugere udelukkende eksponeres for indhold, der understøtter deres eksisterende holdninger. Algoritmerne er designet til at maksimere brugerengagement ved at levere indhold, der er i overensstemmelse med brugerens tidligere interaktioner og præferencer. Dette kan medføre, at brugerne uden at være klar over det, bliver fanget i en boble, hvor de kun ser nyheder og opdateringer, der passer til deres egen verdensopfattelse (Ibid.).

Et potentielt udfald ved dette fænomen finder sted, hvis der opstår mistillid til højere stående instanser, medier og almene borgere med polær ideologisk overbevisning. Når folk kun bliver eksponeret for ensidige synspunkter, kan det forstærke mistillid og polarisering i samfundet. Denne mistillid kan underminere tilliden til traditionelle medier og offentlige institutioner, hvilket er kritisk i et demokrati, hvor en fælles forståelse af fakta og virkelighed er nødvendig for konstruktiv dialog og beslutningstagning. Ekkokamre kan dermed ikke kun påvirke individuelle opfattelser, men også have vidtrækkende konsekvenser for den demokratiske proces og samfundets sammenhængskraft (Ibid.).

Confirmation bias

Teoretikeren bag confirmation bias, Peter Cathcart Wason, var interesseret i at skubbe til ideen om, at mennesker er rationelle og logiske tænkere. Wason var en pioner inden for kognitiv psykologi ved University College London. Han undersøgte, hvorfor mennesker ofte laver systematiske fejl i logisk tænkning og udviklede flere eksperimenter, såsom 2-4-6-eksperimentet, som er grundlaget for denne teoretiske ramme. Wason indførte begrebet *confirmation bias*, som indebærer, at individer har en tendens til at foretrække information, der bekræfter deres eksisterende meninger og tro, uanset om disse er korrekte eller ej (DBpedia, u.å).

I en videnskabelig artikel, illustreres det nu populære 2-4-6 eksperiment, som til den dag i dag stadig er et meget anerkendt eksperiment, og som senere har inspireret til en lang række andre undersøgelser (Evans, 2016).

I eksperimentet 2-4-6 får deltagerne tallene 2-4-6 oplyst, og deres opgave er at undersøge hvilken regel der gælder for disse tal. Deres første indskydelse kunne være at reglen er at tallene stiger med 2, og at talrækken dermed fortsætter 8-10-12. Det er dog her at confirmation bias opstår, da efter forsøgspersonerne havde udset sig denne regel, negligerede de andre mulige regler. Reglen kunne blot være at tallene stiger, eller at talrækken kun indeholder lige tal. Forsøgspersonerne udforskede ikke alternativer til den regel de først udpegede, hvilket videre kan oversættes til hvordan folk er tilbøjelige til ikke at undersøge alternativer til deres forudindtagede holdninger og overbevisninger (Ibid.).

Vi har valgt at inddrage netop denne teori, da confirmation bias er en psykologisk mekanisme, der får folk til at søge og fortolke information på en måde, der understøtter deres allerede eksisterende overbevisninger og perspektiver. Dette fænomen får folk til at blive blinde for alternative holdninger og forhindrer dem dermed i at være modtagelige, overfor andre synspunkter og perspektiver. Netop dette er essentielt når vi i opgaven analyserer og diskuterer, hvordan blandt andet ekkokamre kan have en effekt på demokratiet. (Ibid.).

I et demokratisk samfund er det afgørende, at borgere er i stand til at tage velinformerede beslutninger baseret på en bred vifte af perspektiver og information. Confirmation bias undergraver dette ved at indsnævre informationsfeltet og skabe polariserede grupper, der

ikke engagerer sig i konstruktiv dialog med hinanden. Dette kan føre til øget splittelse og mistillid mellem forskellige segmenter af samfundet. Ved at forstå mekanismerne bag confirmation bias kan vi bedre forstå, hvordan ekkokamre dannes og opretholdes, og dermed hvordan de kan påvirke den demokratiske proces negativt (Ibid.).

Derfor er det relevant at inddrage Wasons arbejde og teorier i vores analyse, da de giver os en dybere forståelse af de psykologiske processer, der ligger til grund for fænomenet ekkokamre, og hvordan disse kan true demokratiet ved at forstærke eksisterende overbevisninger og forhindre åben og kritisk diskussion.

Luhmann – “Theory of Trust”

Niklas Luhmann, en fremtrædende tysk sociolog og videnskabsteoretiker, var professor ved universitetet i Bielefeld, Tyskland. Han er bedst kendt for sin banebrydende systemteori, der analyserer det moderne samfund og dets delsystemer som *autopoietiske enheder*, hvilket indebærer delsystemer som er selvstændige, selvopretholdende enheder, der bidrager til samfundets funktionalitet (Dalberg-Larsen, 2009).

Luhmann præsenterede i sin bog “Trust and Power”, der oprindeligt er udgivet i 1979 og senere oversat til engelsk i 2017. I hans teori “Theory of Trust” erklærer han *tillid* som værende en mekanisme, der kan reducere kompleksiteten af sociale interaktioner. Ifølge Luhmann (2017), så kan gensidig tillid reducere behovet for at analysere adskillige udfald og andre sociale aktørers hensigt. Dette gør sociale aktører mere handlingsdygtige, og i samfund, der i stigende grad bliver mere komplekse og forbundne, og hvori mange er afhængige af hinanden på kryds og tværs, kan det være vanskeligt at navigere rundt for sociale aktører uden en basal tillid til andre sociale aktører eller strukturer, som de skal forholde sig til (Ibid.). Ydermere differentierer han også mellem to forskellige typer tillid, som er system-tillid og personlig tillid. *Personlig tillid* omhandler tilliden til andre individer, og er primært baseret på forhold mellem individer samt tidligere erfaringer (Ibid.). *System-tillid* vedrører derimod individers tillid til mere komplekse systemer, såsom demokratiske institutioner, juridiske instanser og finansielle markeder (Ibid.). System-tillid kræver ikke en dybdegående forståelse af disse komplekse systemer, men

hviler alligevel på en generel tro på at det fungerer og at det fungerer effektivt (Ibid.). Luhmanns teori har flere lag end dette, men vi vil i opgaven udelukkende forholde os til ovenstående dele af den, hvorfor vi ikke vil inkludere mere her i teoriafsnittet.

Robert Dahl – “Democracy and its critics”

I 1989 udgav Robert A. Dahl, sin bog “Democracy and its critics”, hvori han blandt andet opstiller fem kriterier for det ideale demokrati. Robert Dahl, var en fremtrædende amerikansk politolog og professor ved Yale University. Dahl er berømt for hans betydelige bidrag til opfattelsen af demokratiets pluralistiske natur (Svensson, 2009).

Teorien forholder sig til vores problemformulering, idet vi udover tillid til demokratiet, også forsøger at undersøge om borgere potentielt kan miste evnen til effektivt at indgå i et *demokrati*. Dahls kriterier for et demokrati lyder således:

1. *Effektiv deltagelse* forstås som lige og effektive muligheder for at tilkendegive sine meninger overfor andre.
2. *Lige stemmer* forstås som lige og effektive muligheder for at stemme, herunder at alle stemmer vægtes lige højt.
3. *Oplyst forståelse* forstås som lige og effektive muligheder for at lære om konsekvenserne og alternativerne til et forslag.
4. *Lige kontrol over politiske dagsorden* forstås som muligheden for at bestemme hvilke emner der skal tales om og hvordan de skal på agendaen.
5. *Inklusivt medborgerskab* forstås som at alle, eller de fleste, borgere med permanent opholdstilladelse, bør sikres alle fire ovenstående rettigheder.

Vi vil senere, i analysen, udpege to af ovenstående kriterier som værende i risiko i forbindelse med deepfake-teknologien, hvorfor de kort nu vil blive uddybet. Punkt nummer fire, den lige kontrol over den politiske dagsorden, skal sikre at alle demokratiets medlemmer skal kunne have indflydelse på, hvad der fylder i den demokratiske debat.

Dette vil sige at man skal kunne komme med politiske forslag, eksempelvis en underskriftsindsamling i Danmark, hvilket sikrer at alle borgere har en stemme, og mulighed for at være politisk engageret udover deres stemmeret (Ibid.). Det andet kriterie vi vil fremhæve, er den oplyste forståelse. Her skal der sikres et lige og åbent informationsgrundlag, hvorpå alle demokratiets medlemmer skal kunne undersøge, hvad der er bedst for dem, og dermed kunne have mulighed for at varetage egne interesser bedst muligt (Ibid.).

Metode

I følgende afsnit vil vi redegøre for de udvalgte metoder, som vi gør brug af i vores analyse af både deepfake-teknologien og den potentielle indflydelse det kunne have på demokratiet. Vi vil først gennemgå, hvordan vi har gjort brug af TRIN-modellen som rammeværktøj, og særligt hvordan og hvorfor vi har omstruktureret den, samt udeladt enkelte trin. Herefter, vil vi redegøre for vores litteratursøgning og forklare vores tilgang til den eksisterende litteratur. Til sidst, vil vi fremlægge vores metodiske fremgangsmåde i vores kvalitative undersøgelse, der har til formål at understøtte vores undersøgelse af den tidligere nævnte tillid, og/eller mistillid, til medborgere og demokratiet, som følge af deepfake-teknologi. Dette vil gøres på baggrund af Brinkmann & Tanggaard (2010).

TRIN-modellen som rammeværktøj

TRIN-modellen har i vores opgave rammesat vores tilgang til analysen af deepfake-teknologien. Modellen er udviklet af Thomas Budde Christensen, Erling Jelsø og Niels Jørgensen, og har til formål at kunne give studerende værktøjer til at lave en selvstændig analyse af en given teknologi (Jørgensen, 2018). Ifølge Niels Jørgensen (2019), er TRIN-modellen ikke en videnskabelig metode, men blot et værktøj der kan belyse de forskellige aspekter af en teknologi. Dette har også været dens formål i vores opgave, men vi ser det alligevel nødvendigt at redegøre for, hvordan vi har brugt den og ligeledes, hvorfor vi har valgt at gøre som vi har.

TRIN-modellen består af seks trin, som hver især forholder sig til forskellige aspekter af en teknologi (Jørgensen, 2018):

1. Teknologiers indre mekanismer og processer
2. Teknologiers artefakter
3. Teknologiers utilsigtede effekter
4. Teknologiske systemer
5. Modeller af teknologier
6. Teknologier som innovation

Trin 1 (teknologiers indre mekanismer og processer), har været meget centralt for vores opgave. Trinnet er defineret således: *”de centrale af de mekanismer og processer i en teknologi, som bidrager til at opfylde teknologiens funktion(er)”* (Ibid.). Semesterbindingen påkræver en dybdegående analyse af en teknologi, hvorfor netop dette trin har været så vigtigt. Vi har til analysen af deepfake-teknologiens trin 1, foretaget et litteraturstudie, da vi ikke selv havde ressourcer til at konstruere en deepfake fra bunden, grundet teknologiens omfattende natur. Vi har derfor skulle basere vores analyse af teknologien på den eksisterende litteratur og vi vil senere i metode-afsnittet uddybe, hvordan vi helt konkret har tilgået dette.

Trin 2 (teknologiers artefakter) forholder sig til de fysiske genstande, tjenester eller viden der forekommer eller skabes i forbindelse med en teknologi (Ibid.). Dette trin har vi valgt ikke at gå i dybden med i forbindelse med vores projekt, da vi er mere interesserede i teknologiens indre mekanismer, samt dens påvirkning på et samfund på et større plan. Vi vil dog hurtigt gennemgå, hvor meget computerkapacitet en deepfake, og produktionen heraf kræver. Dette står dog ikke lige så centralt som førstnævnte trin, da vi som sagt ikke er overvejende interesserede i de fysiske aspekter af deepfake-teknologien.

Trin 3 (teknologiers utilsigtede effekter) er modsat trin 2, en meget vital del af vores opgave. Dette trin omhandler teknologiens funktioner og formål, men også de utilsigtede effekter, hvilket altså er effekter, der forekommer grundet teknologien og som vurderes som værende negative (Ibid.). Vi forholder os ikke slavisk til trin-modellen, og har valgt at omstrukturere den, så vi indledningsvist i analysen vil gennemgå de mere teknologiske

aspekter, for så at følge op med et afsnit af teknologiens mulige påvirkning på menneskers tillid til hinanden og demokratiet. Dette sidste afsnit er i høj grad inspireret af trin 3, men vi vælger ikke at kalde det netop utilsigtede effekter, da vi har indset at misbrug og følgerne af dette, ofte kan være den tilsigtede effekt. Sammenhængen mellem vores afsnit, og trin 3, er dog at det handler om de følger der kommer af en teknologi. Vi vil derudover argumentere for at risikoen for misbrug til dels er en utilsigtet effekt ved deepfakes, og dermed bliver koblingen mellem trin 3 og vores afsnit endnu engang cementeret.

Trin 4 (teknologiske systemer) indgår ligeledes i vores analyse, men er dog en del af den mere TSA-bundne analyse. Dette trin har til formål at rammesætte de forskellige systemer, der indgår i en teknologi og belyse samspillet mellem dem (Ibid.). Deepfakes, som vi vil illustrere i analysen, består af en række teknologiske komponenter og systemer, der konstant arbejder sammen for at levere så troværdige deepfakes som mulige. Dette trin har vi koblet sammen med særligt trin 1 og trin 2, for at skabe en samlet og dynamisk analyse af deepfake-teknologien. Ydermere, vil vi også klargøre, hvilke systemer deepfake-teknologien indgår i, før den kan have indflydelse på borgeres tillid til hinanden og de demokratiske institutioner.

Trin 5 (modeller af teknologier) har til formål at fremhæve særlige egenskaber ved en teknologi (Ibid.). Dette trin opfordrer til at skabe en *"... abstrakt, visuel eller fysisk repræsentation af et fænomen eller en genstand, hvor særligt udvalgte egenskaber ved fænomenet eller genstanden søges gengivet og/eller undersøgt"* (Ibid.). Denne tilgang har vi også inkluderet i vores analyse, og den er særligt nyttig til at kunne visualisere og overskueliggøre vores redegørelse af en kompleks teknologi. Den er ikke, på samme måde som vores tilgang til de andre trin, begrænset til én del af opgaven, men vi anvender i stedet denne tilgang flere steder i opgaven, for netop at kunne fremhæve og visualiserer de forskellige aspekter af både deepfake-teknologien, men også dens mulige indvirkning på menneskers tillid til demokratiet og hinanden.

Modellens sidste trin, trin 6 (teknologier som innovation), vil ikke fylde så meget i opgaven. Dette trin berører spredningen og tilgængeligheden af en teknologi, og hvordan dette ændrer sig over tid. Denne tilgang vil også kunne være enormt interessant at fokusere mere på, men vi har valgt af afgrænse os til kunstig intelligens på nuværende

tidspunkt, hvorfor vi under vores problemstilling ikke finder det lige så interessant at forsøge at forudsige kommende tendenser og/eller konsekvenser af deepfakes. Vi vil kort komme ind på det dels i analysen, men også i diskussionen, men vil ikke gå mere i dybden med det end det.

Studie af den eksisterende litteratur

Vi har valgt at gennemføre et litteraturstudie som vores primære metode, til vores teknologiske analyse. Vi har systematisk indsamlet relevant litteratur og organiseret vores resultater i et skema. Indledningsvis opnåede vi en generel viden om begrebet kunstig intelligens, gennem kurset Teknologiske Systemer og Artefakter, samt ved hjælp af litteratursøgning på hjemmesider og Det Kongelige Bibliotek. Efterfølgende har vi uddybet vores litteratursøgning og identificeret relevante bøger og videnskabelige artikler. Dette har givet os et dybere kendskab til de teknologier, der ligger til grund for skabelsen af deepfakes.

Denne omfattende søgning har gjort os i stand til at opstille de aspekter af teknologien bag deepfakes, som vi skal kunne forstå og formidle i vores opgave. Da vi havde klarlagt, hvilke emner der er relevante for vores projekt, oprettede vi en tabel, hvor disse emner blev placeret i den øverste række. Efterfølgende har vi gennemgået yderligere peer-reviewed litteratur og markeret felterne, når litteraturen dækkede de ønskede aspekter af deepfake-teknologien.

Efter en nøje gennemgang af al litteratur og med alle hovedemner dækket, konkluderede vi, at vi havde tilstrækkelig litteratur til at kunne udføre en fyldestgørende analyse. Al den omtalte litteratur vil blive præsenteret i tabel 1 på næste side.

I tabel 1 nedenfor har vi defineret de udvalgte emner og teknologier vandret i den øverste række. Øverst i venstre hjørne har vi kategoriseret vores kilder som artikler (A), bøger (B) og undervisningsmateriale (U). I den lodrette kolonne til venstre har vi listet den udvalgte litteratur og angivet deres respektive kategorier.

Inde i skemaet har vi markeret de emner, som hver enkelt kilde dækker. Vi har, som vist i tabellen, flere overlap af emner, hvilket er en bevidst beslutning der sikrer en kildekritisk

tilgang, hvor vi har kunne krydsreferere litteraturen. Vi har dog alligevel valgt at visualisere hvor vi primært har fået information fra, og hvordan vi har afgrænset litteraturen. Dette er gjort med en variation i afkrydsningen, hvor "V" indikerer de emner indenfor den givne litteratur, som vi hovedsageligt har valgt at inddrage i rapporten, mens "X" indikerer at teksten også til dels dækker emnet, men at vi ikke har benyttet det lige så ekstensivt. Fælles for de to er dog, at de har indgået i en generel indsamling af viden, der har dannet et mere omfattende informationsgrundlag, og dermed et nødvendigt overblik over deepfake-teknologien.

Artikler = (A) Bøger = (B) Undervisning = (U)	Neurale Netværk	CNN	GAN	HPC	Auto encoders	NLP	Maskinlæring	AI
(U)-Vigtige begreber om maskinlæring	X	X					V	
(U) - Kunstige neurale netværk og deep learning – en ultrakort introduktion	V	V						
(A)-Generative adversarial network (GAN)	X	X	V				X	X
(B) - A Primer on Generative Adversarial Networks	X	V	V		X	X	X	X
(B)(U) - Artificial Intelligence: A Modern Approach	X	X					V	V
(B) - Introduction to HPC with MPI for Data Science				V			X	
(B) - Computing Power and the Governance of AI				V			X	X
(A) - What is high-performance computing (HPC)?				V				X
(A) - What is an autoencoder?	X	X			V			X
(A) - Floating-point operations per second(FLOPS).				V				

(B) - <i>Advances in Computers, Volume 11</i> (pp. 353-355)	X	X			V		X	X
(B) - <i>Internet of Multimedia Things (IoMT)</i> (pp. 144-145)	X	X			V			X
(A) - <i>What is NLP?</i>						V	X	X
(A) - <i>NLP vs. NLU vs. NLG</i>						V	X	X

Tabel 1, skema over litteraturvalg. Al litteraturen fremgår i litteraturlisten.

Vi har benyttet artikler fra både hjemmesider og Det Kongelige Bibliotek. Hjemmesiderne har fungeret som introduktion til vores litteratursøgning, hvilket har givet os en grundlæggende forståelse af den teknologi, der ligger til grund for kunstig intelligens og deepfake-teknologien. Vi har primært udvalgt artikler fra veletablerede IT-virksomheder som Techtarget, Google og IBM. Disse kilder er førende inden for feltet, og vi har derfor vurderet dem som troværdige (Jungco, 2024). Efterfølgende har vi benyttet peer-reviewed tidsskrifter, hvor vi har udvalgt de videnskabelige artikler, som vi har fundet relevante for vores projekt.

Med udgangspunkt i den viden vi har tilegnet os gennem artiklerne, har vi udvalgt specifikke emner inden for teknologien, som vi har søgt dybdegående litteratur om. Vi har gennemgået udvalgte afsnit fra diverse bøger om emnerne High Performance Computing (HPC), Autoencoders, Generative Adversarial Networks (GANs), Computer Science m.m. Bøgerne er skrevet af professorer og forskere inden for feltet og giver en dybdegående forståelse af teknologien, hvilket har været fundamentalt for at kunne udarbejde vores analyse. Bøgerne er udgivet af anerkendte forlag som Springer Inc. (Springer, 2023), hvilket øger kildernes troværdighed.

I kurset Teknologiske Systemer og Artefakter har vi modtaget undervisning om kunstig intelligens (AI), maskinlæring, neurale netværk, deep-learning og Convolutional Neural Networks (CNN). Det har derfor været oplagt at benytte undervisningsmaterialet til dette projekt. Dele af materialet er udarbejdet af professor Henning Christiansen fra Roskilde Universitet, og han har desuden introduceret os for Russell & Norvigs bog om kunstig

intelligens. Vi har valgt at anvende denne bog, da Russell & Norvig har bidraget betydeligt til AI-forskningen.

Interviews af Tanggaard og Brinkmann

Vi har i vores kvalitative undersøgelse, i form af interviews, taget udgangspunkt i Tanggaard og Brinkmann (2010). I denne har vi fundet vores fremgangsmåde, samt hvordan vi efterbehandler vores interviews.

I dette projekt valgte vi at fokusere på kvalitative interviews, som yderligere indsamling af empiri for projektet. Dette fungerede som en understøttelse af vores egen selvstændige analyse, i koblingen mellem deepfake-teknologien og demokrati. Ved at være i direkte dialog med informanterne, fik vi en bredere forståelse for vores problemfelt. Dette var ydermere for at kunne inkludere andre perspektiver end vores egen, da vi havde en forudgående tolkning og antagelse af problemet, samt årsagerne bag.

Tanggaard og Brinkmann præsenterer tre forskellige interviewstrukturer; det løst struktureret interview, det semi-struktureret interview og det stramt struktureret interview (Tanggaard & Brinkmann, 2010).

Det løst struktureret interview, bruges primært i feltarbejde og er som oftest en uformel samtale. Den er som regel vanskelig og kræver stor træning fra interviewers side. Det løst struktureret interview kan være en fordel, da den tilegner sig så meget viden om interviewpersonen som muligt, i det interviewene bliver gennemført, når det passer interviewpersonen og kun når interviewpersonen har noget at fortælle (Ibid.). Dette blev fravalgt i vores projekt, grundet at vi gerne ville være sikre på at komme ind på specifikke emner under interviewene.

Modsat det løst struktureret interview, er det stramt struktureret interview styret af spørgsmål fra interviewer. De spørgsmål er sat sådan op, at det kan være svært at give et uddybende svar, da de kan minde om et spørgeskema, hvor interviewpersonen eksempelvis skal vurdere tilfredsheden ud fra en skala eller svare på ja/nej spørgsmål. Det stramt struktureret interview kan være en fordel, hvis man nemt vil kunne kategorisere og kvantificere interviewpersonernes svar (Ibid.). Årsagen til at det stramt

struktureret interview er blevet fravalgt i vores projekt, er på grund af *pleaser-effekten*, der betegner en situation, hvor interviewpersonen forsøger at svare, hvad de tror forskeren gerne vil høre (Ibid.). Herudover krævede vores problemformulering også en mere dybdegående og fri samtale om emnet, for at afdække facetter af problemet, vi ikke selv havde overvejet.

Den sidstnævnte interviewmetode er det semi-struktureret interview, som vi har valgt at benytte i dette projekt, da vi på forhånd vidste, hvad vi ville søge viden om. Vi ville dog samtidigt forholde os åbent overfor de perspektiver, som informanterne kunne komme med (Ibid.).

På linje med Tanggaard og Brinkmanns metodiske gennemgang af det semi-struktureret interview, udarbejdede vi en interviewguide, med det formål at skabe en struktur over vores fem interviews, med mulighed for fordybelse ved hvert spørgsmål. Vores interviewguide præsenteres i analysen.

I det semi-struktureret interview må der godt afviges fra interviewguiden, hvis forskeren finder det relevant. Dette kan gøre at forskeren skal være parat til at improvisere spørgsmål, ud fra hvilke svar de pågældende interviewpersoner giver. Interviewguiden kan herefter også hjælpe med at lede interviewet tilbage på sporet, i tilfælde af at man afviger meget fra forskningsspørgsmålene (Ibid.).

Anvendt interviewguide

Vi havde til vores interviews, som nævnt, udarbejdet en interviewguide. Denne guide fungerede som en løs retningslinje for vores interviews. Indledningsvist bad vi om en basal introduktion af vores interviewpersoner, herunder alder og uddannelsesbaggrund. Dette inkluderede også spørgsmål om hvorvidt de var på sociale medier, og videre om de fulgte folkevalgte politikeres profiler herpå. Dette skulle fungere som en blød indledning, der kunne sikre en organisk og afslappet dialog om vores problemfelt.

Vi afspillede herefter interviewpersonerne et lydclip fra en deepfake-video. Efter dette, spurgte vi ind til deres umiddelbare tanker om det. Herefter blev interviewet opdelt efter hvorvidt interviewpersonen havde kendskab til videoen på forhånd eller ej. Her opdelte vi

interviewguiden i to, hvoraf den ene del skulle rettes mod dem der havde kendskab til videoen, og den anden del skulle rettes mod dem der ikke havde. Herudover, kunne vi også undersøge om folk kunne gennemskue om det var en deepfake eller ej, og derimod om teknologien har nået et niveau, hvor den kan efterligne virkeligheden nok til at folk har svært ved at skelne mellem de to.

Havde ikke kendskab til lydklippet

Kendte interviewpersonen ikke til videoen, forklarede vi oprindelsen af videoen, herunder at det var en deepfake. Dette blev fulgt op med at gentage spørgsmålet om deres tanker om videoen, med konteksten in mente.

Dette blev fulgt op med følgende spørgsmål:

1. *"Hvordan har du det med at en politiker har delt en fake video af en opponent?"*
2. *"Giver det dig mindre tillid til demokratiet, idet det kan komme til at ske oftere?"*
3. (HVIS NEJ) *"Hvis det ikke var tydeligt at det var en politiker der havde lavet den, men at politikerens i stedet delte videoen gennem uofficielle kanaler, hvad så?"*

Disse spørgsmål skulle hjælpe med at afdække hvorvidt deepfake-videoer potentielt ville kun influere vores interviewpersoners tillid til det demokratiske system.

Havde kendskab til lydklippet

Havde interviewpersonerne forudgående kendskab til lydklippet, var de højst sandsynligt klar over at det var en deepfake. Dette gjorde det dog ikke mindre relevant at høre dem om deres mening om videoen i en demokratisk kontekst. Vi rettede derfor de samme spørgsmål mod denne gruppe af interviewpersoner, dog med et yderligere spørgsmål der lød: *"Hvordan har du det med at der er andre der er faldet for denne video?"*. Dette fandt vi særligt relevant, da det forholder sig til spørgsmålet om hvorvidt folk er trygge ved at deres medborgere i et demokrati, potentielt ikke kan skelne mellem deepfake og virkelighed, og dermed basere deres holdninger på fiktive scenarier.

Analyse

Vi vil i følgende afsnit analysere deepfake-teknologien isoleret set, ved at kigge på dens individuelle mekanismer og processer, og samtidig analysere på, hvorvidt den potentielt vil kunne influere borgere i et demokrati. Dette gøres ved en firedelt analyse, hvor vi indledningsvis vil gennemgå de mange teknologier der tilsammen udgør en deepfake. Dette afsnit kategoriserer vi som "teknologisk analyse". Herefter følger vi op med "systemer", der har til formål at undersøge hvilke foranstaltninger der skal være, hvis en deepfake skal influere demokratiets borgere. Derefter vil vi analysere vores kvalitative undersøgelser, og afkode resultaterne af vores interviews. Afslutningsvist stiller vi deepfake-teknologien, i kontekst af et moderne samfund, op mod Dahl og Luhmanns teorier om demokrati og tillid.

Teknologisk analyse

Deepfake-teknologien er en relativ nyudviklet teknologi fra 2017, som nu gradvist bliver en integreret del af vores digitale virkelighed (Peters, 2021). Allerede i dag findes der indhold skabt med denne teknologi i form af videoer, stemmer og billedgenereret indhold, som flourer rundt på diverse sociale medier (Ibid.). Teknologien anvendes til en række forskellige formål, herunder, men ikke udelukkende, underholdning (Diakopoulos & Johnson, 2020) filmproduktion (Schomer, 2023) og reklameproduktion (AdCreative.ai, u.å).

Deepfakes er ment til at skabe realistiske digitale gengivelser af data, herunder video, billeder og lyd, der virker så ægte, at de kan forveksles med virkeligheden. Teknologien anvender avancerede metoder inden for kunstig intelligens og *maskinlæring*. Begrebet maskinlæring er kernen i deepfake-teknologien, som involverer algoritmer, der lærer fra store mængder data. Disse algoritmer, som kaldes neurale netværk, forbedrer sin evne til at udføre opgaver ved gentagne gange at analysere data og lære af fejl. Til forskel fra traditionelle computerprogrammer, der følger forudbestemte instruktioner, tilpasser

maskinlæringsmodeller sig selvstændigt for at opnå større nøjagtighed og effektivitet. Programmerne programmeres ikke med faste regler, og dette gør dem i stand til at udføre opgaver, hvor traditionelle metoder fejler. Ved at genkende mønstre og træffe selvstændige beslutninger, bliver disse programmer bedre til at udføre de opgaver, de er designet til (Christiansen, 2024). En underkategori af maskinlæring, ved navn *Deep Learning*, er særligt effektiv til at generere eller manipulere indhold, så det fremstår som ægte (Barney, 2023).

Deep Learning involverer brugen af store neurale netværk med mange lag, til at træne komplekse mønstre i store datamængder (Christiansen, 2024). Disse neurale netværk er bygget op af neuroner, som trænes på store mængder data til blandt andet at genkende mønstre, lave klassificeringer, eller endda generere indhold, der ikke tidligere har eksisteret, hvilket kan benyttes til at skabe såkaldte deepfakes.

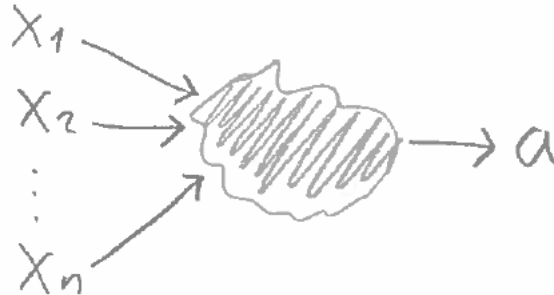
Teknologier som Deep Learning, Neurale netværk, Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), Autoencoders, Natural Language Processing (NLP) og High-performance Computing (HPC) spiller en væsentlig rolle i udviklingen og implementeringen af deepfake-teknologier (Barney, 2023).

For at opnå en forståelse af deepfake-teknologien er hver af disse teknologier blevet analyseret enkeltvis, for at belyse de enkelte dele, der tilsammen muliggør produktionen af deepfakes. Vi vil starte med at analysere den fra det mindste perspektiv og af den grund begynder vi med en dybdegående analyse af neurale netværk, som er kernen i deepfake-teknologiens funktioner.

Neurale netværk

Neurale netværk er en sofistikeret model, der simulerer den menneskelige hjernes funktionalitet og komplekse forbindelser mellem neuroner. Disse modeller anvender kunstige neuroner som grundelementer (Christiansen, 2024).

I ethvert netværk, udfører hvert eneste neuron en elementær operation, der involverer et sæt inputværdier, betegnet matematisk som x_1, x_2, \dots, x_n . Denne operation resulterer i en outputværdi, ofte angivet som a :



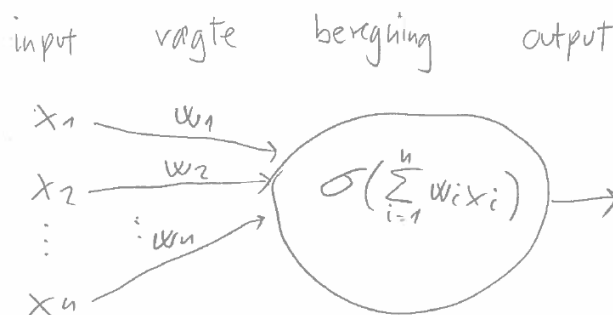
Figur 6, visualisering af et neuron, Christiansen, 2024.

Den præcise funktion, der udføres af et neuron, bestemmes af et antal parametre, kaldet for vægte. Vægtenes matematiske formel repræsenteres som w_1, w_2, \dots, w_n . Neuronets kerneopgave er at beregne en vægtet sum af sine inputværdier:

$$w_1x_1 + w_2x_2 + \dots + w_nx_n = \sum_{i=1}^n w_i x_i$$

Figur 7, matematisk formel for vægtet sum af inputværdi, Christiansen, 2024.

Summen kan være et hvilket som helst tal - stort, lille, negativt eller positivt. For at regulere og stabiliserer den vægtede sum, anvendes en aktiveringsfunktion, som typisk er *sigmoid-funktionen*. Denne funktion stabiliserer outputtet, og begrænser dermed outputtet til at ligge i et bestemt interval, typisk mellem 0 og 1. Hvis inputtet (x) er kraftigt negativt, vil outputtet nærme sig 0 og omvendt vil den nærme sig 1, hvis x er højt positivt. Mindre justeringer i vægtene (w), vil resultere i tilsvarende ændringer i outputtet, hvilket tillader finjusteringer af netværkets adfærd:



Figur 8, illustration af en neurons indre processer, Christiansen, 2024.

Et enkelt neurons evne til at repræsentere komplekse funktioner er begrænset, og derfor består et neuralt netværk af adskillige forbundne neuroner, organiseret i et hierarkisk mønster. Denne struktur indeholder talrige input- og outputneuroner, og når netværket modtager input, føres de gennem lag af neuroner, betegnet som n_1 , n_2 , ..., n_w , hvorved de transformeres til netværkets endelige output (Ibid.).

Netværkets kapacitet til nøjagtigt at frembringe det ønskede output gennem træning, er central for dets funktion. Under træningsprocessen justeres netværkets vægte for at optimere outputtet. Dette opnås oftest med *supervised-learning*, hvor netværket trænes med et datasæt bestående af kendte input - output par (Ibid.). Særligt inden for Convolutional Neural Networks (CNN'er) anvendes denne metode (Ibid.). I konteksten af deepfake-teknologien benyttes metoden *unsupervised-learning*, da data-træningssættet er så stort, at man ikke kan putte en kategori på alle elementer manuelt, og at der er flere ukendte kategorier i datasættet, hvilket betyder, at *unsupervised learning* er den mest optimale træningsform til denne opgave (Kaddoura, 2022). Begreberne *supervised* og *unsupervised*, vil blive uddybet lidt nærmere, længere nede i analysen.

Deep Learning

Samtlige af de ovenstående teknologier er baseret på *Deep Learning*. Deep Learning refererer til neurale netværk med mange lag af neuroner, som er særligt effektive til

opgaver inden for billedbehandling. Disse dybe netværk kan inkludere millioner af neuroner og milliarder af forbindelsesvægte, hvilket skaber meget omfattende modeller. Disse netværk har kapacitet til at håndtere og lære fra store datamængder, hvilket har ført til markante fremskridt inden for feltet. (Christiansen, 2024)

Betegnelsen "deep" i "Deep Learning" refererer til antallet af lag i netværket. I Convolutional Neural Networks (CNN'er) inkluderer disse lag typisk konvolutionelle lag, som er specialiserede til at behandle rumlige data som billeder, samt andre typer af lag, der alle er forbundne og bidrager til at identificere komplekse mønstre og træk i data (Ibid.).

CNN

Convolutional Neural Networks (CNN) er en væsentlig teknologi i GAN-opbygningen og er en type arkitektur inden for Deep Learning, der er særligt velegnet til behandling af billeddata. CNN er en diskriminativ model, som er specialiseret i at analysere og klassificere data gennem konvolutionelle lag, som scanner billeder for at identificere mønstre og karakteristika ved hjælp af en samling af filtre (Kaddoura, 2022). Disse filtre skanner små, lokale områder af billedet for at identificere specifikke træk ved hjælp af konvolutionsprocesser. Hvert filter genererer en såkaldt *feature map*, der fremhæver visse egenskaber i billedet, som eksempelvis kanter, teksturer eller farver (Christiansen, 2024).

Når et CNN er trænet, vil hvert filter specialisere sig til at genkende bestemte træk i de pågældende data. For eksempel kan et filter blive trænet til at reagere på diagonale linjer, mens et andet kan blive trænet til at identificere områder med en bestemt farve. De forskellige feature maps, som netværket producerer, bliver derefter kombineret i efterfølgende lag for at identificere mere komplekse mønstre, indtil netværket kan klassificere hele billeder, som f.eks. at skelne mellem billeder af forskellige objekter eller motiver (Ibid.).

Ved at justere vægte og skævheder for hver pixel, klassificerer CNN-modellen billedet og tildeler en kategorisering, som indikerer, om billedet forestiller en hund eller en kat.

I situationer, hvor opgaven kræver omvendt funktionalitet, nemlig at konstruere billeder på baggrund af specifikke kategorietiketter, træder generative modeller i forgrunden, som generatoren i en GAN (Kaddoura, 2022).

GAN

GAN står for Generative Adversarial Network og er en klasse af maskinlæringsystemer, hvor to neurale netværk konkurrerer i et spil, med målet om at skabe en høj kvalitet af genereret data. De to modeller består som nævnt af en diskriminator og en generator. *Generatoren* har til formål at skabe data, der efterligner de ægte data så tæt som muligt, mens *diskriminatoren* arbejder på at skelne de genererede data fra de ægte. Dette skaber et feedback loop, hvor generatoren løbende forbedres for at gøre de genererede data mere overbevisende. Derved genereres der ny data ud fra det input som generatoren trænes med (Kaddoura, 2022). Denne type model trænes oftest unsupervised, hvilket som nævnt, er en metode til at træne neurale netværk, der foregår uden brug af foruddefinerede eller mærkede data, hvor hvert stykke inputdata er tilknyttet et specifikt output eller et label, som er kendt på forhånd (Yasar, 2023).

Neurale netværk kan trænes på to måder: enten supervised eller unsupervised. Disse metoder adskiller sig ved, at unsupervised learning kun præsenterer maskinen for input, uden at kende det korrekte output på forhånd. Dette gør maskinen i stand til at udforske mønstre i dataen og identificere skjulte strukturer, uden at have et specifikt mål i sigte, for det ønskede output. Ved supervised learning, kender maskinen derimod både til input og outputparret, hvilket betyder at maskinen som trænes, kan lære via sammenhænge mellem inputtet og outputtet, hvilket gør den i stand til at generere regler baseret på det kendte output (Christiansen, 2024).

I denne kontekst, kan input og output meget forsimplet forstås som, at input er den data du giver din maskine til at bearbejde, og output er det data, maskinen spytter ud igen.

Diskriminator

Diskriminatoren fungerer som en klassifikator inden for Generative Adversarial Networks (GANs), og dens opgave er at skelne mellem autentisk data og data produceret af generatoren. Typisk anvendes neurale netværk til at opbygge diskriminatoren, med Convolutional Neural Networks (CNN) som den foretrukne arkitektur for billedrelaterede opgaver (Kaddoura, 2022).

Under træningsprocessen modtager diskriminatoren både data genereret af generatoren og ægte datasæt. Den evaluerer derefter denne data og kategoriserer dem som enten autentiske (markeret med 1) eller genererede (markeret med 0). Til at måle effektiviteten af denne klassificering, anvendes ofte binary cross-entropy, hvilket er en funktion som nøjagtigt udregner forskellen mellem diskriminatorens vurderinger og den faktiske kategorisering af data som "falsk" eller "ægte" (Ibid.).

Træningen indebærer en iterativ proces, hvor generatorens vægte midlertidigt fastholdes, mens diskriminatoren forbedres, indtil generatoren ikke længere kan 'snyde' diskriminatoren. Dernæst fastholdes diskriminatorens vægte, mens generatoren finjusteres for bedre at efterligne ægte data.

Diskriminatorens kritiske rolle i udviklingen af deepfakes er at finjustere generatorens kapacitet til at producere videodata, der er overbevisende nok til at undgå detektion af diskriminatorens algoritme. Det er afgørende, at diskriminatoren trænes med relevante og varierende data for at sikre, at den stiller en høj standard for generatorens output.

$$\text{BCE} = \frac{1}{n} \sum_{i=1}^n y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i))$$

Figur 2, Eksempel på en diskriminators matematiske formel, Kaddoura, 2022.

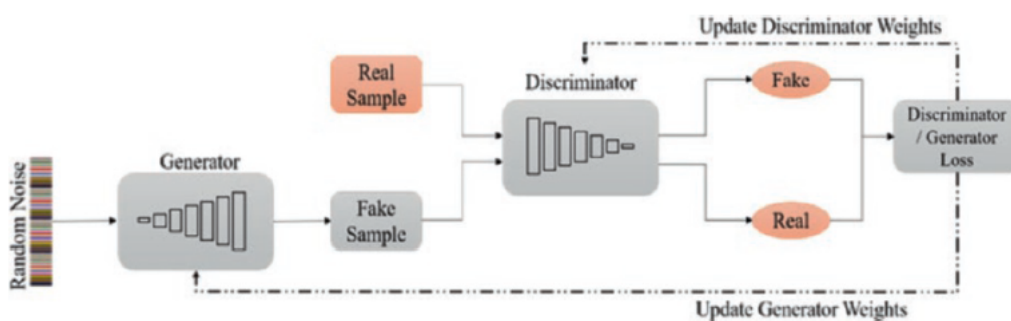
Generator

Generatoren i et Generative Adversarial Network har, som nævnt, til opgave at skabe data, der imiterer ægte data så overbevisende, at det bliver svært at skelne mellem det ægte og det genererede.

Dette opnås gennem en nøje træningsproces, hvor generatoren justerer sine vægte for bedre at efterligne de reelle data. Denne tilpasning sker ved hjælp af *backpropagation*, en central mekanisme i neurale netværk, der justerer netværkets vægte baseret på forskellen mellem netværkets output og de faktiske data. Backpropagation fungerer ved at føre data baglæns gennem netværkets neuroner, fra outputlaget til inputlaget, hvilket gør det muligt at identificere og rette de vægte, der bidrager mest til eventuelle fejl i output. Denne proces af kontinuerlig analyse og justering, gør det muligt for generatoren gradvist at forbedre sin evne til at generere troværdigt og realistisk indhold (Christiansen, 2024).

Generatoren kan tilpasse sig, og lære fra en bred vifte af inputdata, herunder tekst, billeder og videoer, hvilket gør GANs ekstremt alsidig i forhold til at generere forskellige typer af realistisk udseende data, tilpasset specifikke anvendelser eller behov.

Dette er en visualisering af, hvordan en GAN trænes:



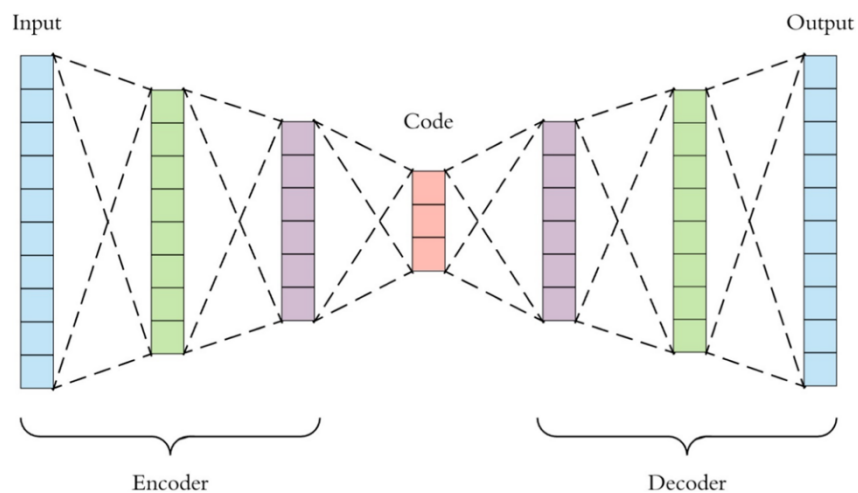
Figur 3, illustrattion af GAN, Kaddoura, 2022

Vægtene i en GANs model justeres baseret på feedback fra diskriminatoren. Når diskriminatoren identificerer falsk input som ægte, tilpasses dens egne vægte for at forfine dens skelneevne. Omvendt, hvis diskriminatoren vurderer genereret data som falsk, opdateres generatorens vægte for at forbedre dens evne til at skabe mere overbevisende falske data som illustreret i figur 3. Denne proces af gentagen justering og tilpasning fortsætter indtil et *Nash Equilibrium* opnås, hvilket er tilstanden, hvor ingen af delene, generatoren eller diskriminatoren, kan forbedre deres præstation yderligere ved træning. I et Nash Equilibrium har generatoren nået et punkt, hvor den producerer data

så realistiske, at diskriminatoren ikke længere effektivt kan skelne mellem ægte og genereret indhold (Kaddoura, 2022).

Autoencoders

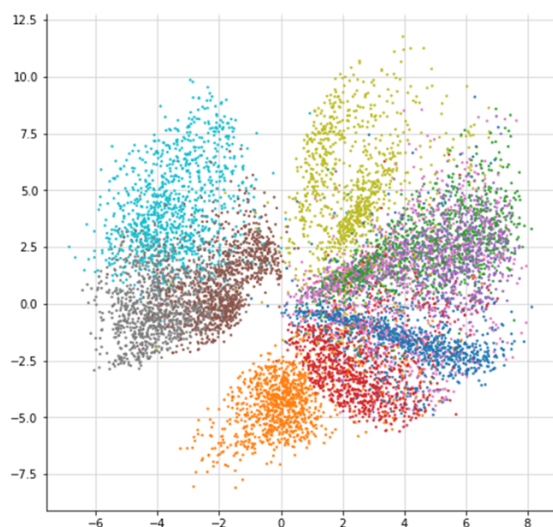
En *autoencoder* er, ligesom CNN, et neuralt netværk. Autoencoders omdanner et input via en *encoder* til et stykke kode, der indeholder inputtets mest fundamentale træk, ved at fjerne dele af den originale data fra inputtet, for derefter at rekonstruere det igen ved hjælp af en *de-coder* (Abirami & Chitra, 2020). Denne proces kan i en forsimplet forstand sammenlignes med, at man tager et billede af et hus for derefter at bygge huset igen ud fra den info du kan få ud af billedet.



Figur 4, Illustration af Autoencoders struktur, Ou, T. Medium (2022).

Som set i figur 4 er strukturen for en autoencoder bestående af et input, en encoder, flere lag af kompression, et stykke kode, en decoder og et output. Dette opnås ved hjælp af lineær og ikke-lineær transformation (Janeczko & Srivastava, 2022) og er opbygget via CNN'er. Via denne proces vil autoencoderen opdage skjulte variabler, der gemmer sig i datastrukturen som den kan bruge til at lave forudsigelser, samt genkende fællestræk i dataene. Distributionen af denne "skjulte" data kaldes for *latent space* hvilket på dansk kan oversættes til latent rum, hvilket er en betegnelse for den samlede mængde skjult data. Distributionen af denne skjulte data kaldes for *latent space*, hvilket på dansk kan oversættes til latent rum, som er en betegnelse for den samlede mængde skjult data. Denne metode er som beskrevet effektiv til at finde mønstre i data, som mennesker ikke

kan se, og kan bl.a. illustreres via et koordinatsystem som ved figur 5:



Figur 5, Latent space distribution af givent datasæt, , Ou, T. Medium,2022 Denne egenskab er essentiel når man bl.a. skal lave deepfakes, da det er denne proces der gør det muligt for en given model at genskabe en persons karakteristika så som et ansigt, en krop, en lokation, mm. Når der laves deepfakes kan der gøres brug af to autoencoders, der via den ovennævnte proces henholdsvis træner en model til at rekreere en kendt persons ansigt, og en anden model til at lave ansigts- og bevægelsesmønstre. Disse to kan tilsammen afbillede eksempelvis en kendt person, der foretager sig en handling der fremstår meget realistisk.

For at der er tale om en decideret autoencoder, er det væsentligt at kigge på om netværket er trænet unsupervised, da der også findes encoder-decoder netværk, som ikke er autoencoders, i tilfælde af, at de er trænet supervised (Bergmann & Stryker, 2023). Det som gør et encoder-decoder system til en autoencoder, er at det per definition automatisk, selv kan finde og kategorisere det data og input som det trænes på.

Natural Language Processing (NLP)

Natural Language Processing er en betegnelse, der bruges indenfor computer science, der som navnet antyder, referer til processeringen af naturligt sprog. Dette henvises i opgaven til dets forkortelse, NLP (Holdsworth, 2024).

NLP er en integreret del af processen, når der skal produceres en deepfake-video, da det er via NLP og nærmere herunder *NLG* (Natural Language Generation) at der skabes menneskelige stemmer til disse videoer (Kavlakoglu, 2020).

Denne proces foregår ved at en given AI-model modtager noget tekst som input fra en bruger, som den derefter via NLP kan strukturere via *tokenization*, hvilket er en metode, hvorpå computeren nedbryder et ord på for at analysere det i dets mest enkelte forstand, for derefter at generere syntetiske stemmer via NLG. For nærmere gennemgang af tokenization henvises der til bilag 7.

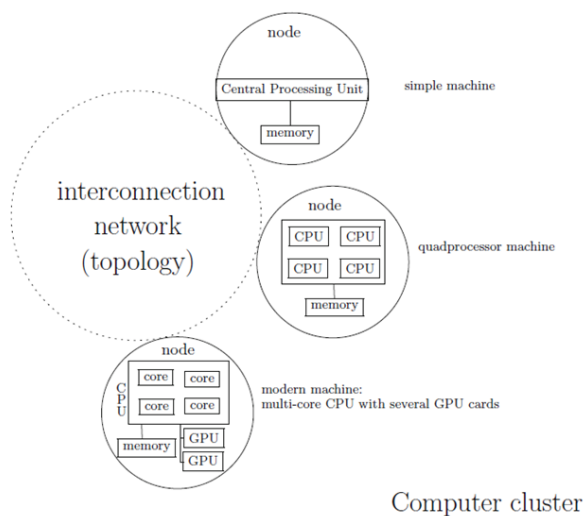
Uden NLP kan man hverken instruere en model i at kreere en deepfake ud fra tekst eller få tilføjet lyd til videoen. NLP-teknologien gør det både nemmere tilgængeligt at lave en deepfake video, samtidig med, at den kan få en given video til at fremstå mere troværdig ved at tilføje lyd til videoen der passer med karakterernes bevægelsestræk.

High-performance computing (HPC)

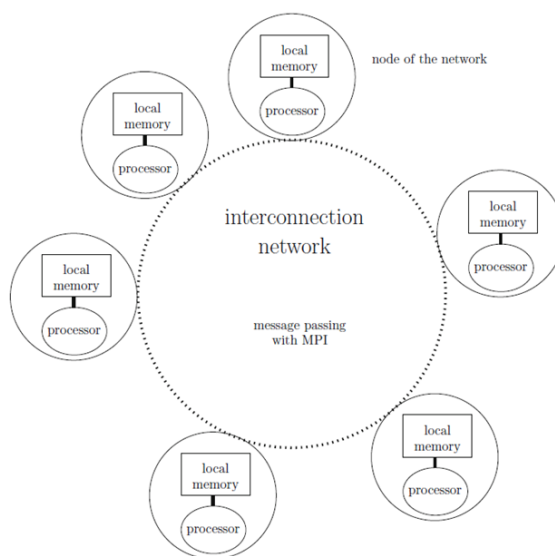
High-performance computing er relevant for deepfake teknologien, da det danner rammerne for, hvilket udstyr og hvilken computerkraft, der skal til for at kunne træne de modeller, der ligger til grund for teknologien. High-performance computing er generelt en af de mest fundamentale teknologier, der har muliggjort at udvikle store AI modeller (Heim et al., 2024). Vi vil i dette afsnit referere til high-performance computing som HPC.

HPC er i simple termer et udtryk for et system med en meget stor mængde computerkraft, der arbejder sammen for at kunne gennemføre opgaver af ekstrem stor karakter (Nielsen, 2015). Et eksempel på sådanne opgaver kunne være at forudse et vejrsystem, kortlægge et genom, energi-forskning, træne en AI, mm. (IBM, u.å).

En HPC-arkitektur består af adskillige *klynger af noder*, som er et begreb der bruges til at betegne de enkelte computere i HPC-systemet (Google Cloud, u.å). Disse noder har forskellige funktioner og udregner parallelt algoritmer, som de udstøder til dataservere, hvis funktion er at gemme outputtet, som vist i figur 9 (ibid.). Dette kaldes for *parallel computing*, som er illustreret i figur 10 og er en væsentlig grund til, at HPC kan udføre opgaver langt hurtigere end normale computere (ibid.).



Figur 9, Illustration af "computer klynge", Introduction to HPC with MPI for Data Science, 2015



Figur 10, Illustration af "parallel computing", Introduction to HPC with MPI for Data Science, 2015

For at kunne træne AI modeller fra bunden, der er på niveau med de "state of the art" modeller, der findes i dag, kræver det HPC, grundet de enorme mængder data, de skal trænes på. En af de fremmeste AI modeller, GPT 4 af OpenAI, har krævet ca. 10^{25} flops at træne (EPOCH AI, u.å). FLOPS står for "floating point operations per second", som er et mål for, hvor mange flydende punkt beregninger (dvs. beregninger med tal, der indeholder decimaler) en computer kan udføre pr. sekund (Sheldon, 2023). Man snakker om FLOPS,

da det er beregninger som plus, minus, gange eller dividere og ikke fx booleske værdier som computere også er i stand til at udregne.

På samme måde, som man kan måle en bils maksimale ydeevne ved X-antal kilometer i timen, så er FLOPS en metode, der gør det muligt at kvantificere ydeevnen af supercomputere, især inden for videnskabelig forskning, ingeniørarbejde og kunstig intelligens.

Ved den førnævnte model, GPT-4, ville det tage ca. 1,6 millioner år at træne den via en standard computer, hvor det med et moderne HPC-system vil tage ca. 3 måneder. For udregninger og nærmere detaljer henvises der til bilag 7.

Vi kan konkludere, at det ikke er muligt med standard computerkræft blot at sætte sig og træne en model fra bunden, tilsvarende til de state of the art modeller, der er offentligt kendte som bl.a. GPT 4 fra OpenAI og Gemini fra google.

Dog er dette ikke nødvendigt, da man som individ kan downloade disse modeller fra internettet, i stedet for selv at skulle træne dem, og på denne måde kan et individ med begrænset adgang til computerkraft stadig få adgang til de store modeller som der findes på internettet. Dette har dog den konsekvens, at man er begrænset af de retningslinjer diverse udviklere har sat for deres modeller, desto mindre man går ind og *finetuner* modellerne, som med andre ord betyder at man finjusterer modellens måde at behandle input-data på, hvilket man kan gøre til den grad ens adgang til computerkraft tillader det. Dette giver dog de samme udfordringer, som hvis man selv skulle træne en model fra bunden. For yderligere information om finetuning henvises der ligeledes til bilag 7.

Systemer

Som vi har klargjort i den teknologiske del af analysen ovenfor, så er deepfake-teknologien, en teknologi hvis succes afhænger af at kunne forbigå filtre, der kan vurdere dem til at være falske. Lykkedes dette, har man produceret en tilfredsstillende deepfake, og man kan atter sprede den på diverse platforme. Vi vil dog ikke mene, at teknologien i sig selv bør bebrejdes, da det som påvist i den teknologiske analyse, kan anvendes til en meget bred vifte af formål. Dette afsnit vil argumentere for at deepfakes, i henhold til

vores problemformulering, ikke kan isoleres, men at det skal ses i lyset af et samspil med en masse andre teknologier og fænomener, og at det først er her, der vil kunne påpeges eventuelle konsekvenser ved en sådan teknologi. Vi har identificeret fem fænomener og teknologier, der indgår i samspil med deepfakes, før den kan anses som en mulig trussel mod demokratiet. Disse er henholdsvis sociale medier, desinformation, nyheder, ekkokamre & confirmation bias, hvoraf flere af dem vil sammenflettes i de uddybende afsnit nedenfor.

Sociale medier og desinformation

Skal deepfakes udgøre en mulig trussel mod demokratiske samfund, kræver det først og fremmest, at folk ser eller hører dem, og altså dermed på en eller anden måde er eksponeret for dem. Deepfakes flourer primært på socialmedie-platforme, som Facebook, Twitter og Instagram, hvor disse medier muliggør en bred deling og spredning af nyheder og information (Olan et al., 2024). Disse platforme har som hovedformål at muliggøre interaktion og kommunikation mellem brugerne på tværs af forskellige delte opslag. Det er et miljø, hvor individer kan dele interesser og opbygge relationer med andre, som har lignende værdier og overbevisninger. Med ubegrænset adgang til at dele information kan mennesker frit udtrykke deres synspunkter og værdier, hvilket også inkluderer spredning af falske nyheder og desinformation, herunder deepfakes (Ibid.). Dette fænomen er især fremtrædende på platforme som Facebook, hvor grupper dannes, og ekkokamre opstår. I kapitlet "Social Media, Echo Chambers, and Political Polarization" af Pablo Barberá udtrykkes bekymring for ekkokamre, da adskillige studier har vist, at eksponering for færre synspunkter i politiske sammenhænge har en tendens til at fremme politisk polarisering. Som beskrevet i artiklen af Femi Olan et al. spredes falske nyheder typisk i specifikke grupper, eller segmenter, for at promovere ideologier, der bekræfter modtagernes eksisterende overbevisninger. Bekymringerne bliver tydelige, når man ser på, hvordan falske nyheder, såsom deepfakes, flourer på sociale medier uden, at borgerne nødvendigvis har kendskab til deres falskhed. En verdensomspændende undersøgelse, foretaget af virksomheden iProov, viste at 71% af deltagerne ikke havde kendskab til begrebet deepfakes (Reddy, 2023). Af de 29% som kendte til teknologien,

mente 43% at de ikke ville kunne genkende en deepfake, hvis de blev præsenteret for en (Ibid.).

Ydermere, er sociale medier oftest designet med indholds rangeringssystemer baseret på algoritmer, der har til formål at generere indhold, der tiltrækker millioner af visninger, ved at dele nyheder eller information rettet mod et specifikt publikum, hvilket betyder at eksponering af desinformation har potentiale til at accelererer (Olan et al., 2024). Ifølge DeepMedia, er spredningen af visuelle deepfakes tredoblet fra 2022 til 2023, imens forekomsten af auditive deepfakes er ottedoblet (Konopliov, 2023). Det er altså primært på sociale medier, at individer kan opleve at blive eksponeret for videoer, billeder og lydclip der er forfalskede.

Nyheder

Spredningen af deepfakes foregår primært på sociale medier, men vi finder det også interessant at undersøge, hvor folk går hen for at få nyheder. En undersøgelse foretaget af Kim Christian Schrøder og Mark Ørsten, til en større rapport om digitale nyheder af Reuters Institute, viser at vi i Danmark får 81% af vores nyheder online i 2018, herunder sociale medier (Newman et al., 2018). Under disse 81%, udgør sociale medier ca. 46%, og selvom denne kategori er faldet, er den stadig væsentligt højere end printede aviser, som udgør ca. 27% i 2018 (Ibid.). Den printede avis er også markant mere reguleret, og har et større hold fagfolk bag hvert udgivet eksemplar, hvilket står i klar kontrast til sociale medier. I USA er statistikkerne lignende, dog med et større skel mellem sociale medier og den printede avis, hvor sociale medier står for 45% af folks nyheder, og den printede avis står for kun 21% af folks nyheder (Ibid.). Her er differencen altså 24%, hvor den i Danmark er 19%. Det er dog værd at bemærke et fald i sociale medier som nyhedskilde i både Danmark og USA, og at TV stadig udgør den største nyhedskilde i begge lande. I Danmark er TV dog konsekvent faldet fra 2013 frem til 2018 (Ibid.), hvor den i USA er stagneret på omkring 74% (Ibid.). Selvom sociale medier ikke udgør den største nyhedskilde, mener vi stadig at mellem 40% og 50% er meget højt for en ureguleret nyhedskilde, hvori deepfakes kan spredes, særligt hvis samme mennesker udelukkende får nyheder derfra. Ydermere er strukturen af sociale medier opbygget på algoritmer, der har mulighed for at holde

informationsstrømme nogenlunde ensformige, et fænomen vi her i opgaven referer til som ekkokamre.

Ekkokamre & Confirmation Bias

Ekkokamre spiller også en central rolle i aktiveringen af deepfakes som en demokratisk trussel. Det gør den i samspil med confirmation bias, som vi har redegjort for i teori afsnittet længere oppe. Samspillet mellem disse to, kan skabe en arena hvori folk bliver eksponeret for deepfakes, der stemmer overens med deres egne holdninger og meninger, mens meget tyder på at disse mennesker er mere villige til at overse at en video er en deepfake. Som nævnt tidligere, er disse ekkokamre baseret på komplekse algoritmer, kaldt *anbefalingsalgoritmer*. Anbefalingsalgoritmer, også kendt som recommender systems, er en subkategori under maskinlæring og er en teknologi der bruger algoritmer til at anbefale og sortere i forslag til en given bruger baseret på den brugers tidligere ageren online (Nvidia, u.å). Disse algoritmer findes i forskellige former og kan skabe ekkokamre på forskellige måder, blandt andet ved den algoritme kaldt *collaborative filtering*, som anbefaler nyt indhold til en bruger baseret på hvad lignende brugere synes godt om (Desai, 2023). Alle sociale medier benytter sig af disse algoritmer for automatisk at opdatere deres brugere med nyt indhold, og grundet algoritmernes natur vil det også betyde at de hurtigt kan blive eksponeret for meget af det samme indhold.

Mange deepfakes, er på nuværende tidspunkt, ikke visuelt helt realistiske sammenlignet med lydfileerne, som kan fremstå utroligt ægte (Mai, 2023), hvorfor der kan være grund til bekymring for at information og fakta udvandes, og at man vil kunne skræddersy usande budskaber, til et særligt publikum, der er mere modtagelige over for selv samme budskab. Man vil kunne skabe narrativer baseret på forfalsket video-, billede- og lydmateriale, og plante disse i ønskede fora.

Delkonklusion

Vi har ovenfor forsøgt at nuancere samtalen om deepfakes, og vi har fundet frem til at deepfakes isoleret set ikke udgør en trussel mod demokratiet. Ser man dog deepfakes i

lyset af de fem fænomener beskrevet ovenfor, kan man indlede samtalen om hvorvidt deepfakes kan være et demokratisk problem. Det kræver altså et større væv af teknologier, som er at finde i en moderne demokratisk og teknologisk verden. Viser man på nuværende tidspunkt en deepfake til kritisk tænkende individer, vil det højst sandsynligt ikke få en større konsekvens. Eksponerer man dog folk for deepfakes i isolerede og polariserede grupper, med medlemmer der er mere tilbøjelige til at tro på narrativer, der stemmer overens med deres egne overbevisninger, mener vi at det potentielt kan influere demokratiet. Hvilke konkrete aspekter af demokratiet det kan udfordre, vil vi uddybe efter vores analyse af den kvalitative undersøgelse.

Interviews

Den 26. april 2024 offentliggjorde politikerens, Morten Messerschmidt, en video af Danmarks statsminister, Mette Frederiksen, der afholder et pressemøde. Dette pressemøde var ikke hændt i virkeligheden, men var en video genereret af deepfake-teknologi (Lerhard, 2024).

Denne video, på trods af at den var skiltet som værende en deepfake, skabte røre blandt befolkningen, men også blandt politikere. Efter videoen blev lagt ud på sociale medier, blev der indgået en aftale blandt Præsidiets og gruppeformændene for Folketingets partier, at der skal laves et kodeks, for hvordan kunstig intelligens, herunder deepfakes, skal og må bruges. Oveni dette skal det undersøges, hvordan politikere kan tackle risikoen for at andre kan bruge kunstig intelligens til at sætte dem i dårligt lys til fremtidige valgkampe (Folketinget, 2024). Denne situation har skabt en unik mulighed for at undersøge, hvordan stemmeberettigede danskere forholder sig til spredningen af deepfakes og falske informationer på sociale medier.

Formålet ved interviewene er at få mulighed for at få indsigt i hvordan danskere på sociale medier forholder sig til deepfakes, og yderligere om det påvirker deres tillid, både til hvad de ser på sociale medier, men også deres generelle tillid til danske politikere. Vi har valgt at interviewpersonerne skal høre lyden fra deepfake-videoen af Mette Frederiksen, men ikke se den, så de kan fokusere mere på hvad der bliver sagt. Denne

overvejelse kommer også af, at vi internt i gruppen har vurderet videoen til at være mindre troværdig rent visuelt.

I nedenstående interviews, omtaler vi begrebet desinformation som misinformation, da vi på tidspunktet for vores interviews, ikke havde skelnet mellem de to begreber.

Interview 1

Facebook er den primære kilde til nyheder

Interviewpersonen, som fremover i rapporten vil blive omtalt ”S1”, indledte interviewet ved at fortælle at hun var 73 år og havde en baggrund som tresproglig korrespondent og folkeskolelærer. Herudover har S1 haft en profil på Facebook siden år 2008. S1 erkendte også at hun, ifølge hende selv, bruger alt for meget tid på platformen, og særligt om natten, hvilket påvirker hendes søvnvaner (Bilag, s. 1-2).

Interviewet bevægede sig herefter hen mod hvor hun primært fik sine nyheder fra, hvor hun siger at Facebook er hendes primære kilde til nyheder. Her følger hun særligt med i Politiken, Berlingske og DR Nyheder. Udover dette ser S1 også nyheder på fjernsynet, hvor hun foretrækker TV2 News, da hun mener de har det bedste format og den bedste formidling af nyheder.

I kølvandet på dette stillede vi (jf. Interviewguiden, bilag 6) spørgsmål vedrørende hendes holdning til politikere på sociale medier. Hun udtrykte at hun synes at politikere på diverse social medie-platforme, virker mere moderne og interesserede i at nå ud til folk, også selvom S1 ikke delte politiske holdninger med dem. Her brugte hun Alex Vanopslagh som eksempel, som hun mente var et frisk pust i det politiske landskab, selvom hun ikke sympatiserede med hans politiske holdninger (Bilag 1, s. 2-4).

Dette må være AI

Lydklipet bliver afspillet og S1 smågriner lidt og nævner at det må være AI, selvom hun erkender at lydklipet lyder meget som Mette Frederiksen, men dog kan S1 også godt regne ud at det ikke er Mette Frederiksen som udtaler sig i lydklipet (Bilag 1, s. 4).

Hun udtrykte sin bekymring over manglende gennemsigtighed og troværdig i forhold til deepfakes, og disse kan medføre vrede og protest i offentligheden. S1 mente at det er afgørende for samfundet at beskytte sig mod deepfakes samt misinformation, og dette kræver en indsats fra IT-eksperter, men også at befolkningen skal forholde sig kritisk til det de ser på sociale medier (Bilag 1, s. 5-8).

Afslutningsvis i interviewet reflekterede S1 over hendes egen sårbarhed som ældre i forhold til svindel, samt hendes tillid til den yngre generations evne til, at være opmærksom og skarp på hendes og andre ældre borgeres vegne (Bilag 1, s. 9-11).

Interview 2

IT-konsulenten

S2 er 58 år gammel, med faglig baggrund som elektronikmekaniker, IT-konsulent og tillidsrepræsentant.

Derefter fortalte S2 at han er på sociale medier som Facebook, Instagram og til tider LinkedIn. Ifølge ham, bruger han op til en time om dagen på sociale medier. Han bruger de sociale medier til at se på kattevideoer, holde sig opdateret på sit netværk og sine yndlingskunstnere. S2 oplyser også at hans primære kilde til nyheder på sociale medier er gennem LinkedIn, og ellers finder han sine nyheder på TV2 News, DR og de lokale nyhedsmedier (Bilag 2, s. 1-3).

S2 følger desuden også politikere på sociale medier, og han nævner at politikere på sociale platforme giver et frisk pust, men han er dog skeptisk overfor det glansbillede som politikerne tegner (Bilag 2, s. 2-3).

Den er nedladende og manipulerende

Efter afspilningen af lydklippet, nævnte S2 med det samme at det er en deepfake som Morten Messerschmidt har lavet af Mette Frederiksen. Han fandt den nedladende, manipulerende og racistisk. S2 udtrykte bekymring overfor politikere, der skaber falske videoer for at skade deres politiske modstandere (Bilag 2, s.4).

Ved spørgsmålet om tillid til politikere og sociale medier generelt, svarede S2 at hans tillid ikke ændres efter at han så videoen, men at han finder politikere mindre interessante, hvis de bruger hadsk retorik eller skaber falske videoer (Bilag 2, s. 5).

Der blev også spurgt ind til bekymringen for deepfakes potentiale til at manipulere opfattelsen af virkeligheden og han udtrykte en tydelig bekymring for, hvordan teknologien kan bruges til at sprede misinformation og opfordrede til teknologiske løsninger for at beskytte samfundet (Bilag 2, s. 5-7).

Sidste kommentar fra S2 omhandlede den hurtige udvikling af teknologi og hans bekymring for dens potentiale (Bilag 2, s.8).

Interview 3

Sociale medier - primært for underholdningens skyld

Interviewet med den næste interviewperson, som i rapporten bliver omtalt som S3, er 20 år gammel og har gennemført en folkeskoleuddannelse. S3 er aktiv på sociale medier som Facebook, Snapchat, Instagram, X og YouTube. YouTube er primært for underholdningens skyld. Han har været aktiv på sociale medier i cirka 8-10 år og bruger dagligt 5 timer, hvilket kan stige til 8 timer på fridage (Bilag 3, s. 1-2).

På Facebook ser S3 primært tilbud fra supermarkeder og diverse videoer. S3 støder også nogle gange på nyheder fra medier som BT, Ekstra Bladet og TV2 Nyheder på Facebook. Derudover bruger han TV2 Nyheds-appen til at læse nyheder. Der blev spurgt ind til om S3 følger politikere på sociale medier, hvilket han kunne svare ja til. Han følger diverse politikere, men mest for underholdningens skyld. S3 mener at politikere som bruger sociale medier, appellerer til de unge, dog har han ikke en stor tillid til politikere generelt (Bilag 3, s. 2-5).

Uenig med beslutningen om at afskaffe yderligere helligdage

Efter afspilningen af lydklippet, blev der spurgt ind til S3's tanker om det der blev sagt på den fiktive pressekonference. Han gav udtryk for, at han normalt ikke er enig i Mette

Frederiksens holdninger, men synes dog at hun håndterede coronakrisen godt som statsminister. Dog var S3 meget uenig i hendes beslutning om at afskaffe yderligere helligdage. Han mente, at danskerne arbejdede meget, især i betragtning af landets høje skattetryk, og derfor mente han, at det var vigtigt at bevare helligdagene, selvom han ikke selv ville bemærke forandringen (Bilag 3, s. 6).

Efter at interviewereren afslørede, at lydklippet var en deepfake, kommenterede S3, at det ellers lød som noget, Mette Frederiksen godt kunne finde på at sige (Bilag 3, s. 7).

Han udtrykte sin bekymring over, hvordan manipulerede videoer kunne påvirke tilliden til politikere og den information der bliver delt. S3 mente at det ville blive svært at vide, hvad man skulle tro på, hvis politikere, samt andre begyndte at bruge deepfakes til at sprede misinformation (Bilag 3, s. 8-11).

Interview 4

Bruger sociale medier til at se opslag fra sin omgangskreds

S4 er 23 år gammel og har gennemført en ungdomsuddannelse. Hun har været aktiv på Facebook, Instagram og Snapchat i cirka 10 år. Hun bruger cirka 3-4 timer om dagen på sociale medier, og S4 bruger dem til at se opslag fra sine venner og hendes personlige netværk.

Hun bliver ofte eksponeret for nyheder på sociale medier og primært omhandlende globale begivenheder og politik. Dog opsøger hun ifølge hende selv, ikke sine nyheder fra sociale medier og ej heller fra traditionelle medier, som tv og aviser, da hun ikke er videre interesseret i nyheder (Bilag 4, s. 1-3).

Hun følger to politikere på sociale medier, Mette Frederiksen og Lars Løkke Rasmussen. S4 er positiv over politikere som bruger sociale medier til at nå ud til unge vælgere, da de traditionelle medier ikke når lige så langt ud til den yngre generation (Bilag 4, s. 3).

Det er barnligt

Efter lydklippet blev afspillet var S4's første kommentar, at det er forrykt at fjerne helligdage og indføre muslimske helligdage. Men hun kunne godt spotte at det var en deepfake, da det ikke er noget Mette Frederiksen kunne finde på at sige. S4 mente at politikere der laver falske videoer, opfører sig barnligt og misbruger sociale medier til mudderkast frem for at sprede deres egne politiske budskaber. På trods af lydklippet mente hun ikke, at det påvirkede hendes generelle tillid til politikere, da hun allerede var kritisk over for politisk adfærd på sociale medier (Bilag 4, s. 4-5).

Senere i interviewet udtrykte S4 en bekymring af deepfakes potentiale til at vildlede befolkningen, men hun mente, at folk bør være kildekritiske og ikke tro på alt det de ser på sociale medier. Ifølge hende, er det i sidste ende brugeren af sociale mediers ansvar at være kritisk, da dem bag sociale medier har ikke ansvaret for at forhindre spredningen af misinformation, medmindre det direkte skader konkrete individer (Bilag 4, s. 6-8).

Interview 5

Sociale medier er den primære kommunikationsmiddel

Personen i interviewet vil blive omtalt som S5. Hun er 36 år og har en EUD Business. S5 er aktiv på sociale medier som Facebook, Instagram og Snapchat. Hun fortæller at hun har været på Facebook siden år 2008 og har haft Instagram siden 2010. Hun bruger ca. 4 timer om dagen, men nævner det er svært konkret at vurdere, da Messenger og Instagram er hendes primære kommunikationsmiddel med sin omgangskreds (Bilag 5, s. 1-3).

S5 bruger derudover sociale medier til at holde sig opdateret på nyheder. Hun følger for eksempel TV2 News, BT og Se og Hør. Udover nyheder, så har hun også fulgt politikere på sociale medier, som Mette Frederiksen, men hun stoppede med at følge med, da der ikke kom relevant indhold op. På trods af dette, synes S5 at politikere på sociale medier, virker mere personligt og gør deres budskaber mere forståelige og relaterbare (Bilag 5, s. 3-6).

Det må være en aprilsnar

S5 reaktion til lydklippet, er at det ikke kan være rigtigt og det må være en aprilsnar, da indholdet i klippet lød usandsynligt. Efter det blev fortalt at det var en deepfake, sagde S5 at hun synes det er både sjovt og skræmmende, fordi det kan skabe store konflikter og udfordre politikernes personlige sikkerhed. Hun mener at man skal bruge sin sunde fornuft, men er bekymret for den potentielle skade og misinformation spredningen af deepfakes kan medføre (Bilag 5, s. 6-10).

S5 har generelt lav tillid til det indhold hun støder på sociale medier, da folk ofte viser sig fra deres bedste sider og der er meget manipulation. Af den årsag er hun meget bekymret for sociale mediers magt og dens påvirkning, som det har på folks selvværd og virkelighedsopfattelse (Bilag 5, s. 11-12).

S5 nævner der bør være mere fokus på oplysning, især til unge, om eksistensen og farerne ved deepfakes og manipulation. Hun foreslår at, information om deepfakes spredes bredt, herunder til ældre via foreninger. Desuden anerkender S5, at sociale medier har nogle mekanismer til anmeldelse af upassende indhold, men finder det svært at finde en balance mellem beskyttelse og frihed på platformene. Hun ser et stort dilemma i, hvordan sociale medier regulerer indhold og definere grænser for, hvad der acceptabelt og hvordan man skelner mellem uskyldige sjove videoer og skadelige manipulationer (Bilag 5, s. 11-14).

Konklusioner fra interview

De fem interviews giver et indblik i, danske borgeres holdninger til brugen af deepfakes i politiske situationer, samt deres tillid til indholdet på sociale medier og politikere generelt. Der tegner sig et klart mønster på tværs af interviewene, af en generel bekymring for deepfakes potentiale. Samtlige interviewpersoner nævner, at de er deciderede bekymrede. Dette skal dog ses i lyset af at alle på nær én, ikke tog fejl af lydklippet, og var i stand til at identificere deepfaken.

S3 kunne ikke skelne fakta fra fiktion i dette tilfælde, og udtrykker derudover at han generelt ikke sympatiserer med Mette Frederiksens politik. Vi mener ikke at én ud af fem er en generaliserbar statistisk, dog finder vi det stadig interessant at den ene person der udtrykte størst modstand mod Mette Frederiksens politik, også var den der antog at

lydklippet var ægte. Dette kan tyde på en sammenhæng mellem confirmation bias og evne til at identificere en deepfake.

Vi kan dog også ud fra resultaterne, konkludere at lydklippet ikke synes helt ægte. Dette kan både være et resultat af det indholdsmæssige aspekt af lydklippet, eller den generelle kvalitet af deepfakes. Sidstnævnte kan vi dog i højere grad eliminere, da der var bred enighed på tværs af interviewene om, at det lød meget som Mette Frederiksens stemme. Vi har dog ikke haft mulighed for at kontrollere disse variable, da vi ikke havde adgang til lignende videoer og lydclip i kontekst af dansk politik, med en mindre satirisk natur.

Vi vil dog endnu en gang fremhæve de mange bekymringer vi kunne uddrage fra undersøgelsen. Særligt ældre borgere i et demokrati er ifølge flere af vores interviewpersoner i farezonen. Dette kan skyldes at deepfake-teknologi, i sin nuværende form, er et meget nutidigt fænomen. Den yngre generation vil i højere grad blive eksponeret for teknologi på dette niveau tidligt i livet, hvorimod den ældre generation vil skulle tilpasse sig. Der er bred enighed blandt interviewpersonerne om at deepfakes kræver en stillingtagen, herunder at sørge for at informere ældre borgere om teknologien. Herudover foreslås der også teknologiske løsninger, og der bliver påpeget at en kritisk sans, overfor det man bliver eksponeret for på sociale medier, er nødvendig.

Afslutningsvis vil vi derudover påpege, at confirmation bias kan være svært at have indsigt i. Flere af interviewpersonerne affejr lydklippet som værende en deepfake, og det kan potentielt være et resultat af politisk holdning. Det auditive indhold, gav udtryk for en ekstrem politisk holdning, som mange ikke ville forbinde med Mette Frederiksen, og udfaldet af vores kvalitative undersøgelse kunne potentielt have været anderledes, hvis den afbilledet politiker havde været en anden.

Demokratisk analyse

I følgende kapitel, vil vi inddrage de to teorier først introduceret i teori-afsnittet, af Robert Dahl (1989) og Niklas Luhmann (2017), der ligger til grund for vores definition af et velfungerende og effektivt demokrati. Robert Dahl opstiller nogle forudsætninger han mener er essentielle for et velfungerende demokrati, hvilket vi her i opgaven, som nævnt,

har adopteret som vores ideelle demokrati. Niklas Luhmann forholder sig ikke specifikt til demokratiske samfund, men vi mener at hans begreb om social kompleksitet (Luhmann, 2017), er meget relevant for det moderne demokrati, hvor sociale interaktioner er blevet flere og mere komplicerede. Disse to teorier vil vi sidestille med deepfakes i kontekst af de fem fænomener og teknologier, vi redegjorde for i analyse-afsnittet ”systemer”.

Tillid i et demokrati

Som nævnt i teori-afsnittet, så baserer Luhmanns teori sig blandt andet på begrebet social kompleksitet, som er et udtryk for, hvor effektive og handledygtige borgere er i et samfund (Luhmann, 2017). Er den sociale kompleksitet høj, har man ikke meget tillid til hinanden, hvorfor det kan være nødvendigt at skulle gennemskue og analysere ens medborgeres, og de demokratiske institutioners, intentioner og måder at handle på. Ifølge denne teori, kan mennesker på sigt blive handlingslammede, da sociale interaktioner vil kræve enorm betænkningstid, grundet en grundlæggende mistillid på tværs af samfundet (Ibid.). Denne teori kan holdes op mod deepfake-teknologiens funktioner, og derigennem undersøge om deepfakes og systemerne omkring det, potentielt kan forværre menneskers tillid til demokratiet, samt evne til at indgå i det.

Vi vil herunder argumentere for, at deepfake-teknologien i et større system, udgjort af de omtalte fænomener og teknologier i ”systemer” afsnittet, har mulighed for at øge den sociale kompleksitet markant. Deepfakes kan, som nævnt, udvande og sløre grænsen mellem sandt og falsk. Dette muliggøres særligt ved brugen af NLP, der som nævnt er en teknologi der har til formål at kunne matche lyd og billede, så det fremstår så ægte som muligt. Dette betyder altså at der ikke længere er konsensus om de objektive fakta, da man vil kunne overbevise folk om usande informationer, ved at få dem til at fremstå som ægte, samt komme fra kilder som modtagerne er mere tilbøjelige til at tro på. Med dette følger flere konsekvenser.

Først og fremmest kan det fører til yderligere politisk polarisering og mistillid over det politiske spektrum, som videre potentielt kan føre til en eskalering af konflikter og markant forringe den sociale sammenhængskraft. Dette kunne ses under det

amerikanske præsidentvalg i 2020, hvor præsidentkandidat Joe Biden blev anklaget for at have et ringe mentalt helbred. Dette blev også konsensus blandt mange af hans modkandidats tilhængere, som følge af flere deepfake-videoer af Biden der fremstår kognitiv ustabil. Videoerne havde overskrifter som ”Joe Biden falder i søvn i løbet af et interview”, ”Farer vild i løbet af et kampagneevent” og ”Taler over sig” (Carnahan, 2020). Havde man set disse videoer og havde vanskeligt ved at skelne mellem deepfake eller ej, ville man stille sig enormt kritisk overfor Joe Biden som præsident, samt overfor hvem end der kunne finde på at stemme på ham, da han på videoerne tydeligvis ikke havde mental kapacitet til at være statsoverhoved. Bygger to politiske grupperinger deres argumenter på vidt forskellige sandheder, vanskeliggør det muligheden for saglig og konstruktiv diskussion, og videre at nå til en fælles løsning. Kan man ikke nå til fælles løsninger, samt grundlæggende er mistroisk overfor ens medborgere, vil dette fænomen på sigt potentielt kunne destabilisere det deliberative demokrati.

Herudover, kan deepfakes igen øge den sociale kompleksitet, hvis der ikke kan skelnes mellem fakta og fiktion. Bliver deepfake-fænomenet alment kendt, og er folk klar over dens tilstedeværelse, vil man muligvis begynde at sætte spørgsmålstegn ved videoer der er ægte. Dette fænomen blev fremhævet i et studie foretaget af studerende på University College Cork, der havde til formål at undersøge, hvordan personer på platformen X reagerede på deepfakes der var blevet delt herpå. Studiet viste, at grundet denne viden om at de ville støde på deepfakes, blev de adspurgtes tillid til det viste materiale undermineret, uanset om det var deepfakes eller ej (University College Cork, 2023). En konsekvens heraf kan være, at folk klassificerer ægte videoer som deepfakes, mens de omvendt vil kunne klassificere deepfake-videoer som ægte. Dermed vil det igen være vanskeligt at differentiere hvad der sandt og hvad der falskt, hvilket videre kan hæve den sociale kompleksitet, og nødvendiggøre grundigere analyse af ens medborgeres intentioner og handlinger.

Derudover, kan den øgede sociale kompleksitet, som resultat af intern mistillid og manglen på konsensus om fakta, også ende ud i at folk i højere grad søger mod populistiske, og i værste fald autoritære, politiske bevægelser og institutioner, der forsøger at gøre krav på sandheden (Hansen, 2018). Her kan der opstå simple løsninger på meget komplekse problemer, hvilket kan være attraktivt for borgere i et demokratisk

samfund med høj social kompleksitet (Ibid.). Med dette kan der potentielt opstå endnu mere splittelse, da én politisk fløj kan beskyldes den anden for at overse eller bevidst negligere en, i deres øjne, indlysende løsning.

Ydermere, kan deepfakes på sigt også anvendes som en politisk strategi til manipulation og desinformation. Med det accelererende tempo deepfakes har udviklet sig i, er det ikke utænkeligt at deepfake videoer kommer til at fremstå fuldstændigt realistiske inden for nærmeste fremtid. Dette bringer anledning til bekymring, da hvis det på sigt ikke bliver reguleret, vil kunne skabe grobund for diverse smædekampagner og såkaldt mudder-kast på tværs af politiske partier, på et hidtil uset niveau. Kan man fremføre videoer af politikere der fremstår i ulovlige eller på anden vis dårlige situationer, vil det kunne have indflydelse på deres fremtoning og offentlige persona, som videre kan svække deres position i et demokrati. Dette vil endnu en gang øge den sociale kompleksitet, da folk vil kunne overbevises om, at deres medborgere stemmer på utroværdige politiske kandidater, der har indgået i meget kompromitterende situationer.

Ovenstående konsekvenser kunne tyde på at deepfakes, på de rette platforme, potentielt kan øge den sociale kompleksitet. Dette kan, ifølge Luhmann (2017), føre til den førnævnte handlingslammelse, som vi mener er harmfult for et effektivt demokrati.

Dahls demokratiske kriterier

Udover en lav social kompleksitet, har vi også baseret vores succeskriterier for et demokrati ud fra Robert Dahls kriterier for demokratiet (Dahl, 1989). Ud af de fem kriterier beskrevet i teori-afsnittet om Robert Dahl, har vi gennem vores teknologiske analyse af deepfake teknologien, identificeret særligt to af dem, der potentielt trues ved fremkomsten af mere avancerede deepfakes. Disse to er som nævnt, oplyst forståelse og lige kontrol over den politiske dagsorden (Ibid.).

Den oplyste forståelse, som vi har gjort rede for i teori-afsnittet om Robert Dahl, kan potentielt opleve udfordringer, på grund af udbredelsen af deepfake-teknologien. Som belyst i vores teknologiske analyse, er en deepfake først færdigproduceret, når den har gennemgået tilpas mange filtreringsprocesser, hvor den bliver tiltagende realistisk hver gang. Som belyst i den teknologiske analyse, er GAN, Autoencoders og NLP, processer

der tilsammen kan udgøre utroligt realistiske afbildninger, hvis formål er at skabe videoer der kan synes ægte. Deepfake-teknologien er altså afhængig af at kunne efterligne den virkelige verden så meget som muligt, og kan dermed være enormt effektiv i spredningen af desinformation. Desinformation, som vi også har redegjort for, ser vi som et klart modspil til en oplyst forståelse (jf. ”fake news på sociale medier” i teori), da det netop ikke klargøre og synliggøre alternativer til et politisk forslag (Ibid.), men i stedet kan skabe alternativer baseret på usande informationer. Dette står i klar kontrast til hvad Dahl fremhæver som én af fem kriterier for et effektivt demokrati (Ibid.).

Lige kontrol over den politiske dagsorden, er som nævnt også et kriterie, vi ser der, med stor sandsynlighed, vil blive udfordret ved fremkomsten af deepfake-teknologi i samspil med sociale medier, algoritmer og confirmation bias. Dette er dog ikke så ligetil som den førnævnte oplyste forståelse, men kigger man på mulighederne med deepfake-teknologi, er dette kriterie potentielt også i risikozonen. Kan man forvrænge virkeligheden, og eksempelvis generere forfalsket videomateriale af en politiker i dårligt lys, vil man kunne rette fokus mod denne fiktive skandale, i stedet for reelle problemer. Her foregår påvirkningen meget ubevidst, men kontrollen over den politiske dagsorden er i dette tilfælde i større grad i hænderne på de personer der generede videoen. Dette kan også koncentrere kontrollen over den politiske dagsorden ned på en specifik gruppe mennesker, der har kapacitet til at skabe deepfakes, der kan skabe røre i befolkningen.

Endnu en gang kan det tyde på, at deepfake-teknologien under de rette omstændigheder, vil kunne influere og destabilisere et demokrati, efter Dahls definition. Det kræver, som nævnt, at flere faktorer spiller ind, men er disse faktorer på plads, ser vi altså særligt den oplyste forståelse og den lige kontrol over den politiske dagsorden særligt udfordrede af deepfake-teknologien.

Delkonklusion

Deepfake-teknologien, og særligt de aspekter af den der har til formål at sikre en vis realisme og troværdighed, såsom GAN, NLP og Autoencoders, kan vanskeliggøre tilliden mellem medborgere, samt udfordre nogle af demokratiets grundstøtten. En teknologi der kan anvendes til at fremstille fiktion som fakta, rejser flere spørgsmålstejn til

informationslandskabet. Dahls oplyste borgerskab, kan pludselig splittes til to eller flere oplyste borgerskab, der alle tror at deres version af sandheden er den rigtige. Dette kan videre føre til en grundlæggende mistillid på tværs af disse borgerskab, der kan øge den sociale kompleksitet, der potentielt vil kunne føre til meget ineffektive samfund, hvor folk ikke længere er handlingsdygtige nok til at kunne indgå i et demokrati. Dette skal også ses i lyset af den accelererende fart, som teknologien udvikler sig i, og bliver særligt NLP og Autoencoders mere avancerede, vil man på sigt, uden regulering, kunne snyde selv de mest årvågne demokratiske borgere.

Diskussion

Vi vil i følgende afsnit gennemgå og diskutere de teoretiske og metodiske overvejelser der ligger til grund for vores rapport. Med dette følger der også en diskussion om den generelle vinkel vi har taget, samt overvejelser om hvilke alternativer der kunne være. Vi vil ydermere perspektivere til eventuelle fremtidige rapporter om samme emne.

Teoretiske overvejelser

Vi har som teoretisk grundlag for vores analyse af demokratiet valgt at inddrage Robert Dahl og hans kriterier for det ideale demokrati. Til trods for at Dahls teori er ældre, finder vi, at disse kriterier stadig i betydelig grad er relevante for vores nutidige forståelse af et demokrati, samt de forventninger, der er forbundet hermed. Især har vi lagt vægt på to af Dahls kriterier, som beskrevet i teori- og analyseafsnittet, da vi vurderer, at disse er særligt relevante i forhold til, hvordan deepfake-teknologi kan påvirke demokratiet, som diskuteret i vores analyse.

Det var oplagt for os at bruge disse kriterier, fordi de, for vores opgave, danner et teoretisk grundlag for hvordan man kan anskue et demokrati. Dette betyder at vi kan komme med en teoretisk funderet analyse, fremfor at den beror på vores egen subjektive anskuelse af et demokrati. Vi er bevidste om, at Dahls teori forsøger at opstille de ideelle forudsætninger for et demokrati, i en tid hvor medier var en markant anderledes størrelse

i dag, og hvor information blev formidlet markant anderledes end i dag. Som vi argumenterer for i løbet af rapporten, ændrer teknologi på flere facetter af et demokratisk samfund, hvorfor det er vigtigt at overveje om Dahl stadig er den rigtige teoretiker at citere, i snakken om det ideelle demokrati. Vi mener dog, at selvom teknologi udvikler sig i et accelererende tempo, er der sket få ændringer i de grundlæggende demokratiske institutioner og praksis, hvorfor vi mener at det er relevant at netop inddrage en ældre demokratisk model, og stille den overfor de nutidige teknologiske fænomener.

Havde vi valgt at undersøge vores problemstilling gennem pragmatismens perspektiv, en filosofisk retning der prioriterer praksis og erfaring som grundlaget for viden og sandhed, kunne vi have nået frem til andre konklusioner. Ved at inddrage pragmatisme ville vi have haft en mere løsningsorienteret tilgang til deepfake-teknologien og analyseret, hvordan denne teknologi kunne integreres i det moderne demokrati og styrke de eksisterende demokratiske processer. Dette ville indebære en dybdegående undersøgelse af de praktiske konsekvenser af deepfake-teknologiens anvendelse. Pragmatisme fokuserer på positive muligheder, hvilket kunne have ført os til at identificere måder, hvorpå deepfake-teknologi altså kunne bidrage til demokratisk innovation. Derudover ville vi have fokuseret på erfaringsbaseret viden, kontinuerlig forbedring og en handlingsorienteret tilgang, hvilket ville have påvirket vores analysemetoder og konklusioner markant. Da dette ikke var vores ønske med denne rapport, fandt vi at Dahls teorier og kriterier opfyldte de krav vi havde til vores teori.

Trods at teorien "Theory of Trust" også er en ældre teori, har den alligevel en essentiel relevans i dag, da social kompleksitet er stigende på grund af blandt andet teknologiske fremskridt og politisk polarisering. Vi synes, det er vigtigt at inkludere Luhmanns teori i projektet for at få en dybdegående forståelse af de mekanismer, som forenkler social kompleksitet.

Uagtet af, at Niklas Luhmanns teori blev udtænkt før den nuværende teknologiske tid, har dens fundamentale ideer stadig en væsentlig indflydelse på samfundet i dag. Teorien stammer oprindeligt fra 1979, og det er dermed nødvendigt at belyse de betydelige teknologiske ændringer siden da, og deres indflydelse på et samfund og de sociale

interaktioner heri. Men trods disse ændringer, mener vi at det er en fordel at inkludere teorien i vores projekt, da vi nu lever i en verden med stadig mere komplekse sociale interaktioner, som følge af teknologiske forandringer og politisk polarisering. Dette nødvendiggør yderligere mekanismer, der kan mindske den sociale kompleksitet.

I tilfælde af, at vi undlod at inkludere teorien, havde vi skulle definere tillid på anden vis. Eftersom vi inkluderer begrebet i vores problemformulering, samt af flere omgange har det stående som et centralt fænomen hele vejen igennem rapporten, ser vi os nødsaget til at få klargjort og defineret et meget abstrakt begreb. Vi tog udgangspunkt i Luhmanns "Theory of Trust" da vi fandt hans teori om social kompleksitet relevant for både den verden vi lever i, men også den problemstilling vi vil undersøge. Alternativt kunne vi havde søgt mere mod psykologisk litteratur og empiri, hvilket havde belyst andre aspekter, mere rettet mod den enkeltes færden i et avanceret demokrati, i kontrast til vores mere institutionelle og kollektive analyse. Ydermere kan Luhmanns teori kan hjælpe med at overskueliggøre konsekvenserne af mistillid og videre, hvordan man kan reducere den sociale kompleksitet, som er essentielt for et demokrati i en moderne og teknologiske verden.

I takt med den voksende tilstedeværelse af sociale medier, fandt vi det relevant at benytte teori om begrebet ekkokamre, da fænomenet har fået en stigende betydning og indflydelse på vores samfund og demokrati. Ekkokamre bidrager til en ubalance i den offentlige debat, hvor brugere isoleres fra andre perspektiver end deres egne. Vi ville undersøge dette fænomen, da det vil kunne skabe en arena for spredning af desinformation og deepfakes og dermed udnyttes til at manipulere politikere og danne mistillid til medier og demokratiet. Dette kan true fundamentet for demokratiet, da borgere i stigende grad ville træffe beslutninger baseret på fordrejede eller falske oplysninger.

Da vi kun kunne have fænomenet ekkokamre med i et begrænset omfang i denne opgave, kunne det være interessant i et fremtidigt studie at gå mere i dybden med fænomenet, da det indgår som en forudsætning for at deepfakes kan påvirke borgernes tillid til, samt evne til, at indgå i et demokrati. Da vi skulle begrænse os rent teoretisk, ville det være

interessant at inddrage begrebet *cross-cutting* i en fremtidig opgave, da det er en kontrast til begrebet *ekkokamre*. *Cross-cutting* omhandler eksponering af forskellige synspunkter, som oftest ikke er ens egne, hvilket kan modvirke polarisering og i sidste ende kan det eventuelt modvirke de *ekkokamre* som der opstår på sociale medier. Disse to begreber vil i samspil kunne lave en nuanceret analyse af spredning af desinformation på sociale medier, da *ekkokamre* er et begreb der debatteres meget, særligt om hvorvidt det overhovedet eksisterer. På baggrund af vores teknologiske analyse af algoritmer på sociale medier, og den eksisterende litteratur om *ekkokamre*, mente vi dog at vi kunne understøtte begrebet tilstrækkeligt.

Metodiske overvejelser

I forbindelse med valg af metode har vi bevidst valgt at anvende litteratursøgning som den primære tilgang til vores teknologiske analyse. Baggrunden for dette valg er, at det ud fra vores perspektiv har været den mest fyldestgørende måde at analysere teknologien bag skabelsen af deepfakes på. Vi havde i overvejelserne at benytte en kvalitativ metode i form af interviews med eksperter inden for feltet. Dog vurderede vi, at denne metode sandsynligvis ikke ville bidrage med væsentlig ny information, som vi ikke selv kunne finde gennem en omfattende litteratursøgning.

Derudover overvejede vi at konstruere en deepfake-video fra bunden, for at opnå en dybere forståelse af de underliggende systemer og processer, så vi kunne videreformidle denne viden i vores teknologiske analyse. Imidlertid kræver dette, som nævnt i vores teknologianalyse, en betydelig computerkapacitet i form af High-Performance Computing (HPC) samt tid til at træne et Generative Adversarial Network (GAN) til dette formål. Da denne mulighed synes langt fra at være realistisk, nåede vi frem til at litteratursøgning ville være mest passende til dette.

Vi har dog valgt at anvende en kvalitativ tilgang, i forbindelse med den mere STS-baserede demokratiske analyse, hvor vi interviewede fem personer om deres viden og holdninger til deepfake-teknologien i et demokrati. Det mente vi, var en nødvendighed for at få et indblik i forskellige perspektiver og yderligere nuancer, som vil kunne bidrage til vores analyse som helhed. I vores optik var dette relevant, da vi dermed selvstændigt har kunne

indsamle empiri, der understøtter vores teori om tillid og demokrati, og derigennem undersøge hvilke demokratiske udfordringer vi som individer potentielt kan stå over for, i forbindelse med deepfake-teknologiens udbredelse.

I vores interviews tog vi udgangspunkt i en deepfake-video, hvor Danmarks statsminister Mette Frederiksen er hovedpersonen. Vi valgte at gøre brug af denne, da videoen er gået viralt i Danmark, og vi fandt den dermed yderst relevant, da det er en dansk politiker, Morten Messerschmidt, som har delt den. Selvfølgelig var det visuelle overbevisende falsk, hvorimod Mette Frederiksens stemme, altså det auditive, nemt kunne narre selv den skarpeste. Dog var indholdet af videoen for nogen tydeligt satirisk, hvilket gjorde det nemt at skelne mellem hvorvidt det var en deepfake eller ej.

Vores empiri understøttede, at selvom det var nemt at gennemskue satiren, så berørte det stadig vores interviewdeltagere på det følelsesmæssige plan, at det nu er muligt at skabe materiale, som kan være så overbevisende, at det på sigt kan være en udfordring at skelne mellem fakta og fiktion. Dermed stillede vi os selv spørgsmålet om, hvorvidt udfaldet af vores empiri kunne have set anderledes ud, i tilfælde af, at vi havde valgt at inddrage en deepfake-genereret video med større troværdighed. Vi er med den overbevisning, at resultatet med stor sandsynlighed havde været et helt andet, men vi valgte at benytte videoen af Mette Frederiksen, da det også var et interessant aspekt at blive klogere på, hvilke følelser det vækker i individer at blive eksponeret for falsk genereret indhold. Herudover var det også interessant at undersøge, hvordan det potentielt kan påvirke tilliden til demokratiet, at en politiker netop vælger at dele en deepfake-video af en anden politiker.

I vores metodiske overvejelser reflekterede vi desuden over at anvende kvantitative metoder i form af spørgeskemaer, hvor vi kunne have vedhæftet den pågældende video og stillet spørgsmål, hvis svar kunne kvantificeres. Denne metode ville give os mulighed for at nå ud til et betydeligt større antal mennesker, og dermed opnå et bredere empirisk grundlag for vores analyse og konklusioner. Spørgeskemaer indebærer dog en risiko for misforståelser og fejlfortolkninger af de spørgsmål der stilles, hvilket kan introducere målefejl i den indsamlede empiri.

I betragtning af, at tillid er et fundamentalt og centralt begreb i vores rapport, vurderede vi, at vores empiri ville stå markant stærkere ved at møde vores deltagere ansigt til ansigt. Denne tilgang giver netop mulighed for at indsamle menneskelige og konkrete detaljer, som kun kan observeres gennem direkte menneskelig interaktion. Dette er af særlig betydning, da vores mål har været at opnå en dybere forståelse af det psykologiske aspekt omkring, hvorvidt deepfake-teknologien kan påvirke befolkningens tillid til, samt evne til, at indgå i et demokrati.

Konklusion

På baggrund af vores analyse og diskussion kan vi konkludere, at deepfakes i sig selv ikke nødvendigvis udgør en direkte trussel mod demokratiet. Men når deepfakes ses i sammenhæng med andre teknologier og de nævnte fænomener, kan de potentielt udgøre et demokratisk problem på grund af kompleksiteten i informationslandskabet. Deepfakes, typisk videoer eller lydoptagelser, manipuleres ved hjælp af avancerede kunstige intelligens-teknologier for at fremstå autentiske og virkelige. Disse teknologier muliggør skabelsen af realistiske, men falske repræsentationer af mennesker og begivenheder, ved at analysere og genskabe visuelt og auditivt indhold.

Deepfake-teknologien består, som belyst i vores teknologiske analyse, af flere komponenter, der arbejder sammen for at kunne skabe realistiske afbildninger af virkeligheden. Af disse teknologier er særligt GAN, NLP og Autoencoders, de komponenter hvis primære formål er at sikre en vis troværdighed, og dermed gøre det svært at skelne en deepfake fra virkeligheden.

Vi vil dog afslutningsvist igen påpege, at deepfake-teknologien ikke isoleret set bør anses som et problem. Mens ser man den i samspil med fænomener som sociale medier, ekkokamre og confirmation bias kan denne teknologi være med til at underminere troværdigheden af rigtige videoer, mens den kan skabe narrativer baseret på falske videoer. Dette kan videre føre til en generel mistro til informationslandskabet, da man som borger i et demokratisk samfund, ikke længere er sikker på hvad der er sandt og hvad der er falskt. En generel mistro til informationslandskabet, vil også føre til en splittelse på

tværs af samfundet, da folk vil basere deres opfattelse af fakta på vidt forskellige basis. Dette kan potentielt føre til en mistillid blandt medborgere, samt til de folkevalgte politiske leders funktionsevne og troværdighed. Der kan ydermere spredes en generel utryghed, hvilket er belyst af vores interviews, hvori samtlige interviewpersoner udtrykker en bekymring for deres medborgeres evne til at kunne skelne mellem deepfake eller ej.

Den endelige konklusion er dermed, at deepfakes i et større system af de nævnte fænomener, som finder sted i en moderne og teknologisk verden, potentielt kan influere demokratiske borgers evne til at indgå i, samt deres tillid til, det demokratiske system. Det tyder på at i et folkestyre, hvori folk har forskellige versioner af sandheden, vil der kunne være grobund for en grundlæggende mistillid og konflikter på tværs af det politiske spektrum, hvilket vil vanskeliggøre saglig debat og muligheden for at finde frem til fælles løsninger.

Litteraturliste:

Artikler

Abirami, S., & Chitra, P. (2020). Advances in Computers, Volume 117 (s. 353-355). Elsevier Inc. Lokaliseret på: <https://doi.org/10.1016/bs.adcom.2019.09.007>

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social network analysis and mining*, 13(1), 30. Lokaliseret på: <https://doi.org/10.1007/s13278-023-01028-5>

Diakopoulos, N., & Johnson, D. (2020). "Anticipating and addressing the ethical implications of deepfakes in the context of elections". *Sage Journals*. Lokaliseret på: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3474183

Evans, J. S. B. T. (2016). "Reasoning, Biases and Dual Processes: The Lasting Impact of Wason (1960). Lokaliseret på: <https://journals-sagepub-com.ep.fjernadgang.kb.dk/doi/epub/10.1080/17470218.2014.914547>

Fallis, D. (2014): A Functional Analysis of Disinformation. Lokaliseret på: <https://core.ac.uk/download/pdf/19961024.pdf>

Hansen, A. D. (2018). Populism and Democracy: Laclau's theory of Populism: The Royal Road to Totalitarianism?. 1-17. Paper presented at ECPR 2018 Joint Sessions of Workshops, Nicosia, Cyprus. Lokaliseret på: https://rucforsk.ruc.dk/ws/portalfiles/portal/63706585/adh_paper_Laclau_populism_and_democracy_cyprus.pdf

Löblich, M., & Venema, N. (2021). Echo chambers: A further dystopia of media-generated fragmentation. Lokaliseret 23/05/2024 på: <https://www.degruyter.com/document/doi/10.1515/9783110740202-010/html>

Mai, K. T., Bray, S., Davies, T., & Griffin, L. D. (2023). Warning: Humans cannot reliably detect speech deepfakes. PloS One, 18(8), e0285333–e0285333. Lokaliseret på: <https://doi.org/10.1371/journal.pone.0285333>

Newman, N., Fletcher R., Kalogeropolous, A., Levy, A. L. L., Nielsen, R. K. (2018). Reuters Institute Digital News Report 2018. Lokaliseret på: <https://www.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf>

Olan, F., Jayawickrama, U., Arakpogun, E. O., Suklan, J., Liu, S. (2022). "Fake news on Social Media: the Impact on Society". Information Systems Frontiers (2024) 26:443-458.

Lokaliseret på: <https://link-springer-com.ep.fjernadgang.kb.dk/article/10.1007/s10796-022-10242-z?fromPaywallRec=false>

Smith, M. H., & Zadeh Chair, L. A. (u.å). "Stuart J. Russell". Berkeley University of California. Lokaliseret på: <https://vcresearch.berkeley.edu/faculty/stuart-russell>

The Government, Local Governments Denmark, Danish Regions (2016). "A Stronger and More Secure Digital Denmark – Digital Strategy 2016-2020". Agency for Digitisation.

Lokaliseret på: https://digst.dk/media/16165/ds_singlepage_uk_web.pdf

Hjemmesider

Adobe (u.å). *Hvad er generativ AI og hvordan virker det?* Lokaliseret [8/3/2024] på: <https://www.adobe.com/dk/products/firefly/discover/how-generative-ai-work.html>

AdCreative.ai (u.å). *AI-Værktøj til Annoncering*. Lokaliseret [15/4/2024] på: <https://da.adcreative.ai/>

Als, R. (2022). *Dagens tegning: Afgrundsdyb enighed*. Politiken. Lokaliseret [20/3-24] på: <https://politiken.dk/debat/dagenstegning/art8816493/Afgrundsdyb-enighed>

Barney, N. *What is deepfake AI?* Techtarget. Lokaliseret [04/04/2024] på: <https://www.techtarget.com/whatis/definition/deepfake>

Bergmann, D., & Stryker, C. (2023). "What is an autoencoder?" IBM. Lokaliseret [04/04/2024] på: <https://www.ibm.com/topics/autoencoder>

Carnahan, D. (2020). *Faked videos shore up false beliefs about Biden's mental health*. The Conversation. Lokaliseret [11/04-2024] på: <https://theconversation.com/faked-videos-shore-up-false-beliefs-about-bidens-mental-health-145975>

Dalberg-Larsen, J. (2009). *Niklas Luhmann*. Den Store Danske. Lokaliseret [27/5-2024] på: https://denstoredanske.lex.dk/Niklas_Luhmann

DBpedia (u.å). *About: Peter Cathcart Wason*. Lokaliseret [1/6/2024] på: https://dbpedia.org/page/Peter_Cathcart_Wason

Desai, U. (2023). *Recommendation Systems Explained: Understanding the Basic to Advance*. Medium. Lokaliseret [28/04/2024] på: <https://utsavdesai26.medium.com/recommendation-systems-explained-understanding-the-basic-to-advance-43a5fce77c47>

EPOCH AI (u.å). *Training Compute of Notable Machine Learning Systems Over Time*. Lokaliseret [16/05/2024] på: <https://epochai.org/data/epochdb/visualization>

Eskildsen, E. (2022). *Desinformations uhyggelige historie: Fra opdigtede nazister til folkedrab på 'kakerlakker'*. Videnskab.dk. Lokaliseret [07/5-2024] <https://videnskab.dk/kultur-samfund/desinformations-uhygge-historie-fra-opdigtede-nazister-til-folkekrab-paa-kakerlakker/>

Folketinget (2024). *Præsidiets indkalder til møde om deepfake*. Lokaliseret [23/5/2024] på: <https://www.ft.dk/da/aktuelt/nyheder/2024/05/praesidiet-indkalder-til-moede-om-deepfake>

Google Cloud (u.å). *What is high performance computing?* Lokaliseret [06/05/2024] på:
<https://cloud.google.com/discover/what-is-high-performance-computing>

Heim, L., Anderljung, M., Bluemke, E., & Trager, R. Cen.
Computing Power and the Governance of AI. Centre for the Governance of AI.
Lokaliseret [01/06-2024] på: <https://www.governance.ai/post/computing-power-and-the-governance-of-ai>

Holdsworth, J. (2024). *What is natural language processing (NLP)?*. IBM. Lokaliseret
[23/05/2024] på: <https://www.ibm.com/topics/natural-language-processing>

IBM (u.å). *What is high-performance computing (HPC)?* Lokaliseret [06/05/2024] på:
<https://www.ibm.com/topics/hpc>

Jungco, K. (2024). *100 Top AI Companies Trendsetting In 2024*. Datamation. Lokaliseret
[29/03/2024] på: <https://www.datamation.com/featured/ai-companies/>

Kavlakoglu, E. (2020). *NLP vs. NLU vs. NLG: The differences between three natural
language processing concepts*. IBM. Lokaliseret [23/03/2024] på:
[https://www.ibm.com/blog/nlp-vs-nlu-vs-nlg-the-differences-between-three-natural-
language-processing-concepts/](https://www.ibm.com/blog/nlp-vs-nlu-vs-nlg-the-differences-between-three-natural-language-processing-concepts/)

Konopliov, A. (2023). *Fake News Statistics & Facts (2023)*. Redline Digital. Lokaliseret
[11/04-2024] på: <https://redline.digital/fakenews-statistics/>.

Ledford, H. (2024). *Deepfakes, trolls and cybertroopers: How social media could sway
elections in 2024*. Nature. Lokaliseret [02/03-2024] på:
<https://www.nature.com/articles/d41586-024-00274-7>

Lerhard, J. (2024). *Efter hård kritik: Morten Messerschmidt forsvare falsk video med Mette Frederiksen*. Politiken. Lokaliseret [08/05-2024] på:

<https://politiken.dk/kultur/medier/art9879175/Morten-Messerschmidt-forsvarer-falsk-video-med-Mette-Frederiksen>

Nvidia, N. (u.å) *Recommendation System*. Lokaliseret [28/04/2024] på:

<https://www.nvidia.com/en-us/glossary/recommendation-system/>

Ordnert (2003). *desinformation*. Lokaliseret [13/05/2024] på:

<https://ordnet.dk/ddo/ordbog?query=desinformation>

Ou, T. (2022). *A Simple AutoEncoder and Latent Space Visualization with PyTorch*.

Medium. Lokaliseret [2/03/2024] på: <https://medium.com/@outerrencedl/a-simple-autoencoder-and-latent-space-visualization-with-pytorch-568e4cd2112a>

Peters, R (2021). *Deepfake*. Den Store Danske. Lokaliseret [31/4/2024] på:

<https://denstoredanske.lex.dk/deepfake>

Reddy, R. (2023). *24 Deepfake Statistics - Current trends, growth, and popularity (December 2023)*. ContentDetector.AI. Lokaliseret [11/4-2024] på:

<https://contentdetector.ai/articles/deepfake-statistics#:~:text=Experts%20expect%20s>

Schomer, A. (2023). *GENERATIVE AI IN FILM & TV: A SPECIAL REPORT*. Variety.

Lokaliseret [22/05/2024] på: <https://variety.com/vip-special-reports/generative-ai-in-film-tv-1235792168/>

Sheldon, R. (2023). *Floating-point operations per second (FLOPS)*. Techtarget.

Lokaliseret [09/05/2024] på: <https://www.techtarget.com/whatis/definition/FLOPS-floating-point-operations-per-second>

Springer (2023). *About Springer*. Lokaliseret [03/04-2024] på:

<https://www.springer.com/gp/about-springer>

Stanford University (u.å). *Peter Norvig*. Lokaliseret [3/6/2024] på:

<https://hai.stanford.edu/people/peter-norvig>

Svensson, P. (2009). *Robert A. Dahl*. Den Store Danske. Lokaliseret [3/5/2024] på:

https://denstoredanske.lex.dk/Robert_A._Dahl

University College Cork (2023). First ever study of wartime deepfakes reveals their impact on news media. Lokaliseret [11/04-2024] på:

<https://www.ucc.ie/en/news/2023/first-ever-study-of-wartime-deepfakes-reveals-their-impact-on-news-media-.html>

Walmsley, R. (u.å). *Cass R. Sunstein*. Harvard Law School. Lokaliseret [03/06/2024] på:

<https://hls.harvard.edu/faculty/cass-r-sunstein/>

Yasar, K. (2023). *Generative adversarial network (GAN)*. Techtarget. Lokaliseret

[9/4/2024] på: <https://www.techtarget.com/searchenterpriseai/definition/generative-adversarial-network-GAN>

Undervisningsmateriale

Christiansen, H. (2024, February 6). *Vigtige begreber om maskinlæring* [Teknologiske Systemer og Artefakter]. <https://acrobat.adobe.com/id/urn:aaid:sc:EU:a1ef3c8f-425a-4259-ab92-3b2fcc5fbf56>

Christiansen, H. (2024, February 6). *Kunstige neurale netværk og deep learning – en ultrakort introduktion* – [Teknologiske Systemer & Artefakter]. <https://acrobat.adobe.com/id/urn:aaid:sc:EU:a8c928d2-12b5-456e-a528-ae81eb078e4f>

Jørgensen, N. (2018). *Digital signatur. En eksemplarisk analyse af en teknologis indre mekanismer og processer*. Roskilde Universitet. Lokaliseret på: <http://webhotel4.ruc.dk/~nielsj/research/publications/indre-mekanismer.pdf>

Jørgensen, N. (2019) *TRIN-modellen er ikke en metode* – [Teknologiske Systemer & Artefakter]. https://ruc-thirdroom.dk/wp-content/uploads/2019/05/Bilag_3.pdf

Bøger

Barberá, P. (2020). Social Media, Echo Chambers, and Political Polarization. In N. Persily & J. A. Tucker (Eds.), *Social Media and Democracy* (s. 34–55). chapter, Cambridge: Cambridge University Press.

Chen, R. H., & Chen, C. C. (2022). *Artificial intelligence* (First edition.). Boca Raton, Florida: CRC Press.

Dahl, R. A. (1989). *Democracy and its critics*. New Haven: Yale University Press.

Janeczko, B., & Srivastava, G. (2022). *Internet of Multimedia Things (IoMT)* (s. 144-145). Amsterdam: Elsevier Inc.

Kaddoura, S. (2022). *A Primer on Generative Adversarial Networks* (1st ed.). New York City: Springer Cham.

Luhmann, N., Morgner, C., & King, M. (2017). *Trust and power* (H. Davies, J. Raffan, & K. Rooney, Trans.; English ed). United Kingdom: Polity Press.

Nielsen, F. (2015). *Introduction to HPC with MPI for Data Science* (s. 3-20). Switzerland: Springer International Publishing Switzerland.

Nielsen, F. (2015). *Introduction to HPC with MPI for Data Science* [Photograph]. Switzerland: Springer International Publishing Switzerland.

Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (s. 1152). United Kingdom: Pearson.

Tanggaard, L., & Brinkmann, S. (2010). *Interviewet: Samtalen som forskningsmetode. I L. Tanggaard, & S. Brinkmann, Kvalitative metoder: en grundbog* (s. 33-42). København: Hans Reitzels Forlag.