

AI

SCENARIER. ETISKE OVERVEJELSER & ANBEFALINGER

INDHOLD

INDTRODUKTION & EXECUTIVE SUMMARY

TVÆRGÅENDE PRINCIPPER

SUNDHED

Scenarie 1: Personlig medicin

Scenarie 2: Når robotter diagnosticerer & opererer

Scenarie 3: Implanteret overvågning & androider

TRANSPORT

Scenarie 1. Smart city

Scenarie 2: Førerløse biler

FINTECH

Scenarie 1: AI fintech assistenter

Scenarie 2: Nye betalingsformer

Scenarie 3: Algo traders styrer de finansielle markeder

ANBEFALINGER

DEFINITIONER

LITTERATUR

Rapporten er bestilt af SIRI-kommissionen og udarbejdet af Pernille Tranberg og Gry Hasselbalch fra tænkehandletanken DataEthics.eu

SIRI-kommissionens øvrige rapporter kan findes her: <https://ida.dk/ida-star/siri-kommissionen>

Introduktion & executive summary

Kunstig intelligens, AI, er et ombejlet digitaliseringsemne. Fremtidsscenariet, hvor alt og alle er forbundne via intelligente maskiner og netværk, der nonstop udveksler data, som afgør vores muligheder, definerer og guider os, er ikke længere en scifi fortælling, men tæt på en realitet. Alt fra biler over lygtepæle til kaffemaskiner bliver 'intelligente', og vores kroppe bliver i stigende grad koblet til de intelligente netværk.

AI er under hastig udvikling i alle former for innovation og tiltrækker nogle af de største private og offentlige investeringer. Samtidig er det blevet et globalt konkurrencefelt lande og regioner imellem.

SIRI-kommissionen blev skabt i sommeren 2016 af Ida Auken, MF for Radikale Venstre og Thomas Damkjær Petersen, formand for IDA. Formålet er at skabe debat om kunstig intelligens og den betydning det har for det danske samfund, både i forhold til vækst, arbejdsmarked, kompetencekrav og etiske udfordringer. SIRI-kommissionen har gennem de sidste 2 år gået i dybden med kunstig intelligens i de tre brancher transport, sundhed og finansområdet. Vi har dog fundet det relevant at gå yderligere i dybden med de etiske dilemmaer og udfordringer, der er dukket op under arbejdet. Det er formålet med denne rapport.

AI bygger på ideen om at forudse mønstre og risici, effektivisere processer og vejlede os. Mulighederne er enorme på sundheds-, transport og det finansielle område. Men i kølvandet på en højhastig udvikling, hvor lovgivning og normer halser efter, opstår etiske dilemmaer. Vi udfordres som mennesker, vores rettigheder, normer og regelsæt får ny betydning, og nye samfundsproblemstillinger tager form.

I rapporten introduceres fremtidsscenarier inden for de tre områder, som SIRI-kommissionen har haft særligt fokus på. Rapporten peger på de etiske implikationer og dilemmaer for hvert fokusområde, men også tværgående principper og ikke mindst anbefalinger til handling.

Vi bør

- sikre, at vi som borgere er bedst muligt rustede til at leve i verden med konstant påvirkning af AI,
- ikke udvikle AI med det formål at erstatte mennesket, men sikre, at AI altid komplementerer mennesket og 'kun' er en cobot, altså en robot, der samarbejder med mennesket,
- sikre, at mennesket altid har det sidste ord og at et menneske altid ved, om han/hun taler med en maskine eller med et menneske,
- altid tænke privatlivsrisici og dataetiske implikationer for individet ind i design og udvikling fra begyndelsen,
- sikre danskerne retten til anonymitet - som man har ret til en hemmelig adresse,
- udvikle en standard for, hvordan man forklarer en algoritme,
- som virksomheder, organisationer og myndigheder, udvikle egne dataetiske principper, der går længere end datalovgivningens krav,
- stille krav til, at der målrettet arbejdes for at mindske problematiske bias i data
- tænke lighed ind i udviklingen og designet af AI tjenester og systemer, således også ressourcetsvage får gavn af udviklingen.

Tværgående principper

Nedenstående principper bygger på eksisterende diskussioner (jvf. litteraturlisten), men tager udgangspunkt i de specifikke etiske dilemmaer i scenarierne inden for de tre områder transport-, sundhed- og finanssektoren.

MENNESKET I CENTRUM

En AI tjeneste har en vis grad af autonomi. Dens analyser, forudsigelser og beslutninger hjælper os med at forstå komplekse situationer baseret på store datasæt, som et menneske ikke vil kunne sætte sig ind i. Men samtidig har vi et stigende behov for, vi som borgere har viden og kendskab nok til, hvordan AI virker til at kunne bevare kontrollen og kunne gennemskue, den påvirkning, vi bliver udsat for. Hvordan sikrer vi, at vi som mennesker altid har forrang for institutionelle og kommercielle interesser? At mennesket er i centrum og har den primære gavn af databehandlingen? Og hvordan sikrer vi, at befolkningen vi er bedst muligt rustet til at leve i en verden med konstant påvirkning af algoritmer?

INDIVIDUEL DATAKONTROL

Dataprofiler bestående af mange forskellige persondata er ofte en forudsætning for, at en intelligent personaliseret teknologi fungerer optimalt. Privatlivsrisici og dataetiske implikationer for individet bør derfor tænkes ind i design og udvikling helt fra begyndelsen af en udviklingsproces, hvor privacy by design-principper er udgangspunkt og dermed bliver indbagt i designet.¹ Det er vigtigt altid at sikre forbrugerens og borgernes egenkontrol af data. Et forsikringsselskab eller en udbyder af lån kan fx have interesse i at få indsigt i de detaljerede dataprofiler, der opbygges via de nye AI tjenester, der kan bruges til fastsættelse af præmier eller lån. Men hvor går grænsen og hvordan sikres det, at kunden også har taget et informeret valg? Hvordan sikres individuel datakontrol, handlekraft og prioritering af det enkelte menneskes selvbestemmelse?

GENNEMSKUELIGHED

AI drevne processer og beslutninger skal kunne forklares, så mennesker forstår de sociale, etiske og samfundsmæssige risici og konsekvenser.² Der skal ikke blot gives basal information om fx databehandling. Det skal dokumenteres og forklares, hvordan en given algoritmisk beslutning er truffet; hvilke kriterier og parametre, der ligger til grund for fx en kreditvurdering, forsikringspræmie eller tildeling af social ydelse. Hvordan sikres det, at mennesker har kontrol over de AI drevne processer?

ANSVARLIGHED

Ansvarlighed dækker en række mulige dataetiske tiltag, som tager udgangspunkt i GDPR, såsom anonymisering, minimal eller ingen brug af metadata, zero-knowledge (sletning af data, selv om man godt må bevare dem), ingen salg eller deling af identificerbare data til tredjepart, ansvarlig organisatorisk forankring samt auditering. Det store spørgsmål bliver, hvem der har ansvaret for disse nye AI-baserede beslutningstagere? Er det menneske-lægen eller robot-lægen? Den menneskelige chauffør eller producenten af den selvkørende bil?

En dataetisk virksomhed eller organisation påtager sig ansvaret for sine produkter og tjenester og sikrer, at såvel samarbejdspartnere og underleverandører lever op til samme dataetiske principper som en selv. Og den har en person eller enhed, der tager sig af etisk håndtering af data. Spørgsmålet er, om de AI-

¹ <https://dataethics.eu/e-t-phone-home-det-intelligente-legetoj/>

² <https://dataethics.eu/midt-i-en-algoritme-oekonomi/>

baserede maskiner skal have ansvar på samme måde som en advokat? Skal de have juridisk status med et nyt juridisk rammeværk?

LIGEVÆRDIGHED

Gennem den digitale udvikling har der været voksende samfundsskel mellem befolkningsgrupper (Digital divides). AI udviklingen giver os nye muligheder, og den kan være med til at udligne skel. Men de kan ligeså vel skabe nye kløfter, hvis lighed ikke tænkes ind i designet og regelsættet fra begyndelsen. I et nyt og forholdsvist ureguleret AI marked er det først og fremmest de ressourcetsvage, der må betale med deres data og som udsættes for intens dataprofilering. På det finansielle marked er det dem med stærkest computerkraft, der kan håndtere de hurtigste og skarpeste handler. Forbrugerbeskyttelse og de nye AI tjenesters kapacitet hænger herudover ofte sammen med, om man kan betale for tjenesten. Algoritmer er objektive matematiske formler. Men når vi fodrer selv-lærende algoritmer med data, er de klassificeringer de foretager, ikke nødvendigvis objektive, men kan reproducere eller skabe magtfordelinger i samfundet. Et datasæt kan indeholde bias, hvorfor en AI dreven beslutning kan være diskriminerende. Bias kan også forekomme i designet af en algoritme, som kan kategorisere og stemple mennesker på en måde, der diskriminerer mellem fx befolkningsgrupper. Hvordan sikres lighed i AI?

TRANSPORT

Scenarie 1. Smart City

Du har - som efterhånden alle andre - ikke kørekort, men benytter dig af eldrevne førerløse biler, busser og tog, der er styret som offentlig transport af myndighederne (der findes private alternativer, som kun et mindretal har råd til). De er udstyret med kameraer med ansigtsgenkendelse og sensorer, der monitorerer passagerernes vægt og varmeafgivelse, lugten fra deres mad, størrelsen af bagagen, om de er alene eller holder én i hånden. Du er langtidssygemeldt fra dit arbejde og på offentlig ydelse. Det betyder, at kommunen samkører dine data med data fra andre offentlige registre - herunder politiets og data indsamlet via sensorer fra lygtepæle og andre offentlige steder - for at tjekke, om du fortsat er berettiget til ydelserne samt for at forudse, om du er i risikozonen for at lave offentlig uorden og hvornår du er klar til at gå på arbejde igen.

Borgerens rolle i denne version af en 'smart' by er begrænset til at være modtagere af en muligvis effektiv og sammenhængende service, hvor forsøg på snyd systemet mere eller mindre bliver elimineret pga den effektive data-overvågning. Et systematisk overvågningsystem som ovenstående er et fantastisk effektivt redskab, som de fx forsøger sig med i Kina³, til at regulere offentlig forsørgelse, men det er også en voldsom indgriben i borgernes personlige integritet og privatliv. Hvis vi vil undgå, at smart cities fx bruges til at regulere offentlige ydelser med, skal følgende problemstillinger løses:

- 1) **Interesser.** Ønsker om effektiv offentlig service må ikke sættes over borgernes privatliv og egenkontrol.
- 2) Er der **proportionalitet** i overvågningen? Hvor effektiv man skal være i at fange dem, der snyder versus masseovervågning alle modtagere af offentlig forsørgelse (en særdeles aktuell problemstilling i forhold til, at kommunernes må overvåge borgerne for at bekæmpe social snyd)? Er databehandlingen, der sikrer færre trafikdrab, så værdifuld, at det er okay at kompromittere alle bilisters privatliv ved at skanne alle nummerplader ved indkørslen til en motorvej?
- 3) Det er vigtigt at være opmærksom på mulige **bias** i de data, der bruges i samkøringen af offentlige databaser.
- 4) **Privatliv.** Der er risiko for, at nogle vil føle deres privatliv krænket. Også selv om de har givet samtykke. De mistroiske vil måske tilbageholde data om sig selv, slukke for wifi og bluetooth, der indsamler vigtige data, eller lyve om data for at undgå overvågningen.
- 5) **Datakontrol og transparens.** Det er vigtigt at give borgerne kontrol over egne data. Til at forstå dem og ikke mindst til at udfordre dem. Skal en borger have lov til at slette sine sundhedsdata - hvor går grænsen? Må man fx slette alle sine spor, hvis man har haft cancer?
- 6) **Forklarlighed.** Såfremt der bruges algoritmer til at analysere data og forudse mønstre, kan de så forklares, så mennesket forstår det?
- 7) **Fri zoner.** Er 'smart' altid den bedste løsning? Eller er der brug for teknologiske frizoner i det offentlige rum?⁴

³ <https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html>

⁴ <https://dataethics.eu/smar-te-byer-eller-smarte-mennesker-2/>

Scenarie 2. Førerløse biler

Du skal købe en bil. Du vælger Mercedes Benz⁵, som sælger sig på, at hvis din algoritme-styrede bil på vej ind i et trafikuheld i et splitsekund skal vælge mellem at slå dig eller en fodgænger ihjel, så vil den altid forsøge at redde bilens ejer. Hvad du ikke ved er, at 'LDS Services'⁶ leverer GPS-systemet til Mercedes Benz og LDS Services derudover lever af at sælge anonymiserede lokationsdata til bl.a. politiet (for bedre at placere automatiserede fart-tjek) og private aktører, der er interesserede i lokations-mønstre. Men det er ganske nemt at re-identificere brugere på lokationsdata⁷, hvis der ikke er tale om en tilstrækkelig anonymisering af data. Derfor er din lokation blevet kortlagt over de seneste to år og solgt til utallige virksomheder indenfor ad tech branchen, så fx hver gang du kører ind i på p-pladsen i et shoppingcenter, bliver du 'serviceret' med personaliserede reklamer.

Brugen af førerløse biler kan utvivlsomt blive en kæmpe gevinst for samfundet. Færre uheld, ingen problemer med alkoholpåvirket kørsel, bedre miljø pga transportoptimering osv. Men med førerløse biler følger en række etiske dilemmaer.

- 1) **Ulighed og privatliv.** I de første design af førerløse biler indkorporeres 16 videokameraer og over 20 forskellige sensorer⁸. Vil det fremadrettet kun være de mest velhavende, der har råd til private førerløse biler, hvor de selv er i kontrol over data - og vil der være bilmærker, der sikrer dem privatlivets fred? Eller skal dette reguleres? Og vil de offentlige førerløse transportmidler sikre, at alle får ret til at kontrollere egne data?
- 2) Skal bilproducenter have lov til at markedsføre sig på, at de altid vil sikre deres **køberes liv over andres liv**? Apropos Moral Machine⁹, der på en forsimplet vis giver os buddet på, hvem en førerløs bil skal slå ihjel i tilfælde af en ulykke, og maskinen på et splitsekund skal beslutte hvem den skal ramme: To ældre mænd eller en dreng og en mand? En overvægtig kvinde eller en atletisk kvinde? En hund eller en kat? En hjemløs eller en læge? Og så videre.
- 3) **Ansvar.** Hvem har ansvaret ved en ulykke i førerløse biler? Bilproducenten, bilejere, bilbrugere?
- 4) **Fra eje til dele.** Ifølge eksperter er det at eje en bil en stærk tradition¹⁰, der giver følelsen af kontrol og frihed. Kan man få danskerne til at opgive den kontrol for, at de førerløse biler bliver så sikre og bæredygtige som muligt?
- 5) Er det okay, at man skanner en nummerplade ved indgangen til et indkøbscenter, for dermed at kunne **personalisere reklamer** til dem, der træder ind i centret?

⁵ <http://fortune.com/2016/10/15/mercedes-self-driving-car-ethics/>

⁶ Opdigtet navn

⁷ https://www.wu.ac.at/fileadmin/wu/d/i/ec/Privacy_Brosch%C3%BCre/Brochure_2702.pdf

⁸ <https://www.nytimes.com/2018/08/16/business/cars-internal-data-networks.html>

⁹ <http://moralmachine.mit.edu/>

¹⁰ Selvflyvende droner er på vej - og de laver helt om på vores liv og hverdag, Politiken 30.7.2018 <https://politiken.dk/vi-den/Tech/art6643992/Selvflyvende-droner-er-p%C3%A5-vej-%E2%80%93-og-de-laver-helt-om-p%C3%A5-vores-liv-og-hverdag>

SUNDHED

Scenarie 1. Personlig medicin

Du går til din læge, som har din sundhedsjournal via sundhed.dk og de oplysninger, du derudover selv afgiver til lægen. Du har fået udarbejdet en fuld DNA-profil, som du har givet lægen adgang til. I flere år har du brugt en skridttæller fra apoteket, hvis data du også har overført til lægen. Ligeledes kan lægen indhente lokationsdata fra din telefon og se, hvor socialt stærk du er ved at vurdere dit sociale netværk via Facebook og andre sociale medier. Med disse data kan lægen vurdere, hvilke risici, du har for at udvikle sygdomme, før du får symptomer og hvilken medicin og personlig adfærd, der vil virke bedst til forebyggelse af sygdommen.

Dette er drømmen i personlig medicin (precision medicin). Den rummer store potentialer for meget bedre sundhedsbehandling, for de AI-drevne løsninger kan - men en vis sandsynlighed - forudsige mønstre i patientdata og foretage risikovurderinger baseret på komplekse data profiler. Fx har læger haft stor succes i klinisk præcisions-medicin med at bruge eksisterende kræftbehandling på helt nye måder baseret på DNA-profiler¹¹. Og det rummer store muligheder for at ændre vores sundhedsvæsen til at fokusere mere på forebyggelse end helbredelse. Men der er lige så mange etiske implikationer:

- 1) Der kan være forskelle i **interesserne** i de AI drevne løsninger: Står patienternes ønsker altid over lægens, systemets eller andre interesserter?
- 2) Det er vigtigt at være opmærksom på mulig **bias** i de data, der bruges i personaliseret medicin. I rapporten Fairness in Precision Medicine (Ferryman, Pitcan, 2018)¹² fremhæves fem områder, der bør analyseres for mulig bias:
 - a) genetiske data, idet virksomheder, der sælger DNA-test, kan klassificere forskelligt
 - b) elektroniske patientjournaler, som kan mangle data, være klassificeret forskelligt eller være svære at fortolke
 - c) diversiteten i deltagere og i datatyper, hvor der i små befolkninger kan være problemer i forhold til at generalisere i alle typer af patienter, ligesom befolkningsdata-grupperingen kan være alt for simplificerede
 - d) historiske bias kan ligge i tidligere undersøgelser, hvor man har rekrutteret for snævert og fx glemte en befolkningsgruppe for at det er repræsentativt, og endelig
 - e) analytiske bias som kan opstå, fordi researchernes bias kan så skinne igennem.
- 3) **Ulighed**. Mindre ressourcestærke patienter kan muligvis ikke få gavn af personlig medicin i samme grad som dem, der er ressourcestærke og har overskud til at gå op i egen sundhed. Derfor kan personaliseret medicin være til mest fordel for de borgere, der er højtuddannede, tech-savy, selvkørende, informationssøgende og dansktalende.
- 4) **Privatliv**. Der er risiko for, at nogle ikke følger rådene i personlig medicin, hvis de føler sig gået for tæt på, udsponeret imod og deres privatliv dermed krænket. Også selv om de har givet deres samtykke. Der er desuden risiko for, at data ender i hænderne på uvedkommende, såsom forsikringsselskaber.
- 5) **Individ-fokus eller struktur-fokus**. Fra et samfundsmæssigt perspektiv er der risiko for, at man med personlig medicin kommer til at fokusere for meget på individet frem for de strukturelle udfordringer, der skaber vores sundhedssektor. Og at man lægger for stort et ansvar over på individet.

¹¹ <https://www.nature.com/articles/s41698-017-0016-z>

¹² https://datasociety.net/pubs/pm/DataSociety_Fairness_In_Precision_Medicine_Feb2018.pdf

- 6) **Datakontrol og transparens.** Det er vigtigt at give patienter kontrol over egne data. Selv om der er en bekymring for, at individuelt ansvar marginaliserer flere, er der flere fordele ved at give patienter adgang til egne data, så længe data kan forstås og bruges til at handle på.

Scenarie 2: Når robotten diagnosticerer & opererer

Efter at være blevet skannet med det samme (fordi AI sænker priserne og omkostningerne ved skanninger kommer man hurtigere til) får du konstateret brystkræft. Ved hjælp af AI er sandsynligheden for, at skanningen er korrekt øget fra 85% til 99,5%¹³. Du bliver også screenet for neurologiske problemer; din stemme bliver analyseret, og så forudser systemet, om der er Parkinson's på vej. Det er der ikke. Men du skal opereres, og det sker ved hjælp af en robot, der hjælper den menneskelige læge med at sy og slå knuder på syningerne.

Udover samme problemstillinger som ved scenarie 1, er der følgende etiske overvejelser;

1. **Overdiagnoser.** Der er risiko for at AI-baserede systemer er over-forsigtige og dermed øger behovet for unødige test og behandlinger. Det modsatte kan også være tilfældet.
2. **Transparens** i AI-diagnoser. Er det overhovedet menneskeligt muligt – som GDPR lægger op til – at forstå logikken bag den AI-foreslåede behandling?
3. **Ansvarlighed.** Hvem har ansvaret for, hvis der er skader som følge af AI-diagnoser og AI-baseret behandling?
4. **Adgang til et menneske.** Har en patient altid ret til at få adgang til et menneske? Eller få en second opinion af et menneske, eftersom maskinen heller ikke altid har ret¹⁴? Og skal et menneske altid træffe den endelige beslutning¹⁵?
5. **Tillid-mistillid.** I nogle tilfælde vil de AI-baserede systemer være udviklet og implementeret af kommercielle virksomheder. Dette kan medføre mistillid til sundhedsvæsenet.
6. **Sundhedspersonalet** kan føle sig truet på deres autoritet, hvis AI-baserede systemer overtager. Det kan betyde, at personalet frigøres fra automatiserede opgaver og kan fokusere på andre vigtige opgaver. Men det kan også føre til, at man i stedet ansætter færre folk med færre kompetencer. Og hvad sker der så, hvis teknologien fejler, og der ikke er menneskelige kompetencer til at tage over?

Scenarie 3. Implanteret overvågning & androider

Du får indopereret en bluetooth og wif-fi connected chip i din søns hjerne for at blive varslet om symptomer på hjernekræft, fordi han er i højrisikogruppen. Da dit barn endnu er pre-teen og ofte bliver væk, samtykker du til en GPS i chippen, så du kan se, hvor han befinder sig. Da han bliver voksen, lader han chippen blive siddende, for så kan han altid findes, såfremt der sker ham noget. Samtidig er det nemt at tilføje nye funktioner til chippen, så han kan bruge den som ID- og betalingskort. Han får også på et tidspunkt en robotarm efter en ulykke i sin soldatertid, og så spiser han piller, der skærper hans opmærksomhed og hjerterytme. Endelig designer han sit eget barn i samarbejde med sin kone, da gen-redigering er blevet tilladt. Da han dør som 105-årig, genopliver hustruen ham som en klonet android.

I dag forskes der i de fleste af ovenstående scenarier. Udover problemstillinger i kølvandet på scenarie 1 og 2 og her et par stykker mere:

¹³ <https://www.fastcompany.com/40570350/5-ways-to-ensure-ai-unlocks-its-full-potential-to-serve-humanity>

¹⁴ <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/>

¹⁵ <https://www.montrealdeclaration-responsibleai.com/the-declaration>

1. **Indopereret tracking.** Er det etisk forsvarligt at tillade tracking af børn (eller ældre eller umyndige for den sags skyld) og er deres samtykke - hvis muligt - nok?
2. **Gen-redigering.** Allerede i 2018 begynder det at blive mere accepteret med gen-redigering¹⁶ (*gene editing*) - hvor skal grænserne gå for det?
3. **Definition af mennesket.** Hvad er et menneske, og hvor går grænserne mellem menneske og robot? Vil vi - som nogle, fx Yuval Harari, påstår¹⁷ - inden for et par årtier opleve en ny race af supermennesker, der er lige så forskellige som mennesker og gorillaer er i dag?
4. **Android-etik & robotrettigheder**
 Især i Japan produceres menneskelignende robotter - androider. Nogle gange er det en basal klon af ens udseende som Henrik Schärfe¹⁸ har gjort det. I science fiction film genoplives mennesker som androider vha alle de data, man har genereret, mens man er i live. Det fordrer svar på en lang række etiske dilemmaer:
 - A. Er det okay at producere robotter, der ligner mennesker?
 - B. Er det okay at producere kopier af mennesker og gå videre end udseendet og reproducere hjerne og personlighed? Og har vi ret til at sige nej til, at andre laver en robotkopi af os selv?
 - C. Skal vi selv bestemme, før vi dør, om den data, vi genererer, må bruges til andre formål efter vores død?
 - D. Har androider menneskerettigheder og ret til statsborgerskab ligesom Sofia¹⁹ fik?
 - E. Hvem har ansvaret for androiders handlinger? Virksomheden, der har produceret den, eller skal de selv have juridisk status?
5. **Italesættelse**
 Den måde vi taler om AI og robotters rolle influerer udviklingen af AI. Nogle taler fx om ukontrollerbar AI med egen fri vilje og om mennesker som 'outdated software'²⁰. Sjældent tænker vi videre over, hvilken filosofisk menneske- og teknologiforståelse, der ligger til grund for vores AI diskurs. AI gennemsigtighed handler også om det sprog, vi bruger, når vi designer, skaber politikker og bruger AI.²¹

¹⁶http://assets.pewresearch.org/wp-content/uploads/sites/14/2018/07/25173131/PS_2018.07.26_gene-editing_FINAL.pdf

¹⁷<https://dataethics.eu/en/wef18-tech-data-enhance-human-life-right-track/>

¹⁸<https://henrikscharfe.com/>

¹⁹<https://www.khaleejtimes.com/region/saudi-arabia/video-sophia-becomes-first-robot-to-receive-Saudi-citizenship>

²⁰<https://www.inverse.com/article/42256-sxsw-2018-dna-out-dated-software-biotech-can-reprogram>

²¹<https://dataethics.eu/en/lets-talk-ai/>

FINANS

Scenarie 1. AI fintech assistenter

Du forvalter din privatøkonomi ved hjælp af en AI drevet personlig assistent, som hjælper med at administrere dine indtægter og udgifter, med at kortlægge din økonomi, samt investere til fremtiden i forhold til dine pensionsordninger. Det er en kompleks proces, hvor dine forbrugsmønstre kortlægges, og du guides mod et bestemt mål, som du selv kortlægger, fx at bo i en villa i nordsjælland eller at gå på pension som 60-årig. Du får hjælp til dine daglige indkøb og til at vælge pensionsfond, og assistenten giver dig besked, hvis du er i gang med et økonomisk risikabelt indkøb. Og så fungerer den samtidig som din personlige algo trader, der deltager på dine vegne på det globale aktie og børsmarked.

Personaliserede AI assistenter kan give brugerne bedre indsigt i samt vejledning til komplekse beslutninger omkring privatøkonomi og mulighed for at deltage i ellers lukkede komplekse makroøkonomiske dynamikker. Det er et område, der allerede er i fuld udvikling.²² Men den kræver regulering, og en række etiske overvejelser bør indgå fra begyndelsen:

- 1) **Interesser og individuel kontrol.** Hvem udvikler tjenesterne? Fintech AI assistenterne vil blive udviklet af traditionelle banker, af stater, startups og nogle større private udbydere. Banker er traditionelt stramt regulerede, men hvad med et forholdsvist underreguleret startup miljø? Hvordan sikres forbrugerinteresserne? Et forsikringsselskab eller en udbyder af lån kan også have interesse i at få indsigt i forbrugernes detaljerede dataprofiler for at fastsætte præmier eller lån. Hvordan sikres, at de målsætninger, de personlige AI tjenester guider forbrugerne ud fra, repræsenterer forbrugerens interesser? Hvilket juridisk regelsæt skal de overholde? Er der behov for udvikling af ny lovgivning på fintech området?
- 2) **Privatliv.** For at en personlig AI tjeneste fungerer optimalt, skal den være sammensat af alle typer data - ikke kun økonomiske data, men fx også data om indkøb, lokation, ønsker og behov, job, uddannelse og politisk overbevisning. Dette indbefatter også en psykologisk profilering. AI bruges i stigende grad til at forudsige mønstre i data og til risikovurderinger i fx. banker.
- 3) **Forklarlighed.** Hvordan sikrer vi, at vi som individer er med på og forstår kriterierne, målsætningerne, beslutningerne og i det hele taget har kontrol over AI assistancen? Eller er det okay, at mennesker ikke forstår AI-beslutningerne?
- 4) **Ulighed og nye digitale kløfter.** Der vil være forskel på de personaliserede AI fintech tjenester. Nogle vil blive udviklet af idealistiske startups eller regulerede banker med forbrugerrettigheder i mente. Men andre vil ikke have samme udgangspunkt. De personaliserede AI tjenester risikerer derfor at skabe digitale kløfter.
- 5) **Ansvar.** Hos hvem eller hvor ligger ansvaret, hvis en personaliseret AI assistent tager fejl eller udfører en dårlig investering på vegne af forbrugeren? Udvikleren? Forbrugeren selv? Eller har AI assistenterne egne rettigheder - og dermed også ansvar? Har de ansvar på samme måde som en bankrådgiver? Skal AI assistenten overhovedet have mulighed for - med fuldmagt - at indgå kontrakter på vegne af forbrugeren?
- 6) **Kill switch.** Bør der være mulighed for, at en AI tjeneste, der er løbet ud på et autonomt sidespor, eller som bare ikke længere agerer i forhold til sit oprindelige formål, kan lukkes helt ned?

²² Se f.eks.: *Another robo-adviser enters the fray: Tiller lets you invest in passive AND active funds - so how does it compare?* <http://www.thisismoney.co.uk/money/investing/article-5941775/Tiller-lets-invest-passive-active-funds-worth-it.html>

Scenarie 2 Nye betalingsformer

I supermarkedet kan du vælge at betale på flere måder. Du kan betale med danske kroner i kontanter, men du kan også vælge at betale med en alternativ digital valuta med egen værdi, som fx bitcoin, du kan betale med personlige data eller med anmeldelser, som du skriver på supermarkedets online forum. AI er en naturlig del af de alternative betalingssystemer og bruges til risikoanalyser og forudsigelser omkring din adfærd, når du betaler med data og af banker og stater til at bekæmpe kriminalitet.

I dag betaler vi primært med en valuta, hvor værdien er reguleret af staten, og transaktionerne og bevægelserne fastsat af bankerne. Men vi ser i stigende grad nye valutaer med alternative værdier fastsat af virksomheder. Sociale medier, hvor vi 'betaler med' persondata er et eksempel. Digitale management systemer (fx Paypal) eller crypto valuta (fx Bitcoin) er et andet. Der er også udviklet betalingssystemer, hvor man betaler eller får fastsat sin forsikringspræmie efter, hvordan man opfører sig, pay-as-you-behave²³. AI bruges her til analyse af forbrugernes data - risikoanalyser og forudsigelser, som deres betaling eller ydelser fastsættes efter. Der er i hvert fald en fordel i de alternative betalingssystemer, nemlig at tjenesterne er 'gratis'. Det giver flere ressourc svage mulighed for at få en bankkonto eller kredit. I disse AI drevne betalingssystemer er der en række etiske problemstillinger:

- 1) Snyd, kriminalitet, identitetstyveri.** Udviklingen af alternative betalingsformer er truet af tjenester, hvor kriminelle hackere står bag for fx at stjæle identiteter og penge. Udviklingen af AI systemer vil kunne identificere og til dels eliminere skattesnyd og kriminalitet. Men hvilke interesser favoriserer de? Og hvilken indflydelse vil det anonyme crypto marked have på det vestlige bank-system og staters muligheder for at opkræve skat?
- 2) Gennemsigtighed.** Forstår forbrugerne værdien af de alternative valutaer? Et eksempel er den personlige data, vi betaler sociale medier med. Hvad er en fødselsdato værd? En kronisk sygdom? En seksuel orientering? Dette er en betalingsform, som ikke er reguleret som fx kronen, hvor man kan få indsigt i kursen. EU kommissionen har via Margrethe Vestager talt for at gøre de nye valutaer mere gennemsigtig for forbrugeren²⁴. Men hvordan sikres gennemsigtighed i værdisættelsen af de nye typer valutaer?
- 3) Digitale kløfter.** For de ressourc svage er det ikke ligetil at få adgang til fx lån. Ikke i Danmark og endnu mindre i Indien og USA. De vil derfor også være de første til at vælge de alternative betalingssystemer, som ikke er regulerede og ikke giver den samme forbrugerbeskyttelse. De bliver dermed mere udsatte for betalingsformer baseret på dataprofilering- og analyse.

Scenarie 3 Algo traders styrer de finansielle markeder

Der er ikke længere nogen mennesker på det internationale børsmarked og aktiemarked. Alle handler foretages af algo traders, dvs AI software, der analyserer markedet, og foretager handler. Der findes både algo traders blandt traditionelle handelsvirksomheder, men også helt almindelige mennesker har fået deres egne personlige algo traders.

For få år siden blev de fleste handler på det internationale børsmarked og aktiemarked foretaget af mennesker. I dag stiger andelen af handler foretaget af maskiner (algo traders). Det er AI drevet software, der analyserer markedet ved at ved fx at inddrage sociale medie data og nyheder og så træffe beslutninger og handle på baggrund af data.

²³ <https://dataethics.eu/pay-as-you-behave-aendrer-vores-forsikringsmodel/>

²⁴ https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/competition-big-data-world_en

- 1) **Computerlogik vs menneske.** Den globale økonomi er under forandring; fra at være baseret på beslutninger taget af mennesker til beslutninger taget af maskiner. Det er processer, der går hurtigere, end hvad mennesker kan opfatte og handle på. Algo traders er udviklet til at maksimere profit. Hvor kommer de etiske og sociale overvejelser ind i billedet? Er en algo trader mindre moralsk end et menneske? Kan man udvikle et moralsk baseret algo trading marked? Og har et menneske altid ret til at vide, at de rådgives med en maskine, ikke et menneske?
- 2) **Manipulation og autonome beslutninger.** Et marked kan crashe, hvis der sniger sig forkert eller misvisende information ind i de data, som algo traderen analyserer. Hvordan sikrer vi os mod manipulation med henblik på at crashe et marked? Kan og skal vi altid sikre menneskelig kontrol? Eller skal vi give computerne lov til at tage autonome beslutninger? Hvad vil konsekvenserne være? Økonomiske, sociale og menneskelige?

Anbefalinger

MENNESKET I CENTRUM

- Vi bør ikke udvikle AI med det formål at erstatte mennesket men sikre, at AI altid komplementerer mennesket²⁵ og 'kun' er en *cobot*, altså en robot, der samarbejder (collaborate) med mennesket²⁶.
- Enhver tjeneste og ethvert (IoT) produkt bør have en kill switch, en mulighed for at mennesket kan slå systemet fra, hvis fx en AI tjeneste løber løbsk eller ikke agerer i forhold til sit oprindelige formål. Tjeneren/produktet skal altid kunne fungere manuelt.
- Sikre, at mennesket altid har det sidste ord. Og at et menneske altid ved, om han/hun taler med en maskine eller med et menneske. Selv om GDPR burde sikre dette, er det nødvendigt med en reel menneskelig vurdering for at sikre imod hovedløs automatisering.
- Sørge for at definere smarte byer på en måde, der ikke kun handler om tal og beregninger, men om smarte mennesker, der tager hånd om egen livslang uddannelse, om mangfoldighed og åbenhed, jvf den østrigske beregningsmodel²⁷.
- Privatlivsrisici og dataetiske implikationer for individet bør altid tænkes ind i design og udviklingen af fx de personlige AI assistenter fra begyndelsen. Privacy by design principper og individuel datakontrol bør være obligatorisk, som også GDPR lægger op til.

INDIVIDUEL DATAKONTROL

- Etablering af 'personal data store'²⁸. En udvidet 'min side' eller en personlig cloud, hvor borgeren/forbrugeren har kontrol over sine data. Borgeren/forbrugeren kan selv tilføje flere data for fx at få flere tjenester eller slette data (der ikke skal opbevares pga lovgivning - og det gælder fx også hvis du har haft en kræftdiagnose, men er erklæret rask og ønsker at få det slettet), til at opdatere data og til at downloade data, som man kan tage med sig (portabilitetsprincippet).
- Sikre danskerne retten til anonymitet - fuldkommen som man har ret til en hemmelig adresse. Ens data - herunder DNA-data, som bliver mere og mere værd, skal opbevares under stærk kryptering og eneste adgang er med en dommerkendelse. Anonymitet medfører også, at vi har ret til at kommunikere og handle anonymt.

GENNEMSKUELIGHED

- Udvikle en standard for, hvordan man forklarer en algoritme. En spørgeramme, der sikrer at man kommer hele vejen rundt, som [DataEthics.eu](https://www.dataethics.eu) har gjort det med dataetiske principper²⁹. Man kunne tage udgangspunkt i DTUs Safe AI principper³⁰ og udvikle dem i samarbejde med DTU. Og få inspiration hos finske Silo.ai, som har udviklet en SHAP standard³¹ der forklarer en kompleks algoritme.

²⁵<https://www.fastcompany.com/40570350/5-ways-to-ensure-ai-unlocks-its-full-potential-to-serve-humanity>

²⁶ <https://blog.universal-robots.com/collaborative-robots-ushering-in-industry-5.0>

²⁷ <http://www.smart-cities.eu/?cid=2&ver=4>

²⁸ <https://dataethics.eu/nye-personal-data-stores-give-individer-datakontrollen-tilbage/>

²⁹ <https://dataethics.eu/dataetiske-principper/>

³⁰ <https://dataethics.eu/faq-dataetiske-principper/>

³¹ silo.ai taler på dataethics.eu/conference18 og der vil blive skrevet en artikel om SHAP modellen

De skal ikke blot give basal information om databehandlingen, når en algoritme forklares. Det skal være dokumenteret og forklaret, hvordan en given algoritmisk beslutning er truffet; hvilke kriterier og parametre der fx ligger til grund for tildeling eller afvisning af en offentlig ydelse, et kreditscore eller størrelsen på en bøde.

ANSVARLIGHED

- Udover en forståelig privatlivspolitik, som GDPR kræver, bør virksomheder, organisationer og myndigheder udvikle egne dataetiske principper, hvor de går længere end compliance af datalovgivningen og forklarer i detaljen, hvad de gør med fx metadata og anonymisering. I udviklingen af disse bør man tage udgangspunkt i eksisterende standarder, som IEEEs etiske standarder for AI32 samt etiske principper og guidelines, der bliver lanceret i 2019³³.
- AI etiske og sociale "Impact Assessments" bør gøres obligatoriske ligesom en Data Protection Impact Assessment er det i dag³⁴.
- Støt op om europæiske og globale mærkningsordninger og AI audits af uafhængige tredjeparter, som viser det inden for miljøet og fødevarer. AI design skal kunne revideres af uafhængige instanser.

LIGE VÆRDIGHED

- Det bør være et krav - i fx en mærkningsordning - at der arbejdes for at mindske bias ved manuel sortering og oprydning i data. Det kan mindskes ved at altid være forklarlige og åbne overfor revision.
- Ved personlige tjenester såsom personlig medicin, er det vigtigt at inddrage de større strukturelle spørgsmål.
- Der skal tænkes lighed ind i udviklingen og designet af AI tjenester og systemer, således også ressourcerne får gavn af udviklingen.

AI SOM FÆLLES SAMFUNDSGODE

- Udviklingen af AI begyndte som eksperimentelle silo projekter. I dag har AI teknologier afgørende indflydelse på vores fælles økonomiske, kulturelle, sociale og politiske processer. Men AI teknologier udvikles stadig i siloer; i forskningsuniverser - omend al universitetsforskning er åbent tilgængeligt - eller som forretningshemmeligheder i virksomheder. Det er tid til, at vi definerer AI som en fælles gode i samfundet og ikke som en enkelt institution eller virksomheds ejendom, ligesom vi engang bestemte os for, at elektricitet er noget alle bør have adgang til.³⁵

³² <https://ethicsinaction.ieee.org/>

³³ <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>

³⁴ <https://dataethics.eu/en/time-data-ethics-impact-assessment/>

³⁵ <https://dataethics.eu/en/lets-talk-ai/>

- Etisk teknologidesign er et vidt begreb, men pt. er man på globalt plan i gang med at udvikle design standarderne for AI. Standardorganisationen ISO har fokus på standarder for AI.³⁶ Det samme har IEEE, der med en række standardprojekter (P700s) er i gang med at udvikle globale etiske standarder for udviklingen af AI. Disse bliver udviklet af eksperter fra hele verden inden for forskellige sektorer og er således baseret på mange forskellige interesser.³⁷ En anbefaling er at følge med og deltage i udviklingen af disse standarder.
- AI er et globalt konkurrencefelt regioner imellem. EU Kommissionens strategi for AI, lanceret i maj 2018 med opbakning fra 25 medlemsstater, indbefatter udvikling af policy og etiske guidelines såvel som en større økonomisk investering i feltet. Der er hermed dannet grobund for en europæisk vinkel på udviklingen af AI med særlig fokus på de etiske og menneskelige dilemmaer. En hovedanbefaling er at følge med den europæiske etiske tilgang til AI, der bør ses som et etisk alternativ og dermed også en konkurrencefordel på det globale marked.

³⁶ <https://www.iso.org/committee/6794475.html>

³⁷ Se standarderne her <https://ethicsinaction.ieee.org/>

Definitioner

Artificial intelligence

Kunstig intelligens (AI) er systemer, der udviser intelligent adfærd ved at analysere omgivelserne og handle med en vis grad af autonomi for at opnå specifikke mål.

Kilde: <https://eur-lex.europa.eu/legal-content/DA/TXT/PDF/?uri=CELEX:52018DC0237&from=en>

Bias

Bias er en indbygget skævhed i data og algoritmen. I nogle tilfælde vil den være et etisk problem, fx hvis det betyder man frasorterer kvinder til en stilling. I andre tilfælde vil det ikke være et problem. I denne rapport bruges bias når det er etisk problematisk og bias er lig med indbyggede fordomme og negative stereotyper. Bias kan forekomme i træningsdata, dvs de historiske data, man bruger til at udvikle en selv-lærende algoritme. Bias kan også forekomme i designet af en algoritme, som kan kategorisere og stemple mennesker på en måde, der diskriminerer mellem fx befolkningsgrupper. Dette kan mindskes ved f.eks. manuel sortering og oprydning i data. Og det kan mindskes ved at altid være forklarlige og åbne for revision.

Kilde: <https://dataethics.eu/dataetiske-principper/>

Deep Learning

En form for 'machine learning', hvor data bliver analyseret gennem selv-lærende matematiske netværk ("neurale netværk") udviklet med inspiration fra den måde neuroner opererer på i menneskets hjerne. Når træningsdata processes i netværket tilpasser de forskellige dele i netværket sig, hvilket styrker dets evne til at lave forudsigelser på data.

Kilde: <https://www.wired.com/story/guide-artificial-intelligence/>

Machine Learning

Brugen af data eller erfaring til at forbedre den måde en computer laver forudsigelser eller udføre en opgave. Det handler om at gøre computeren i stand til at udregne eller analysere automatisk med meget stor hastighed ved hjælp af store mængder og ofte skiftende data.

Kilder: <https://www.wired.com/story/guide-artificial-intelligence/> <https://www.bernardmarr.com/default.asp?contentID=1272>

Personlig medicin (precision medicine) betyder at indsamle, integrere og analysere flere forskellige data-kilder med det formål at udvikle individualiserede indsigter og anbefalinger om sundhed og sygdom.

Kilde: <https://datasociety.net/output/fairness-in-precision-medicine/>

Sundhedsdata

Hver gang du har kontakt med sundhedsvæsenet, uanset om det er apoteksbesøg eller hospitalsbehandling, skabes sundhedsdata. Men sundhedsdata kan også opsamles fra f.eks. apps, wearables og insidables. Sundhedsdata kan både være lagret fysisk eller i elektroniske journaler. Både den information der anvendes på individuelt niveau og den der anvendes på nationalt plan, er sundhedsdata.

Kilde: https://danskdesigncenter.dk/sites/default/files/pdf/sundhed_og_det_gode_liv_data_og_kunstig_intelligens_siri-kommissionen_januar_2018.pdf

Flere definitioner: <https://dataethics.eu/faq-dataetiske-principper/>

Litteratur

Indledning:

- Dataetik - den nye konkurrencefordel, kapitel 12
<https://dataethics.eu/wp-content/uploads/DATAETIK-2-udgave-2.pdf>
- EU kommissionens strategi for AI: Kunstig Intelligens for Europa <https://eur-lex.europa.eu/legal-content/DA/TXT/PDF/?uri=CELEX:52018DC0237&from=en>

Sundhed:

- Fairness in Precision Medicine <https://datasociety.net/output/fairness-in-precision-medicine/>
- BioEthics <http://nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf>
- Genetiske test i virksomheder <https://www.nytimes.com/2018/04/15/technology/genetic-testing-employee-benefit.html>

Transport:

- EKSPONERET, kapitel 'Den smarte by kigger med' af Birgitte Kofod Olsen, maj 2018 <https://dataethics.eu/eksponeret/>

Finans:

- Bygget på Interview med Francesco Lapenta, RUC, 14.6.2018
- IEEE P7006 Personal AI Agent principles <http://sites.ieee.org/sagroups-7006/working-group-progress/>
- IEEE P7006 Personal AI Agent PAR http://sites.ieee.org/sagroups-7006/files/2017/05/P7006_PAR_Detail.pdf
- Automated decision making and artificial intelligence - a consumer perspective http://www.beuc.eu/publications/beuc-x-2018-058_automated_decision_making_and_artificial_intelligence.pdf

Tværgående afsluttende etiske principper:

- Statement on Artificial Intelligence, Robotics and Autonomous Systems, EGE https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf
- IEEE P7006 Personal AI Agent principles <http://sites.ieee.org/sagroups-7006/working-group-progress/>
- IEEE Ethically Aligned Design <https://ethicsinaction.ieee.org/>
- DATAETIK - Principper og pejlemærker for virksomheder, myndigheder og organisationer <https://dataethics.eu/dataetiske-principper/>
- Kunstig Intelligens - de dataetiske spørgsmål <https://dataethics.eu/kunstig-intelligens-de-dataetiske-spoergsmaal/>
- **ALGORITHMIC IMPACT ASSESSMENTS: A PRACTICAL FRAMEWORK FOR PUBLIC AGENCY ACCOUNTABILITY**, Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker, 2018
- Montreal Declaration of Responsible AI: <https://www.montrealdeclaration-responsibleai.com/the-declaration>